Analysis and design of gammatone signal models

Stefan Strahl[a)]
International Graduate School for Neurosensory Science and Systems, Carl von Ossietzky University, D-26111 Oldenburg, Germany

Alfred Mertins
Institute for Signal Processing, University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany

(Received 14 January 2009; revised 22 June 2009; accepted 29 July 2009)

An established model for the signal analysis performed by the human cochlea is the overcomplete gammatone filterbank. The high correlation of this signal model with human speech and environmental sounds [E. Smith and M. Lewicki, Nature (London) **439**, 978–982 (2006)], combined with the increased time-frequency resolution of sparse overcomplete signal models, makes the overcomplete gammatone signal model favorable for signal processing applications on natural sounds. In this paper a signal-theoretic analysis of overcomplete gammatone signal models using the theory of frames and performing bifrequency analyses is given. For the number of gammatone filters $M \geq 100$ (2.4 filters per equivalent rectangular bandwidth), a near-perfect reconstruction can be achieved for the signal space of natural sounds. For signal processing applications like multi-rate coding, a signal-to-alias ratio can be used to derive decimation factors with minimal aliasing distortions.
 [DOI: 10.1121/1.3212919]

PACS number(s): 43.60.Hj [DOS] Pages: 2379–2389

I. INTRODUCTION

The earliest theoretical signal analysis model, proposed by Fourier,[1] analyzes the frequency content of a signal using the expansion of functions into a weighted sum of sinusoids. Gabor[2] extended this signal model using shifted and modulated time-frequency atoms which analyze the signal in the frequency as well as in the time dimension. With the wavelet signal model, a further improvement was presented by Morlet *et al.*[3] using time-frequency atoms that are scaled dependent on their center frequency. This yields an analysis of the time-frequency plane with a non-uniform tiling. The time-frequency atoms used in these signal models normally do not assume an underlying signal structure. As the performance of subsequent processing algorithms depends strongly on how well the fundamental features of a signal are captured, it is favorable to use time-frequency atoms that are specialized to the applied signal class. In this paper we are concerned with the signal class of natural sounds such as speech or environmental sounds, which have been found to be highly correlated with gammatone time-frequency atoms.[4,5] The signal-dependent properties of gammatone atoms are their non-uniform frequency tiling of the time-frequency plane and their asymmetric envelope.[6] A gammatone filterbank is furthermore an established model for the human auditory filters.[7–12] Several analysis-synthesis systems have been proposed using gammatone filters in the analysis and time-reversed filters in the synthesis stage,[13–16] including low-delay[17] and level-dependent asymmetric compensation[18] concepts.

Overcompleteness in signal models has advantages in signal coding applications. It enables sparse signal models like matching pursuit[19] (MP) to search for the sparsest signal representation from the resulting infinite number of possible encodings. Overcompleteness further introduces a robustness toward noise.[20,21] Generally, the choice of the number of time-frequency atoms in a signal model, hence the choice of overcompleteness, is nontrivial. In this paper we are therefore also concerned with the trade-off between the achieved performance in the subsequent processing algorithms and the introduced computational load. To derive the minimal number of time-frequency atoms needed to realize an overcomplete gammatone signal model that can adequately analyze the signal space, we use the theory of frames[22–25] which is a generalization of signal representations based on transforms and filterbanks. A second parameter that can control the overcompleteness of the gammatone signal model is the number of removed analysis filter coefficients. Such a decimation of the filter coefficients introduces aliasing distortions that should not only be kept to a minimum but should also be steered to cancel out in the synthesis stage of the filterbank. Therefore we performed a bifrequency analysis[26] in addition to a frame-theoretic analysis of overcomplete decimated gammatone signal models. We show how a signal-to-alias ratio (SAR) can be used to derive optimal sets of decimation factors with minimal aliasing distortions at a given total decimation factor.

This paper is organized as follows. In Sec. II we introduce the analyzed overcomplete gammatone signal models. In Sec. III we present a frame-theoretic analysis of a non-decimated and a decimated overcomplete gammatone signal model by performing an eigenanalysis of the frame operator.[27] We further show how these results can be used to select the optimal number of atoms for an overcomplete

[a)]Author to whom correspondence should be addressed. Electronic mail: stefan.strahl@uni-oldenburg.de

gammatone signal model. In Sec. IV we show how optimal decimation factors with minimized distortion artifacts can be derived using the bifrequency system analysis.[26] We then analyze these theoretically derived optimal parameters in Sec. V in several audio coding examples.

A. Notation

Matrices and vectors are printed in boldface. $\|\cdot\|$ denotes the Euclidean norm of a vector. $\langle\cdot,\cdot\rangle$ is the inner product of a vector space. $\mathbb{Z}$ is the set of all integers, $\mathbb{R}$ is the set of all real, and $\mathbb{C}$ is the set of all complex numbers. $[a,b] := \{x | a \leq x \leq b\}$ represents the set of all numbers between and including a and b. The superscript * denotes the complex conjugate of a complex number and the superscript ^{H} the conjugate transposition of a complex $m \times n$ matrix. The asterisk $*$ denotes convolution. The argument of the maximum of a function $f(x)$ is denoted as $\arg\max_x f(x)$.

II. OVERCOMPLETE GAMMATONE SIGNAL MODEL

A. Gammatone function

In 1960, Flanagan[28] used a gammatone function as a model of the basilar membrane displacement in the human ear. Johannesma[29] further showed in 1972 that a gammatone filter can be used to approximate responses recorded from the cochlear nucleus in the cat. In 1975, de Boer[30] used a gammatone function to model impulse responses from auditory nerve fiber recordings in the cat, which have been estimated using a linear reverse-correlation technique. The term "Gamma-tone" was introduced in 1980 by Aertsen and Johannesma.[31] Patterson *et al.*[8] stated in 1988 that the gammatone filter also delineates psychoacoustically determined auditory filters in humans. A gammatone filter is defined as

$$\gamma[n] = a n^{\nu-1} e^{-\lambda n} e^{2\pi i f_c n}, \tag{1}$$

with the amplitude a and the filter order ν. The damping factor λ is defined as $\lambda = 2\pi b \mathrm{ERB}(f_c)$ (ERB denotes equivalent rectangular bandwidth) with the center frequency f_c. The parameter b controls the bandwidth of the filter proportional to the ERB of a human auditory filter. For humans, the parameters $\nu = 4$ and $b = 1.019$ have been derived using notched-noise masking data.[32] For moderate sound pressure levels, Moore *et al.*[33] estimated the size of an ERB in the human auditory system as $\mathrm{ERB}(f_c) = 24.7 + 0.108 f_c$. The center frequencies of the gammatone filters are equally spaced on the ERB frequency scale.[34] The scale is defined as the number of ERBs below each frequency with $\mathrm{ERBS}(f_c) = 21.4 \log_{10}(0.004\,37 f_c + 1)$. This non-uniform distribution of the center frequencies (see Fig. 1) correlates with the $1/f$ distribution of frequency energy found in natural signals.[35] It is one of the signal-dependent features of a gammatone signal model. The frequency-dependent bandwidth resulting in narrower filters at low frequencies and broader filters at high frequencies is also an important feature of the gammatone time-frequency atoms. In Sec. III we will show that this enables the signal model to form a snug frame. The third signal-dependent feature of gammatone time-frequency atoms is the asymmetric envelope of the gammatone function,[6] which can also be found in natural sounds, exhibiting a short transient followed by an exponentially damped oscillation.

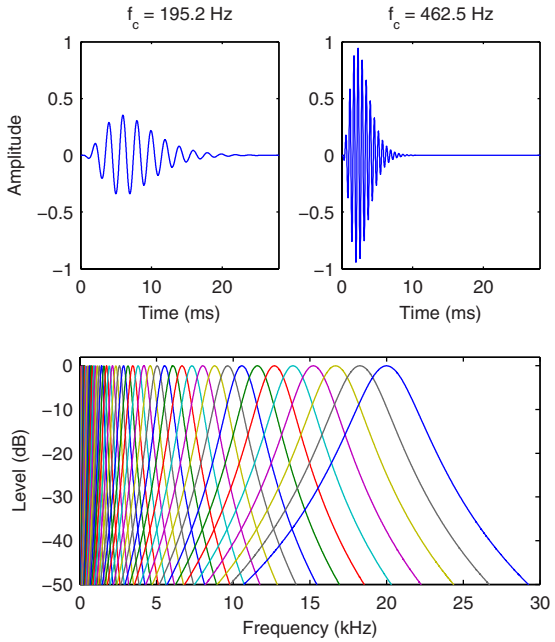

FIG. 1. (Color online) In the upper row the waveforms of two gammatone filters are plotted. The lower row shows the magnitude frequency response of $M = 50$ gammatone filters that are equally distributed along the ERB scale from 20 Hz to 20 kHz.

B. Overcomplete gammatone signal model

To analyze overcomplete gammatone signal models we first have to define a corresponding discrete signal processing system (Fig. 2). The signal $x[n]$ is analyzed with a filterbank where $h_m[n]$, $m \in [0, M-1]$ denotes the impulse responses of M gammatone filters. This splits the full-band signal $x[n]$ into M frequency bands (subbands). In many signal processing applications these subbands are subsampled by decimation factors N_m to remove redundancy from the internal representation and thereby reducing the overcompleteness of the signal model. For the maximally decimated case with $1/N_0 + \cdots + 1/N_{M-1} = 1$, a critical sampling is realized, meaning that the amount of data (samples per second) in the transformed domain and for the original signal is the same. For $\sum_{m=0}^{M-1} 1/N_m > 1$ the signal model is overcomplete, and there are more subband coefficients $y_m[n]$ per time unit than input samples $x[n]$. All subband coefficients $y_m[n]$ are then routed into a subband processing block. In this block, further operations could be performed, for example, a quantization of the subband coefficients controlled by a psychoacoustic model (PAM) or a sparse signal model algorithm like

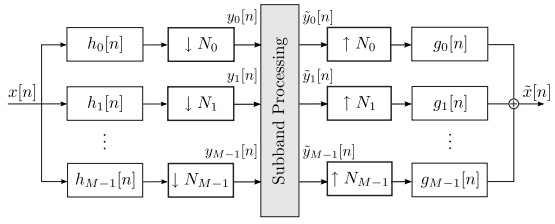

FIG. 2. Discrete signal processing system used to analyze the overcomplete gammatone signal models.

MP (see Appendix A). After the subband processing, the signal $\tilde{x}[n]$ is reconstructed from the M processed subband signals $\tilde{y}_m[n]$ by upsampling with N_m, followed by the synthesis filterbank with the filters having impulse responses $g_m[n]$, $m \in [0, M-1]$.

The analysis presented in this paper is applicable for two different variations in the gammatone signal model. The first variation uses gammatone analysis filters $h_m = \gamma[n]$ and reversed gammatone synthesis filters $g_m = \gamma[-n]$. This is the most commonly used design, for example, in audio coding applications.[13,14,16] The second variation uses reversed gammatone analysis filters $h_m = \gamma[-n]$ and gammatone synthesis filters $g_m = \gamma[n]$. This system can be used to perform a fast MP analysis with a gammatone dictionary (see Appendix A). By choosing the synthesis filters as the time-reverse of the analysis filters the overall filterbank response has a linear phase in both designs.

A gammatone signal model is normally designed to cover only a limited frequency range.[9–14,16,36] Consequently, the analyses in this paper have been conducted using such bandlimited gammatone signal models. We distributed the center frequencies of the gammatone filters equally spaced on the ERB scale within the interval $f_c \in [20, 20\,000]$ Hz, which represents the approximated human hearing range.[37]

III. FRAME-THEORETIC ANALYSIS OF AN OVERCOMPLETE GAMMATONE SIGNAL MODEL

In this section, we will perform a frame-theoretic analysis of the overcomplete gammatone signal model. We will introduce the theory of frames and use it to evaluate the properties of the corresponding frame of a non-decimated and a decimated gammatone signal model. All calculations have been performed with a sampling rate of 96 kHz, and the length of the impulse responses $h_m[n]$ and $g_m[n]$ was 8192 samples or 85.3 ms, respectively.

A. The theory of frames

The theory of frames provides a mathematical framework to analyze overcomplete signal models.[23–25] A *frame* of a vector space $\mathbf{V}$ is a set of vectors $\{\mathbf{e}_m\}$ which satisfy the following *frame condition*:[25]

$$A\|\mathbf{v}\|^2 \leq \sum_m |\langle \mathbf{v}, \mathbf{e}_m \rangle|^2 \leq B\|\mathbf{v}\|^2 \quad \forall\, \mathbf{v} \in \mathbf{V}, \tag{2}$$

with the *frame bounds* $A > 0$ and $B < \infty$. Frames can be seen as a generalization of bases, as the set $\{\mathbf{e}_m\}$ is allowed to be linearly dependent, and Eq. (2) implies that the set $\{\mathbf{e}_m\}$ must span the vector space $\mathbf{V}$. Otherwise it would follow $A = 0$ from $\langle \mathbf{v}, \mathbf{e}_m \rangle = 0$ for $\mathbf{v} \in \mathbf{V} \backslash \mathrm{span}\{\mathbf{e}_m\}$.

The frame condition can also be written as $A\|\mathbf{v}\|^2 \leq \langle \mathbf{S}\mathbf{v}, \mathbf{v} \rangle \leq B\|\mathbf{v}\|^2$ with $\mathbf{S}$ being the *frame operator* defined as

$$\mathbf{S}\mathbf{v} = \sum_m \langle \mathbf{v}, \mathbf{e}_m \rangle \mathbf{e}_m. \tag{3}$$

The frame bound A is the essential infimum and the frame bound B is the essential supremum of the eigenvalues of $\mathbf{S}$.[25] A frame is called *tight* if $B/A = 1$ and *snug* if $B/A \approx 1$. The advantage of a tight frame is that perfect reconstruction can be done by the frame itself:

$$\mathbf{v} = \frac{1}{A} \sum_m \langle \mathbf{v}, \mathbf{e}_m \rangle \mathbf{e}_m \quad \forall\, \mathbf{v} \in \mathbf{V}. \tag{4}$$

The frame bounds for the discrete signal processing system as shown in Fig. 2, are given by the following inequality:

$$A\|\mathbf{x}\|^2 \leq \sum_{m=0}^{M-1} \sum_{k=-\infty}^{\infty} |\langle \mathbf{x}, \mathbf{h}_{m,k} \rangle|^2 \leq B\|\mathbf{x}\|^2 \quad \forall\, \mathbf{x} \in \ell^2(\mathbb{Z}), \tag{5}$$

with $m \in [0, M-1]$, $k \in \mathbb{Z}$, and the vectors $\mathbf{h}_{m,k}$ containing the filter coefficients $h_m(kM-n)$ and $\mathbf{x} \in \ell^2(\mathbb{Z})$ being the vector that contains the input samples $x[n]$.

In general, the smaller the ratio B/A is, the better the numerical properties of the signal model will be. If B/A is close to 1, then the assumption of energy preservation may be used without much error when relating the energy of the subband signals $y_m[n]$ to the energy of the input signal $x[n]$ and the output signal $\tilde{x}[n]$. This is important in audio coding applications, as it guarantees that small quantization errors introduced in the subband signals will result in only small reconstruction errors. It enables a bit allocation optimized for minimum error in the subbands to be near-optimal for the final output signal.

The speed of convergence for algorithms like MP also depends on the frame bounds, as shown in Sec. V. In this context it is to note that the frame realized by a MP decomposition with a dictionary of atoms $\mathbf{e}_k$ is identical to a frame realized by a filterbank with the matched filters $\mathbf{e}_k^*[-n]$, as shown in Appendix A.

The frame operator $\mathbf{S}$ can be represented in the polyphase domain by the $M \times M$ matrix $\mathbf{S}(z) = \tilde{\mathbf{E}}(z)\mathbf{E}(z)$, where $\mathbf{E}(z)$ is the analysis polyphase matrix of the filterbank[38] and the eigenvalues of the frame operator $\mathbf{S}$ equal the eigenvalues $\lambda_n(\theta)$ of the matrix $\mathbf{S}(e^{i\theta}) = \mathbf{E}^H(e^{i\theta})\mathbf{E}(e^{i\theta})$. Bolcskei *et al.*[27] could show that the frame bounds A and B are the essential infimum and essential supremum, respectively, of the eigenvalues $\lambda_n(\theta)$. Thus, the computation of the frame bounds of overcomplete gammatone signal models using their polyphase matrix representations is possible. Note that in the non-decimated case, the frame bounds and respective eigenvalues are related to the ripple in the overall frequency response of the filterbank.

The eigenanalysis of a signal model is only applicable for a limited frequency interval if the corresponding filterbank is non-decimated. For $N_m > 1$, the mapping of the eigenvalues of the frame operator to the analyzed frequency interval is lost. Thereby the essential infimum and essential supremum can only be calculated for the entire frequency range, from zero to half the sampling frequency. This results to a lower frame bound of $A = 0$ for bandlimited signal models, like the here analyzed overcomplete gammatone signal model, where filters do not cover frequencies below 20 Hz and above 20 kHz. To circumvent this problem, we added two additional filters for the frequency intervals not covered by the gammatone filterbank, i.e., a lowpass for the $[0, 20]$ Hz frequency interval and a highpass filter for

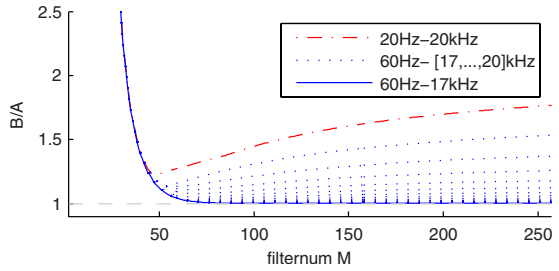

FIG. 3. (Color online) The frame-bound ratios B/A of non-decimated gammatone signal models with the number of filters $M \in [2,256]$ analyzed over the frequency intervals 20 Hz–20 kHz and 60 Hz–[17,20] kHz. For the frequency interval of 60 Hz–17 kHz, the frame-bound ratio converges toward a tight frame for higher filter numbers.

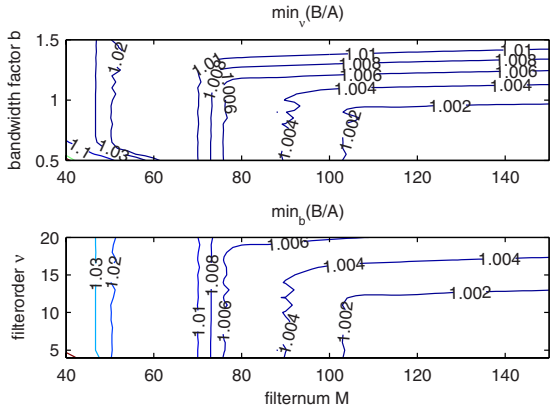

FIG. 4. (Color online) Best possible frame-bound ratios for a fixed bandwidth factor b and filter number M (upper plot) or filter order ν and filter number M (lower plot). The gammatone signal model parameters were $b \in [0.5,1.5]$, $\nu \in [4,20]$, and $M \in [40,150]$ analyzed over the frequency interval from 60 Hz to 17 kHz.

$[20,48]$ kHz. Thereby we could compute A for a decimated gammatone signal model within the limited frequency range. B was computed without additional filters.

B. Analysis of a non-decimated overcomplete gammatone signal model

An overcomplete signal model results in a large quantity of subband coefficients for every filter. To reduce bitcoding and computational costs, it is of interest to know the smallest number M of subbands needed to achieve good frame-bound ratios. As the frame bounds of $\gamma[n]$ are identical to the frame bounds of $\gamma[-n]$, we only need to analyze the frame of the gammatone prototype $\gamma[n]$ itself. The frame bounds A and B of the non-decimated overcomplete gammatone signal model can be computed, as described in Sec. III A, and the respective frame-bound ratios B/A are shown in Fig. 3. The parameters of the analyzed gammatone signal models were $b=1.019$, $\nu=4$ with $M \in [2,256]$ center frequencies between 20 Hz and 20 kHz.

Figure 3 shows that the gammatone signal model does not realize a frame for the frequency interval of its center frequencies. The frame-bound ratio is mainly determined by small eigenvalues of the frame operator $\mathbf{S}$ found at the first and last gammatone filters (see also Fig. 11). The ERB scale distributes the center frequencies of the gammatone atoms in such a way that the overlapping filters result in almost constant eigenvalues. As for the first and the last filters this overlap is not fully realized; the essential infimum of the eigenvalues results in a low lower frame bound A. If we perform the analysis over a reduced frequency interval (see Fig. 3 and Table I), the frame-bound ratio improves and the gammatone signal is able to achieve a snug frame from $M=50$ subbands on. This marginal reduction in the frequency interval is non-critical as it still embeds the class of natural sounds with speech, for example, ranging approximately from 80 Hz to 10 kHz.

TABLE I. Frame-bound ratios B/A analyzed for different bandlimited signals and number of gammatone filters M.

M	Frequency interval	B	A	B/A	Frame
50	[20 Hz, 20 kHz]	1.294	1.046	1.238	Not snug
50	[40 Hz, 17 kHz]	1.294	1.167	1.109	Snug
100	[20 Hz, 20 kHz]	2.462	1.697	1.451	Not snug
100	[60 Hz, 17 kHz]	2.462	2.455	1.003	≈ tight

For $M=50$ the frame bounds are $A=1.167$ and $B=1.294$, which results in a frame-bound ratio of $B/A=1.109$. This means that, depending on the actual signal, the energy of the input or output signal of the filterbank may be different from the subband energy by a factor between 1.167 and 1.294. For higher filter numbers the frame-bound ratio converges toward a tight frame and for $M=100$ a frame-bound ratio of $B/A=1.003$ is achieved.

For applications that allow a deviation from the human gammatone parameters, we also analyzed the influence of the bandwidth parameters $b \in [0.5,1.5]$ and the filter orders $\nu \in [4,20]$ on the frame-bound ratio for the frequency interval from 60 Hz to 17 kHz (see Fig. 4). For $M=50$ gammatone atoms, the best frame-bound ratio $B/A=1.020$ is achieved for a filter order $\nu=11$ and the bandwidth factor $b=0.85$. For $M=100$ the filter order $\nu=12$ and the bandwidth factor $b=0.5$ result in the lowest frame-bound ratio of $B/A=1.003$. The contour plot in Fig. 4 shows that these best frame-bound ratios are located in relatively shallow minima. More generally, we can conclude that for a filter number of $M=50$, snug frames can be achieved with $b>0.7$ and all examined filter orders. For $M=100$ a tight frame is possible with $b \leq 1$, $\nu<13$. Additionally it can be seen that for a small number of filters ($M<50$) larger bandwidths achieve better frame-bound ratios. More interestingly, for a higher number of filters, large filter bandwidths introduce a decline in the frame-bound ratio which is explained in detail in Sec. VI and Fig. 11.

C. Analysis of a decimated overcomplete gammatone signal model

To further reduce encoding and subband processing costs, it is often favorable to remove the redundancy in an overcomplete signal model by downsampling its subband coefficients by factors $N_m>1$. The decimation of the filterbank coefficients can result in distortions, which will worsen the frame-bound ratio of the decimated signal model. Thus, a

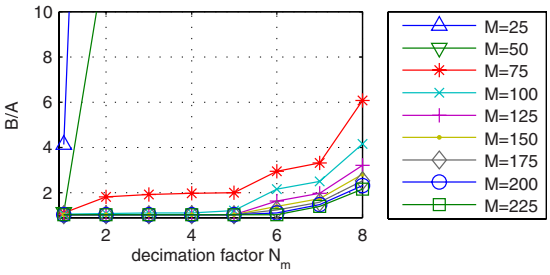

FIG. 5. (Color online) The frame-bound ratios B/A of decimated gammatone signal models with the number of filters $M \in \{25,50,\ldots,200,225\}$ and decimation factors $N_m \in [1,8]$ analyzed over the frequency interval of 60 Hz–17 kHz.

frame-theoretic analysis can be used to analyze the introduced distortions for different decimation factors N_m. We derived frame bounds for a decimated overcomplete gammatone signal model for the frequency interval from 60 Hz to 17 kHz by introducing additional filters to allow the derivation of A, as described in Sec. III A. The resulting frame-bound ratios B/A are shown in Fig. 5. It can be seen that no snug frame can be achieved for $M \leq 75$ filters with an equal decimation of the subband coefficients. For higher filter numbers, a snug frame can be realized up to an equal decimation of the subband coefficients of $N_m=4$, $N_m=5$, and $N_m=6$ for the filter numbers $M=100$, $M \in [125,150]$, and $M \in [175,255]$, respectively.

To derive optimal decimation factors for an overcomplete gammatone signal model, a full search over all possible N_m by computing the corresponding frame-bound ratios would be necessary, which is computationally intractable. It is further to note that distortions that fall into a frequency range where the signal has only little energy will have a minor effect compared to distortions in frequency bands, where most of the signal energy is present. This cannot be exploited by an optimization based on frame-bound ratios due to the lost mapping of the eigenvalues of the frame operator to the analyzed frequency interval. Therefore we introduce and use in Sec. IV an alternative technique to derive optimal decimation factors.

IV. BIFREQUENCY ANALYSIS OF A DECIMATED OVERCOMPLETE GAMMATONE SIGNAL MODEL

To allow the optimization of decimation factors dependent on the applied signal, we will introduce in this section the bifrequency analysis[39] and define a SAR. The bifrequency analysis has the additional advantage that it offers a complete frequency description of the distortions introduced by a decimation of the subband coefficients. This leads to a better insight of the design limitations, i.e., to Conditions I and II as given below. This allows to reduce the computational costs of an optimization of the decimation factors. All results in this section were derived with a sampling rate of 44.1 kHz, which is a common sampling rate in signal processing applications like audio coding. The length of the analyzed impulse responses $h_m[n]$ and $g_m[n]$ has been set to 4096 samples or 92.9 ms, respectively.

A. Bifrequency analysis

An alternative theoretical analysis of the decimated gammatone signal models is possible by the fact that a decimated filterbank can also be understood as a linear time-varying (LTV) system

$$\mathbf{y}[n_y] = \sum_{n_x=-\infty}^{\infty} \mathbf{k}[n_y,n_x]\mathbf{x}[n_x], \tag{6}$$

with a periodic system response $\mathbf{k}[n_y,n_x]=\mathbf{k}[n_y+\ell N, n_x+\ell N]$, $\ell \in \mathbb{Z}$, where $\mathbf{x}[n_x]$ is the input and $\mathbf{y}[n_y]$ is the output sequence. $\mathbf{k}[n_y,n_x]$ denotes the response of the system at the discrete time n_y to a unit sample applied at discrete time n_x. For periodic LTV systems, a bifrequency analysis[39] gives a complete description of the system as well as of its aliasing components. The discrete bifrequency system function[26] is defined as

$$\mathbf{K}[e^{i\omega_y},e^{i\omega_x}] := \frac{1}{2\pi}\sum_{n_y=-\infty}^{\infty}\sum_{n_x=-\infty}^{\infty}\mathbf{k}[n_y,n_x]e^{i\omega_x n_x}e^{-i\omega_y n_y}, \tag{7}$$

relating the input signal spectrum $\mathbf{X}[e^{i\omega_x}]$ to the output signal spectrum $\mathbf{Y}[e^{i\omega_y}]$ with

$$\mathbf{Y}[e^{i\omega_y}] = \int_{-\pi}^{\pi}\mathbf{K}[e^{i\omega_y},e^{i\omega_x}]\mathbf{X}[e^{i\omega_x}]d\omega_x. \tag{8}$$

In the analyzed gammatone signal models, the only periodically time-varying parts are the decimators and interpolators. Therefore, the overall bifrequency map is composed of nonzero unity-slope parallel lines with a constant factor, on whose input and output spectra the effects of the analysis and the synthesis filters, respectively, are projected.[40] The center line represents the time-invariant part of the system; all other lines represent the parts of the system which cause aliasing (see also Fig. 6). As an objective measure of the aliasing distortions in a signal model we used a signal-to-alias (SAR), defined analogous to the commonly used signal-to-noise ratio (SNR). For a given input signal spectrum $\mathbf{X}[e^{i\omega_x}]$ the SAR is defined as

$$\mathrm{SAR}(\mathbf{X}[e^{i\omega_x}]) = -10\log_{10}\left(\frac{T_1^2}{\sum_{n\in\{N_m\}}T_n^2}\right), \tag{9}$$

with

$$T_n = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\delta(n\omega_x-\omega_y)\mathbf{K}[e^{i\omega_y},e^{i\omega_x}]\mathbf{X}[e^{i\omega_x}]d\omega_x d\omega_y, \tag{10}$$

and $\delta(\cdot)$ being the Dirac pulse. The time-invariant part of the system corresponds to T_1, and the aliasing components of the LTV system are represented by the T_n.

To avoid in-band aliasing distortions, N_m must be chosen in such a way that all integer multiples of the decimated Nyquist frequency lie outside the mth passband of a subband [see Fig. 6(b)]. For an aliasing-free signal model this results in the following necessary condition to prevent in-band aliasing.

Condition I. With ω_m^L and ω_m^H being the starting and stopping cutoff frequencies of the mth gammatone filter ($0 \leq \omega_m^L \leq \omega_m^H \leq \pi$) it needs to hold

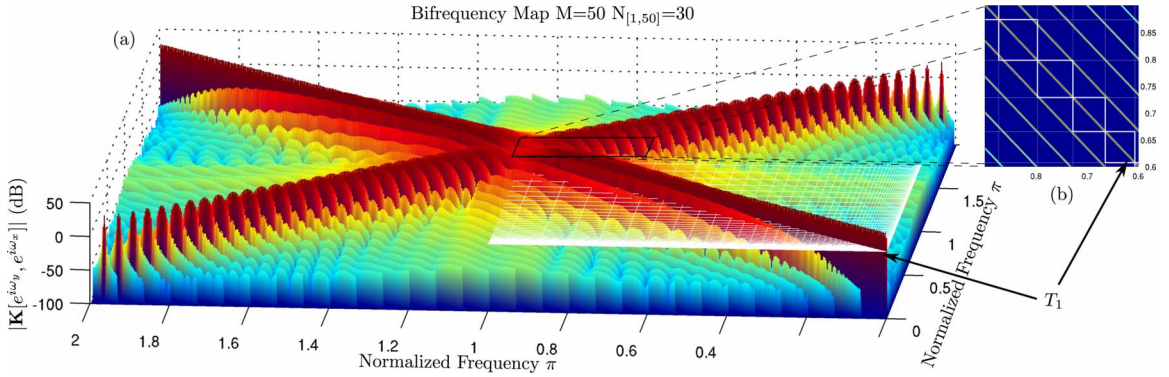

FIG. 6. (Color online) (a) Bifrequency map for a gammatone signal model with the number of filters $M=50$ and the decimation factor of $N_m=30$ in every subband. The axes show the normalized frequency domains associated with the input and output signals. The center line represents the time-invariant part (T_1) that maps the input to the output signal and is independent of any decimation. All other lines are due to aliasing terms ($T_{n>1}$) introduced by a decimation of the subband coefficients. The zoom-in (b) shows that in this example in-band aliasing occurs in the last three filters, in which aliasing components fall into the passband of these filters. The filter's passbands are indicated by a grid of thin white lines.

$$(k\pi/N_m) \notin [\omega_m^L, \omega_m^H] \quad \forall\, k \in \mathbb{N}. \tag{11}$$

This dependency on the bandwidth of the corresponding gammatone filter limits the possible decimation factors to the set which fulfills $N_m < \pi/(\omega_m^H - \omega_m^L)$. In contrast to an ideal bandpass filter, which has a discontinuity in magnitude at the cutoff frequencies, real filters like the gammatone filter exhibit a magnitude response that changes gradually from the passband to the stopbands. A commonly chosen decrease in magnitude to define the cutoff frequency is an attenuation of 3 dB.

Inter-band aliasing can be reduced if the decimation factors are chosen in such a way that an aliasing term of a filter in one subband can be canceled by another aliasing term of a filter in another subband. Such a set of integer decimation factors N_m in which each aliasing term occurs at least twice is called a *compatible set*[38,41,42] and needs to fulfill the following condition.

Condition II. Let $L := \mathrm{lcm}(\{N_m\}_{m=0}^{M-1})$ be the least common multiplier (lcm) of the set of decimation factors $\{N_m\}_{m=0}^{M-1}$. If the set is an apposition of repeated distinct integers $\{\mathcal{N}_1, \mathcal{N}_1, \ldots, \mathcal{N}_1, \ldots, \mathcal{N}_{K-1}, \ldots, \mathcal{N}_{K-1}\}$ with $\mathcal{N}_j \in \{N_m\}_{m=0}^{M-1}$ and n_j denoting the number of $\mathcal{N}_j$ in this set, then it needs to hold

$$\min\left\{ \frac{\mathrm{lcm}\left(\frac{L}{\mathcal{N}_i}, \frac{L}{\mathcal{N}_j}\right)}{\frac{L}{\mathcal{N}_j}} \right\}_{\substack{i=0 \\ i\neq j}}^{M-1} - 1 < n_j. \tag{12}$$

B. Analysis of a decimated overcomplete gammatone signal model

We will use the results from Sec. IV A to show how optimal decimation factors N_m for a given decimated overcomplete gammatone signal model and a given signal spectrum $\mathbf{X}[e^{i\omega_x}]$ can be derived. Let $\mathbf{N} := (N_0, N_1, \ldots, N_{M-1}) \in [1, M-1]^M$ be the M-dimensional vector space of all possible decimation factors for a gammatone signal model. We can reduce the size of $\mathbf{N}$ by allowing only decimation factors that fulfill Conditions I and II. The cutoff frequency was set at 3 dB stopband attenuation. The size of the set of possible decimation factors can be further reduced using the constraint $N_0 \geq N_1 \geq \cdots \geq N_{M-1}$, which is derived from Condition I and the fact that the gammatone signal model has monotone increasing bandwidths. To select decimation factors that form a *compatible set*, the decimation factors can be required to be powers of 2.

To derive for a given degree of overcompleteness $O = \sum_{m=0}^{M-1} 1/N_m$, a set of decimation factors with minimal aliasing distortions, the SAR can be used as a quality measure. To exemplify this, we analyzed an overcomplete gammatone signal model with $M=50$ filters, center frequencies ranging from 20 Hz to 20 kHz, and $N_m \in \{1, 2, 4, 8, 16, 32, 64, 128, 256, 512\}$. We further evaluated if varying the bandwidth of the gammatone filters has an influence on the aliasing distortions. Analyzing Fig. 6, it can be seen that the major aliasing distortions occur in the high-frequency bands due to the nonuniform frequency resolution of the gammatone signal model. For applications like speech or audio coding, where only a small amount of signal energy falls in the high-frequency bands, these distortions will have a minor effect compared to the distortions in the low-frequency band, where most of the signal energy is present. Therefore it is favorable to optimize the decimation factors according to the SAR computed for the specific spectrum of the applied signal class. In this example we used the spectrum of the audio test signal "Tom's Diner" by Vega (svega.wav). Table II shows the SAR achieved by optimal decimation factors (stated in Appendix B), selected from a set of decimation factors that is constructed as described above and that results in the degrees of overcompleteness $O = 1, 2, \ldots, 8$, respectively. They are compared with commonly chosen decimation factors that are inverse-proportional to the bandwidth of the gammatone filters while fulfilling Condition I. The optimized decimation factors achieve a SAR improvement of 4.7 dB on average compared to the commonly chosen decimation factors. This can be seen as a significant improvement, recalling that a SAR improvement of 6 dB means a reduction in the distortion energy due to aliasing components by a factor of 2. As the overcomplete gammatone signal model realizes for $M=50$ only a snug frame, we additionally investigated if the SAR can be improved using different filter

TABLE II. The SAR for svega.wav and a gammatone signal model with $M=50$ filters achieved with optimized decimation factors compared to commonly chosen decimation factors that are inverse-proportional to the bandwidth of the filters while fulfilling Condition I.

SAR (dB)	$O=1$	$O=2$	$O=3$	$O=4$	$O=5$	$O=6$	$O=7$	$O=8$
Optimized N_m	9.5	14.2	15.6	17.5	18.2	18.5	18.9	19.2
Prop. bandwidth	6.2	8.1	10.6	11.3	13.4	14.5	14.6	15.7

bandwidths. It showed that for $M=50$ a deviation from the human bandwidth parameter $b=1.019$ can reduce inter-band aliasing distortions from 1 up to 15.2 dB for $O=1$ and $O=8$, respectively (see Fig. 7). As an increase in the filter bandwidth leads to an increase in the energy in the aliasing components, this reduction in aliasing distortions can be addressed to an optimized cancellation of aliasing terms. So depending on the number of applied gammatone filters, the bandwidth factor b should also be included into the optimization process.

V. APPLICATIONS

In this section we report on the signal reconstruction performance of overcomplete gammatone signal models using the example of audio coding and compare the findings with the theoretical results from Secs. III and IV. We applied a coding scheme whose block diagram is shown in Fig. 2.

In the first experiment, we investigated the signal reconstruction and subband algorithm performance of a non-decimated overcomplete gammatone signal model ($N_m=1$), as analyzed in Sec. III. We tested two signal model variations. In the first variation (GTFB), we evaluated the standard overcomplete gammatone signal model with $h_m=\gamma[n]$, $g_m=\gamma[-n]$ and without subband processing. In the second variation, a sparse overcomplete gammatone signal model was realized with $h_m=\gamma[-n]$, $g_m=\gamma[n]$, and a MP algorithm[19] was performed in the subband processing block. The stopping condition was set to 2000 atoms/s and it was implemented as described in Appendix A. The test signal for this initial audio coding experiment was the commonly used Tom's Diner by Suzanne Vega (svega.wav). In accordance with the theoretically derived results (Fig. 3), the signal reconstruction error decreased for both schemes with an increasing number of filters and saturated for higher filter numbers (Fig. 8). For the overcomplete gammatone signal model (GTFB), near-perfect reconstruction was achieved for $M\geq 100$. For the sparse overcomplete gammatone signal model (MP) the SNR rose to 22.5 dB at $M\approx 70$ and continued to slightly improve further for higher filter numbers until it stayed constant at 23.5 dB for $M\geq 500$ gammatone filters. This shows that the convergence speed of the MP algorithm facilitated also from small frame-bound ratio improvements close to $B/A=1$, as the overcomplete gammatone signal model did not contribute further to the signal reconstruction for $M\geq 100$.

We further evaluated a basic perceptual audio coding scheme by scaling the subband coefficients $y_m[n]$ according to a PAM before performing a fixed quantization.[43] The PAM was realized by the MPEG-2 AAC/MPEG-4 audio standard reference implementation,[44] and a linear 7 bit quantizer was used. The coding and decoding of the scaled and quantized coefficients were assumed to be lossless and therefore omitted. Finally according dequantization and rescaling was performed before the audio signal was reconstructed using the synthesis filterbank. We measured the perceived audio quality of the resulting audio signals relative to the original test signal using a model of auditory perception (PEMO-Q).[45] The estimated perceived audio quality was mapped to a single quality indicator, the objective difference grade (ODG).[46] This is a continuous scale from 0 for "imperceptible impairment," −1 for "perceptible but not annoying impairment," −2 for "slightly annoying impairment," −3 for "annoying impairment" to −4 for "very annoying impairment." As explained in Sec. III, subband processing algorithms like perceptual audio coding rely on the assumption of energy preservation in the signal model. Their performance therefore depends on the achieved frame-bound ratio of the used signal model. As shown in Fig. 9, the GTFB signal model without quantization achieved transparent audio coding from $M>55$ gammatone filters on. Linearly quantizing the subband coefficients to a 7 bit encoding, the ODG converged around $M>45$ to approximately −2.5. Scaling the important subband coefficients before quantization according

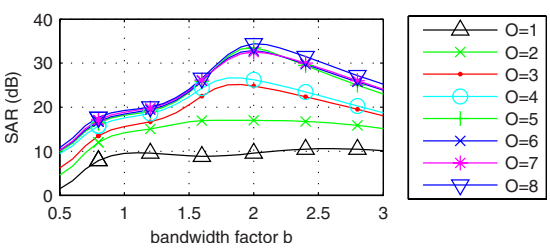

FIG. 7. (Color online) The SAR achieved by optimized decimation factors for a given degree of overcompleteness O and different bandwidth factors b of $M=50$ gammatone filters.

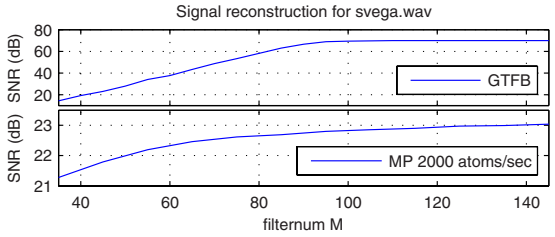

FIG. 8. (Color online) Signal reconstruction experiment using non-decimated overcomplete gammatone signal models for the svega.wav test signal. The upper plot shows the results for a signal model without subband processing (GTFB) and the lower plot shows the achieved SNR for a sparse gammatone signal model based on the MP algorithm.

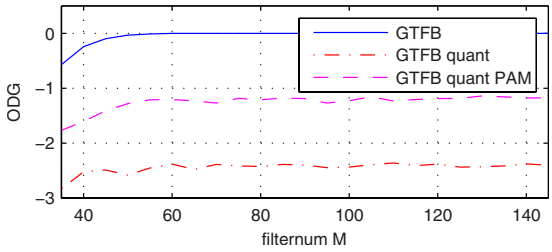

FIG. 9. (Color online) Perceptual reconstruction quality for svega.wav encoded without quantization with a linear quantization and a linear quantization including a PAM.

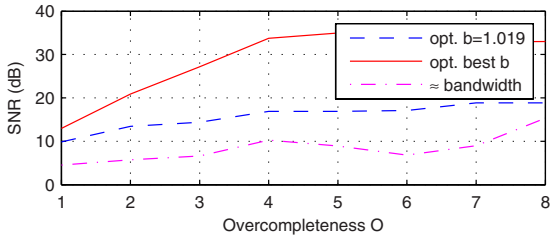

FIG. 10. (Color online) Signal reconstruction experiment using decimated overcomplete gammatone signal models being optimized to maximize the SAR of the test signal (svega.wav), as described in Sec. IV B, compared to commonly chosen decimation factors that are inverse-proportional to the bandwidth of the filters while fulfilling Condition I.

to a PAM showed an improvement in the perceived audio quality until $M \approx 60$ where an ODG of approximately -1.2 is achieved. With the results from Sec. III it can be concluded that for audio coding applications at least a snug frame should be realized by the gammatone signal model. Clearly, to further improve the quality up to an ODG of zero, finer quantization is needed.

In the second experiment, we investigated the signal reconstruction performance of decimated overcomplete signal models with $M=50$ filters and without any subband processing. As a reference signal model we selected commonly chosen decimation factors that are inverse-proportional to the bandwidth of the gammatone filters, while fulfilling Condition I. We compared their achieved signal reconstruction performance with optimized decimation factors for a gammatone signal model having a fixed bandwidth factor $b=1.019$ and for a gammatone signal model where also the bandwidth of the filters was optimized, as described in Sec. IV B. The audio test file was svega.wav, and the results are plotted in Fig. 10. It can be seen that the decimation factors optimized to maximize the SAR of the audio signal as described in Sec. IV B result in a better SNR than the N_m that are increased proportional to the filter bandwidth and fulfill Condition I. It further shows that for the snug frame realized with $M=50$ gammatone filters, a deviation from the human bandwidth parameter $b=1.019$, if allowed in the context of the application, can reduce the aliasing distortions and improve the signal reconstruction performance.

VI. DISCUSSION

Applications that use an overcomplete gammatone signal model can be divided into two groups. The first group is concerned with modeling the auditory system. In these studies, the number of auditory filters is inferred from a reasonable filter spacing determined by the estimated bandwidths of the auditory filters. A common value used is 1 filter per ERB,[9,10,45] which results in 39 filters for the human cochlea, whose basal end corresponds to 38.9 on the ERB scale.[47] The second group of applications is concerned with signal processing tasks, for example, audio coding and speech recognition. Hereby, not an accurate replication of the auditory system is strictly needed, but a maximal performance of the algorithm is desired. Therefore, the number of gammatone filters should be chosen optimizing the performance of the subsequent processing algorithms and the introduced computational load. Most signal processing applications using an overcomplete gammatone signal model so far have used psychoacoustically derived filter numbers, which do not result in a frame (see Table III). As shown in Sec. V, subband processing algorithms like MP or a perceptual quantizer show an improved performance for improved frame bounds.

Note that it is not self-evident that an overcomplete gammatone signal model can achieve a snug frame and converge to a tight frame. The parameters of the gammatone function have been derived from psychoacoustic experiments and are not specifically designed to realize a frame in the mathematical sense. Further analysis of the eigenvalues showed that at higher filter numbers ($M>60$), the frame-bound ratio is determined mainly by the fact that the frequency spacing of the ERB scale does not fully match the filter overlap to the filter bandwidths. This introduces a positive shift of the largest eigenvalues toward higher frequencies (see Fig. 11). Therefore we evaluated if marginal alter-

TABLE III. Examples for gammatone signal model parameters found in the literature. The frame-bound analysis was performed on a limited frequency interval to exclude distortion effects from the first and last filters.

Paper	Interval of center frequencies	M	Given rational	Filter per ERB	B/A	Frame-bound analysis interval	Frame
Ambikairajah *et al.*[36]	50 Hz–7.0 kHz	21	"Ripple within 1.5 dB"	⋯	1.481	100 Hz–7 kHz	Not snug
Brucke *et al.*[56]	73 Hz–6.7 kHz	30	1 filter per ERB	1.0	1.322	70 Hz–6.2 kHz	Not snug
Feldbauer *et al.*[16]	100 Hz–3.6 kHz	50	Frame-bound ratio	2.2	1.003	150 Hz–3.0 kHz	≈ tight
Hohmann[17]	70 Hz–6.7 kHz	30	1 filter per ERB	1.0	1.332	65 Hz–6.3 kHz	Not snug
Kubin and Kleijn[14]	100 Hz–3.6 kHz	20	"Physiologically-motivated"	0.9	1.364	190 Hz–3.1 kHz	Not snug
Lin *et al.*[15]	<4 kHz	25	Not stated	0.9	1.572	35 Hz–4.0 kHz	Not snug
Ma *et al.*[57]	50 Hz–8.0 kHz	64	"Computational costs"	2.0	1.003	100 Hz–6.2 kHz	≈ tight
This study	20 Hz–20.0 kHz	50	Frame-bound ratio	1.2	1.109	60 Hz–17 kHz	Snug
		100	Frame-bound ratio	2.4	1.003	60 Hz–17 kHz	≈ tight

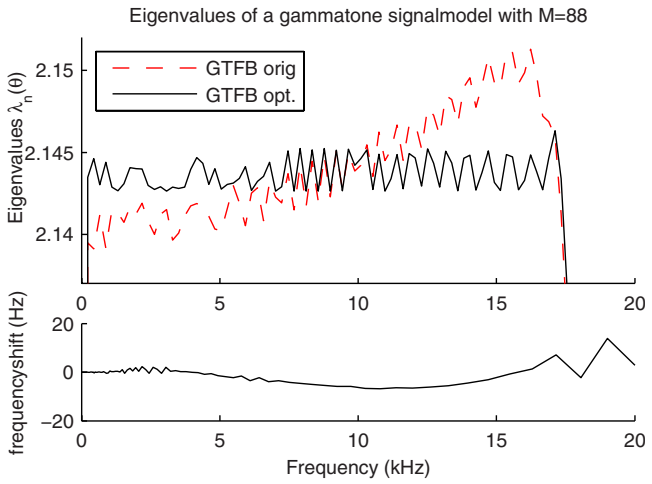

FIG. 11. (Color online) The eigenvalues $\lambda_n(\theta)$ of an overcomplete gammatone signal model with $M=88$ filters being equally spaced on the ERB scale compared to an optimized frequency scale with frequency shifts applied to the ERB scale as shown in the lower row.

ations of the filter's center frequency can improve the gammatone signal model. Using the frame-bound ratio as a cost function, a standard optimization algorithm like the MATLAB function fmincon can be used to derive the frequency shifts necessary to remove the monotonic shift. The derived frequency shifts reduced the center frequencies slightly at middle frequencies, compensating this with a frequency increase at the lower and higher frequencies, see also the example shown in Fig. 11. For this example with $M=88$ the frame-bound ratio could be improved from 1.006 to 1.001 by applying only, relative to the center frequency, marginal frequency shifts. Note that these results are only of theoretical interest, as the gammatone signal already forms an almost tight frame at higher filter numbers M, and the derived optimization does not improve the numerical properties of the signal model at a noticeable level. So for the overcomplete gammatone signal model, the ERB scale itself is already close to the frequency tiling of the time-frequency plane that achieves the best frame-bound ratio.

For a decimated overcomplete gammatone signal model, the derived frame bounds cannot be used to optimize the decimation factors in dependency of the signal spectrum, as explained in Sec. IV. Another possibility to evaluate such bandlimited signal models is the computation of the SAR allowing the optimization of the trade-off between linear amplitude distortions and the amount of aliasing. We could show that the common approach to use decimation factors that are proportional to the bandwidth of the filters is suboptimal. The SAR can easily be computed using a two–dimensional fast Fourier transform (2D-FFT), and we therefore recommend for signal processing applications using a decimated overcomplete gammatone signal model to utilize decimation factors N_m being optimized for the applied signal class.

Note that very long finite-impulse responses and high sampling rates have been used in this study to derive frame bounds that are valid approximations for the analog gammatone filters. Applications using other digital realizations of the gammatone filterbank like infinite-impulse response filters might result in slightly different frame bounds.[48]

A linear gammatone signal model is a valid approximation of the human auditory filters for moderate sound pressure levels. It has been shown that the filter shape of the auditory filter changes with stimulus level,[49] which led to the development of dynamic, non-linear auditory filter models.[50,51] The analysis methods applied in this study cannot directly be applied to such dynamic filters and are therefore not within the scope of this manuscript.

VII. CONCLUSIONS

Using the theory of frames we could derive that from 2.4 filters per ERB on, a non-decimated overcomplete gammatone signal model achieves near-perfect signal reconstruction and that from $M=55$ (1.3 filters per ERB) filters on, a perceptual transparent audio coding is possible. We further showed that by computing a SAR, the decimation factors in multi-rate signal processing schemes can be optimized, balancing the amplitude and aliasing distortions. We showed for an audio test signal that hereby significant improvements can be achieved.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments and corrections, which significantly improved the quality of this manuscript. This work was partly funded by the German Science Foundation (DFG) through the International Graduate School for Neurosensory Science and Systems and the SFB/TRR 31: "The Active Auditory System." The author Stefan Strahl wants to especially thank Astrid Klinge for the inspiring scientific discussions about the manuscript.

APPENDIX A: MP WITH MATCHED FILTERS

MP (Ref. 19) assumes an additive signal model of the form

$$\mathbf{x}=\sum_{i=1}^{K} s_i\mathbf{a}_i, \tag{A1}$$

with the signal vector $\mathbf{x}\in\mathbb{R}^{N\times 1}$, the coefficients $\mathbf{s}=(s_1,s_2,\ldots,s_K)\in\mathbb{C}^K$, and the atoms $\mathbf{A}=(\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_M)\in\mathbb{C}^{N\times M}$ having unit-norm. For an overcomplete signal model, the MP algorithm searches for the sparsest encoding in the infinite number of possible encodings. As mentioned in the Introduction, this sparse signal model resembles the signal analysis performed by the human cochlea.

The algorithm performs a greedy iterative search by selecting at the ith iteration the atom having the largest inner product with the residual $\mathbf{r}_i$:

$$s_{m_i}=\arg\max_{\mathbf{a}_{m_i}\in\mathbf{A}}|\langle\mathbf{r}_i,\mathbf{a}_{m_i}\rangle|^2, \tag{A2}$$

with m_i being the dictionary index of the selected atom at the ith iteration. The new residual is then computed with

$$\mathbf{r}_{i+1}=\mathbf{r}_i-s_{m_i}\mathbf{a}_{m_i}. \tag{A3}$$

If we rewrite the inner products in Eq. (A2) as

$$s_m = \langle \mathbf{r}_i, \mathbf{a}_m \rangle = \sum_{n=1}^{N} r_i[n] \cdot a_m[n] = \sum_{n=1}^{N} r_i[n] \cdot \tilde{a}_m[N-n+1]$$

$$\text{with } \tilde{a}_m[n] = a_m^*[-n] = r_i[n] * \tilde{a}_m[n],$$

it can be seen that the inner products can also be computed using the time reversed atom $\tilde{a}_m$, which is also called a *matched filter*. So we can efficiently compute all inner products using a time-reversed gammatone filterbank. In practical applications of MP the support L of the atoms is often much smaller than the length N of the signal. Therefore most implementations[52–54] divide the signal into overlapping blocks of length L and stepwidth S. With this iterative procedure, only the correlations of the $2L/S-1$ signal blocks which have been altered in the previous iteration need to be recomputed. Using the matched-filter approach we can compute the new correlations of the $2L/S-1$ signal blocks in one step by convolving the $2L$ samples of the whole block once with the matched filterbank. So for a signal of length N and a dictionary size M, we can perform the MP iteration in $\mathcal{O}(MN)$. If MP is performed with a pure gammatone dictionary, we can accelerate the MP algorithm further by precomputing the representations of the gammatone atoms in the filterbank domain and performing the update of the inner products by a simple subtraction in the filterbank domain. For a dictionary of size M, instead of $6M \cdot 2L$ multiplication and $10M \cdot 2L$ additions,[55] the update of the correlations can be done with $M2L$ subtractions.

APPENDIX B: OPTIMAL DECIMATION FACTORS

In Sec. IV B derived optimal decimation factors for svega.wav, $b=1.019$, and $M=50$ are as follows:

$O=1$ $N_{1-10}=128$, $N_{11-33}=64$, $N_{34-49}=32$, $N_{50}=16$,

$O=2$ $N_{1-8}=64$, $N_{9-36}=32$, $N_{37-48}=16$, $N_{49-50}=8$,

$O=3$ $N_{1-24}=32$, $N_{25-40}=16$, $N_{41-50}=8$,

$O=4$ $N_{1-10}=32$, $N_{11-31}=16$, $N_{32-50}=8$,

$O=5$ $N_{1-2}=32$, $N_{3-31}=16$, $N_{32-44}=8$, $N_{45-50}=4$,

$O=6$ $N_{1-20}=16$, $N_{21-44}=8$, $N_{45-49}=4$, $N_{50}=2$,

$O=7$ $N_{1-14}=16$, $N_{15-39}=8$, $N_{40-49}=4$, $N_{50}=2$,

$O=8$ $N_{1-10}=16$, $N_{11-39}=8$, $N_{40-46}=4$, $N_{47-50}=2$.

[1] J. B. J. Fourier, *Théorie Analytique de la Chaleur (The Analytical Theory of Heat)* (Didot, Paris, 1822).
[2] D. Gabor, "Theory of communications," J. Inst. Electr. Eng. **93**, 429–457 (1946).
[3] J. Morlet, G. Arens, I. Fourgeau, and D. Giard, "Wave propagation and sampling theory," Geophysics **47**, 203–236 (1982).
[4] M. Lewicki, "Efficient coding of natural sounds," Nat. Neurosci. **5**, 356–363 (2002).
[5] E. Smith and M. Lewicki, "Efficient auditory coding," Nature (London) **439**, 978–982 (2006).
[6] S. Strahl and A. Mertins, "Sparse gammatone signal model optimized for English speech does not match the human auditory filters," Brain Res. **1220**, 224–233 (2008).
[7] R. Patterson and B. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. Moore (Academic, London, 1986), pp. 123–177.
[8] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE, December 14–15, 1987.
[9] T. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of the effective signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622 (1996).
[10] T. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of the effective signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631 (1996).
[11] R. Patterson, "Auditory images: How complex sounds are represented in the auditory system," Acoust. Sci. & Tech. **21**, 183–190 (2000).
[12] M. Cooke, "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573 (2006).
[13] G. Kubin and W. Kleijn, "Multiple-description coding (MDC) of speech with an invertibleauditory model," in Proceedings of the IEEE Workshop on Speech Coding (1999), pp. 81–83.
[14] G. Kubin and W. Kleijn, "On speech coding in a perceptual domain," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (1999), pp. 205–208.
[15] L. Lin, W. Holmes, and E. Ambikairajah, "Auditory filter bank inversion," in Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS) (2001), Vol. **2**, pp. 537–540.
[16] C. Feldbauer, G. Kubin, and W. Kleijn, "Anthropomorphic coding of speech and audio: A model inversion approach," EURASIP J. Appl. Signal Process. **9**, 1334–1349 (2005).
[17] V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," Acta. Acust. Acust. **88**, 433–442 (2002).
[18] T. Irino and M. Unoki, "A time-varying, analysis/synthesis auditory filterbank using the gammachirp," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (1998), Vol. **6**, pp. 3653–3656.
[19] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," IEEE Trans. Signal Process. **41**, 3397–3415 (1993).
[20] Z. Cvetkovic and M. Vetterli, "Overcomplete expansions and robustness," in Proceedings of the IEEE International Symposium on Time-Frequency and Time-Scale Analysis (1996), pp. 325–328.
[21] H. Bolcskei and F. Hlawatsch, "Oversampled filter banks: Optimal noise shaping, design freedom, and noise analysis," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (1997), Vol. **3**, pp. 2453–2456.
[22] R. Duffin and A. Schaeffer, "A class of nonharmonic Fourier series," Trans. Am. Math. Soc. **72**, 341–366 (1952).
[23] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," J. Math. Phys. **27**, 1271–1283 (1986).
[24] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE Trans. Inf. Theory **36**, 961–1005 (1990).
[25] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, PA, 1992).
[26] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
[27] H. Bolcskei, F. Hlawatsch, and H. Feichtinger, "Frame-theoretic analysis of oversampled filter banks," IEEE Trans. Signal Process. **46**, 3256–3268 (1998).
[28] J. Flanagan, "Models for approximating basilar membrane displacement," J. Acoust. Soc. Am. **32**, 937 (1960).
[29] P. I. Johannesma, "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in *Symposium on Hearing Theory* (Institute for Perception Research, Eindhoven, Holland, 1972), pp. 58–69.
[30] E. de Boer, "On the principle of specific coding," ASME J. Dyn. Syst., Meas., Control **95**, 265–273 (1973).
[31] A. M. H. J. Aertsen and P. I. M. Johannesma, "Spectro-temporal receptive fields of auditory neurons in the grassfrog," Biol. Cybern. **38**, 223–234 (1980).
[32] T. Irino, "An optimal auditory filter," in Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) (1995), pp. 198–201.
[33] B. Moore, R. Peters, and B. Glasberg, "Auditory filter shapes at low center frequencies," J. Acoust. Soc. Am. **88**, 132–140 (1990).
[34] B. Moore and B. Glasberg, "A revision of Zwicker's loudness model,"

Acta. Acust. Acust. **82**, 335–345 (1996).
[35]A. Bell and T. Sejnowski, "Learning the higher order structure of a natural sound," Network Comput. Neural Syst. **7**, 261–266 (1996).
[36]E. Ambikairajah, J. Epps, and L. Lin, "Wideband speech and audio coding using gammatone filter banks," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2001), pp. 773–776.
[37]ISO, ISO 389-7, Acoustics-reference zero for the calibration of audiometric equipment—Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions, International Organization for Standardization, Geneva (1996).
[38]P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Upper Saddle River, NJ, 1993).
[39]L. Zadeh, "Frequency analysis of variable networks," Proc. IRE **38**, 291–299 (1950).
[40]C. Loeffler and C. Burrus, "Optimal design of periodically time-varying and multirate digital filters," IEEE Trans. Acoust., Speech, Signal Process. **32**, 991–997 (1984).
[41]P. Hoang and P. Vaidyanathan, "Non-uniform multirate filter banks: Theory and design," in Proceedings of the IEEE International Symposium on Circuits and Systems (1989), pp. 371–374.
[42]I. Djokovic and P. Vaidyanathan, "Results on biorthogonal filter banks," Appl. Comput. Harmon. Anal. **1**, 329–343 (1994).
[43]B. Edler and G. Schuller, "Audio coding using a psychoacoustic pre- and post-filter," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2000), Vol. **2**, pp. 881–884.
[44]ISO/MPEG, "MPEG-4 Audio Version 2 ISO/IEC 14496-3:1999/Amd.1" (1999).
[45]R. Huber and B. Kollmeier, "PEMO-Q: A new method for objective audio quality assessment using a model of auditory perception," IEEE Trans. Audio, Speech, Lang. Process. **14**, 1902–1911 (2006).
[46]ITU-R Recommendation BS.1387-1, "Methods for objective measurements of perceived audio quality," International Telecommunication Union, Geneva (2001).
[47]B. C. J. Moore, *Cochlear Hearing Loss* (Wiley-Interscience, Malden, MA, 1998).
[48]L. Van Immerseel and S. Peeters, "Digital implementation of linear gammatone filters: Comparison of design methods," ARLO **4**, 59–64 (2003).
[49]S. Rosen and R. J. Baker, "Characterising auditory filter nonlinearity," Hear. Res. **73**, 231–243 (1994).
[50]T. Irino and R. Patterson, "A dynamic, compressive gammachirp auditory filterbank," IEEE Trans. Audio, Speech, Lang. Process. **14**, 2222–2232 (2006).
[51]E. Lopez-Poveda and R. Meddis, "A human nonlinear cochlear filterbank," J. Acoust. Soc. Am. **110**, 3107–3118 (2001).
[52]S. Mallat and Z. Zhang, "The matching pursuit software package (mpp)," ftp://cs.nyu.edu/pub/wave/software/mpp.tar.Z (Last viewed 4/23/2009).
[53]S. E. Ferrando, L. A. Kolasa, and N. Kovačević, "Algorithm 820: A flexible implementation of matching pursuit for gabor functions on the interval," ACM Trans. Math. Softw. **28**, 337–353 (2002).
[54]R. Gribonval and S. Krstulovic, "MPTK, The matching pursuit toolkit," http://mptk.gforge.inria.fr/ (Last viewed 4/23/2009).
[55]T. Herzke and V. Hohmann, "Improved numerical methods for gammatone filterbank analysis and synthesis," Acta. Acust. Acust. **93**, 498–500 (2007).
[56]M. Brucke, W. Nebel, A. Schwarz, B. Mertsching, M. Hansen, and B. Kollmeier, "Silicon cochlea: A digital VLSI implementation of a quantitative model of the auditory system," J. Acoust. Soc. Am. **105**, 1192 (1999).
[57]N. Ma, P. Green, and A. Coy, "Exploiting dendritic autocorrelogram structure to identify spectro-temporal regions dominated by a single sound source," Speech Commun. **49**, 874–891 (2007).