A TIME-VARYING, ANALYSIS/SYNTHESIS AUDITORY FILTERBANK USING THE GAMMACHIRP

Toshio Irino[*] *and Masashi Unoki*[*+]

[*] ATR Human Information Processing Research Labs.
2-2 Hikaridai Seika-cho Soraku-gun Kyoto, 619-02, JAPAN
[+] Japan Advanced Institute of Science and Technology
1-1 Asahidai Tatsunokuchi Nomi Ishikawa, 923-12, JAPAN

ABSTRACT

A time-varying, analysis/synthesis auditory filterbank has been developed using a new implementation of the "gammachirp", which has been shown to be an excellent function for the asymmetric, level-dependent auditory filter. The gammachirp filter is shown to be implemented through a combination of a gammatone filter and an IIR asymmetric compensation filter; which largely reduces the computational cost for time-varying filtering. The gammachirp filterbank is designed using a linear gammatone filterbank and a bank of time-varying asymmetric compensation filters controlled by the sound pressure level estimated at the output of the filterbank. Since the inverse filter of the asymmetric compensation filter is always stable, it is possible to resynthesize signals from time-varying, level-dependent auditory representations. The resynthesis error is only determined by the linear analysis/synthesis gammatone filterbank. The proposed filterbank is applicable to various types of signal processing required to model human auditory filtering.

1 . INTRODUCTION

Extensive efforts have been made to introduce characteristics of the human auditory system into signal processing carried out in telecommunications systems, as exemplified by audio coding. A number of auditory models have been proposed for simulation of the peripheral auditory system [1], but none of them have been successfully used as much as the linear predictive analysis and the Fourier transform. One of the major reasons for this may be that no signal resynthesis procedure is provided with any realistic time-varying auditory model. Although linear auditory filterbanks or wavelets have often been used for signal resynthesis [2,3,4], they cannot account for the dynamic characteristics of basilar membrane motion. To solve this problem, this paper presents a new time-varying, level-dependent, analysis/synthesis auditory filterbank based on the "gammachirp" [5,6].

The gammachirp was analytically derived as a function satisfying minimal uncertainty in joint time-scale representations [7,5]. The gammachirp auditory filter is an extension of the popular gammatone filter [8]; it has an additional frequency-modulation term to produce an asymmetric amplitude spectrum. When the degree of asymmetry is associated with stimulus level, the gammachirp filter can provide an excellent fit to 12 sets of notched-noise masking data from three different studies [6]. The gammachirp is much simpler than non-parametric impulse responses of recent physiological models on cochlear mechanics [1], which have not provided a good fit to human masking data. Nevertheless, the frequency-modulation in the gammachirp is consistent with recent physiological observations of frequency-modulations in mechanical responses of the basilar membrane [9].

There is, however, a problem in the processing speed. Since the gammachirp filter is defined with a time-domain function, it has been implemented as a finite impulse response (FIR) filter. For simulation of dynamic characteristics of cochlea, the filter coefficients have to be recalculated and then multiplied to signal on a moment-to-moment basis. The large number of FIR coefficients, especially at low frequencies, precludes fast time-varying filtering.

The problem would be solved if the gammachirp filter could be implemented with a sufficiently small number of filter coefficients. Therefore, an efficient implementation of the gammachirp filter has been developed as shown in section 2. This implementation divides the gammachirp filter into two cascaded filters: a linear gammatone filter and a time-varying, level-dependent IIR filter, i.e., an asymmetric compensation filter [10]. Section 2 also shows the stability of the inverse filter of this IIR filter that enables signal resynthesis. Section 3 shows an implementation of the gammachirp filterbank controlled by the sound pressure level estimated at the output of the filterbank, and also shows a total analysis/ synthesis procedure.

2 . GAMMACHIRP IMPLEMENTATION

2.1 Definition of the gammachirp

The complex impulse response of the gammachirp is

$$g_c(t) = a t^{n-1} \exp\left(-2\pi b \mathrm{ERB}(f_r)\ t\right) \exp\left(j2\pi f_r t + jc \ln t + j\phi\right) \quad (1)$$

where time $t > 0$, a is the amplitude, n and b are parameters defining the distribution, f_r is the asymptotic frequency, c is the parameter for the frequency modulation, and ϕ is the initial phase; $\ln t$ is a natural logarithm of time, $\mathrm{ERB}(f_r)$ is the equivalent rectangular bandwidth of the filter at f_r, and at moderate levels, $\mathrm{ERB}(f_r) = 24.7 + 0.108 f_r$ in Hz. When $c = 0$, this equation represents a complex impulse response of the gammatone.

Results on fitting to notched-noise masking data have shown that parameter c is a level-dependent parameter whereas n and b are invariant parameters [6]. Because the argument in this paper is valid for any parameter value, a set of typical values is used throughout this paper, i.e., $n = 4$, $b = 1.68$, and

$$c = 3.38 - 0.107 \cdot P_s \quad (2)$$

where P_s is the sound pressure level (dB) of a probe tone at 2000 Hz [6].

2.2 Amplitude spectrum of the gammachirp

The amplitude spectrum of the gammachirp in Eq. (1) [4] is

$$|G_c(f)| = \frac{|a\Gamma(n+jc)|}{|2\pi b\mathrm{ERB}(f_r) + j2\pi(f-f_r)|^n} \cdot e^{c\theta}, \quad (3)$$

$$\theta = \arctan\{(f-f_r)/b\mathrm{ERB}(f_r)\}. \quad (4)$$

The first term in Eq. (3) represents the amplitude spectrum of the gammatone since $e^{c\theta}=1$ when $c=0$. The peak frequency is obtained as

$$f_{peak} = f_r + c \cdot b\mathrm{ERB}(f_r)/n. \quad (5)$$

Thus, the term $e^{c\theta}$ produces a shift in the peak frequency according to Eq. (4) and introduces asymmetry into the amplitude spectrum. When the amplitude of Eq. (2) is normalized, Eq. (2) is rewritten as

$$|G_C(f)| = |G_T(f)| \cdot |H_A(f)|, \quad (6)$$

where

$$|H_A(f)| = e^{c\theta} = \exp[c \cdot \arctan\{(f-f_r)/b\mathrm{ERB}(f_r)\}], \quad (7)$$

and $|G_T(f)|$ is the amplitude spectrum of the gammatone, which is level-independent and invariant since n and b are constant. The gammachirp is represented by two cascaded filters: an invariant gammatone filter and an asymmetric, level-dependent filter. Thus, the gammachirp could be implemented for fast processing if a filter corresponding to $|H_A(f)|$ were designed with a few parameters in a reasonable accuracy, since an efficient implementation of the gammatone is already known [11].

Amplitude characteristics of $|H_A(f)|$ are shown with the solid lines in Fig. 1 when f_r is 2000 Hz, as an example. These summary are the necessary conditions for designing approximation filters.

(a) $|H_A(f)|$ is an all-pass filter when $c=0$, a high-pass filter when $c>0$, and a low-pass filter when $c<0$. The slope and the range of amplitude increase when the absolute value of c increases.

(b) For an arbitrary frequency f_a, the characteristics follow

$$|H_A(f_r+f_a)| = |H_A(f_r-f_a)|^{-1}. \quad (8)$$

(c) $|H_A(f)|$ changes monotonically. Neither a peak nor a dip exists.

2.3 Asymmetric compensation filter

A filter satisfying condition (b) is considered first. FIR filters cannot satisfy Eq. (8) in the strict sense since they only have zeros. It is possible to simulate $|H_A(f)|$ for a reasonable accuracy with a linear-phase FIR filter designed using the Remez algorithm, but this is not effective since the number of coefficients is comparable to that of the original FIR gammachirp and, moreover, the coefficients seem to require a table indexed with parameters b, c, and f_r. Well-known IIR Butterworth and Chebyshev filters also cannot satisfy Eq. (8). IIR filters can satisfy Eq. (8) only when the numbers of poles and zeros are equal and the poles and zeros are symmetrically located at $f_r+\Delta f$ and $f_r-\Delta f$ for a proper frequency Δf. The

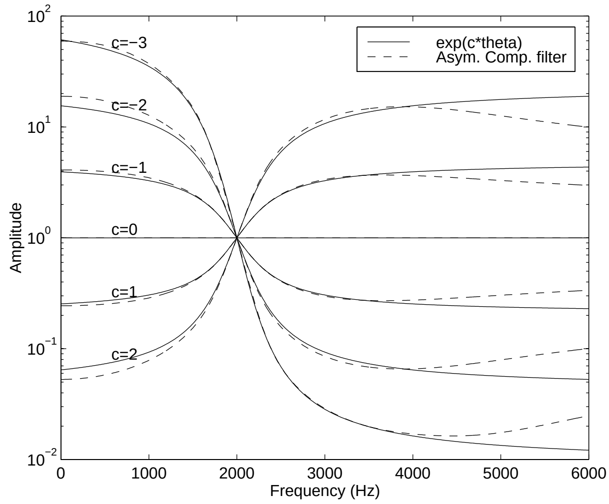

Figure 1. Amplitude spectra of $|H_A(f)|$ (solid lines) and an asymmetric compensation filter $|H_C(f)|$ (dashed lines), where $n=4$, $b=1.68$, $f_r=2000$ Hz, and the amplitude is normalized by the value when $f=f_r$.

amplitudes r of the corresponding poles and zeros are equal and are less than unity for the convergence of IIR filters. Since the bandwidth gets narrower when r gets closer to unity, r might be negatively correlated to the bandwidth parameter $b\mathrm{ERB}(f_r)$. Condition (a) implies that Δf is proportional to c and is positively correlated to $b\mathrm{ERB}(f_r)$. A cascaded second-order digital filter satisfying these properties is defined as

$$H_C(z) = \prod_k H_{Ck}(z) \quad (9)$$

$$H_{Ck}(z) = \frac{(1-r_k e^{j\varphi_k} z^{-1})(1-r_k e^{-j\varphi_k} z^{-1})}{(1-r_k e^{j\phi_k} z^{-1})(1-r_k e^{-j\phi_k} z^{-1})} \quad (10)$$

$$r_k = \exp\{-k \cdot p_1 \cdot 2\pi b\mathrm{ERB}(f_r)/f_s\} \quad (11)$$

$$\phi_k = 2\pi\{f_r + 2^{k-1} \cdot p_2 \cdot c \cdot b\mathrm{ERB}(f_r)\}/f_s \quad (12)$$

$$\varphi_k = 2\pi\{f_r - 2^{k-1} \cdot p_2 \cdot c \cdot b\mathrm{ERB}(f_r)\}/f_s \quad (13)$$

where p_1 and p_2 are positive coefficients and f_s is the sampling rate. The reason for cascading filters with gradually located poles and zeros is to satisfy condition (c). This filter is referred to as an "asymmetric compensation (AC)" filter.

The amplitude spectrum $|H_C(f)|$ is shown by the dashed lines in Fig. 1; four second-order filters were cascaded and the amplitude was normalized at $f=f_r$. The dashed lines are very close to the solid lines when the frequency is less than 4000 Hz. Above 4000 Hz, the disparity gets larger. Although further fitting is possible by increasing the number of cascaded filters, the trade-off between the number of coefficients and the degree of fitting should be considered. Results from the next subsection show that four cascaded second-order filters are sufficient. Then, the numbers of poles and zeros are 16 in total.

2.4 Approximation error

The gammachirp filter is shown to be approximated with the asymmetric compensation filter cascaded to the gammatone filter. The amplitude spectrum of this filter is, by replacing $|H_A(f)|$ with $|H_C(f)|$ in Eq. (6),

$$|G_{CAC}(f)| = |G_T(f)| \cdot |H_C(f)|. \tag{14}$$

This filter $G_{CAC}(f)$ is referred to as an "Asymmetric Compensation - gammachirp" or "AC-gammachirp" in this subsection to be distinguishable from the original gammachirp defined by Eq. (1). Figure 2 shows amplitude spectra of the gammachirp $|G_C(f)|$ in Eq. (6) (solid lines), the AC-gammachirp $|G_{CAC}(f)|$ in Eq. (14) (dashed lines), and the gammatone $|G_T(f)|$ (dotted lines). To improve the fitness, the amplitude $|G_{CAC}(f)|$ was normalized for each second-order filter. The normalizing frequency is closely related to the peak shift in Eq. (5) and for the *k*-th filter,

$$f = f_r + k \cdot p_3 \cdot c \cdot b\mathrm{ERB}(f_r)/n. \tag{15}$$

The coefficients p_1, p_2, and p_3 are set heuristically as:

$$p_1 = 1.35 - 0.19 \cdot |c|, \tag{16}$$

$$p_2 = 0.29 - 0.0040 \cdot |c|, \tag{17}$$

$$p_3 = 0.23 + 0.0072 \cdot |c|. \tag{18}$$

In Fig. 2, the rms (root-mean-squared) error for each filter pair in the range of $|G_C(f)| > -50$ dB is less than 0.41 dB (maximum when $f_r = 500$ Hz). For 90 sets of parameter combinations { n=4; b= 1.0, 1.35, and 1.7; c= 1.0, 0,-1.0, -2.0, and -3.0; f_r = 250, 500, 1000, 2000, 4000, and 8000 (Hz)} , the average rms error is 0.63 dB. The rms error exceeds 2 dB only for three sets when f_r= 8000 Hz and c= -3.0. The rms errors improve little when the coefficients in Eqs. (16), (17), and (18) are optimized using a least squared-error method. It may be possible to improve the fitness further by changing the locations of the poles and zeros defined in Eqs. (11), (12), and (13), but this is beyond the scope of this paper.

The difference in the impulse responses of the original gammachirp and the AC-gammachirp is about -48 dB in rms amplitude, i.e., the responses are almost identical, when the initial phase is chosen properly. The phase spectra have been found to be very close to each other [10]. Consequently, the AC-gammachirp is able to excellently approximate the original gammachirp. Therefore,

$$G_C(f) \cong G_{CAC}(f) = G_T(f) \cdot H_C(f). \tag{19}$$

It is possible to perform fast time-varying, level-dependent auditory filtering since the gammachirp can be implemented with a fast gammatone filter [11] and the IIR filter with 16 coefficients.

2.5 Stability of the inverse filter

The asymmetric compensation filter $H_C(z)$ defined in Eqs. (9) and (10) can be represented as

$$H_C(z) = B(z)/A(z). \tag{20}$$

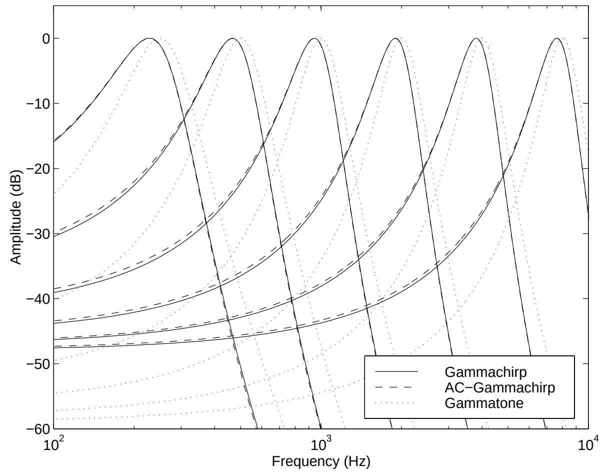

Figure 2. Amplitude spectra of the original gammachirp filter (solid lines), an asymmetric compensation (AC) gammachirp filter (dashed lines), and a gammatone filter (dotted lines), where $n = 4$, $b = 1.68$, $c = -1$, and f_r are 250, 500, 1000, 2000, 4000, and 8000 Hz.

The inverse filter of this filter is always derived simply as

$$H_C^{-1}(z) = A(z)/B(z). \tag{21}$$

This IIR filter is also stable because the amplitudes of all of the poles and zeros are less than unity as defined in Eq. (11). This fact guarantees that the output of the time-invariant gammatone filter can be recovered from the output of the time-varying asymmetric compensation filter, with the precision of numerical errors. This result leads to a discussion of signal resynthesis in the next section.

3. GAMMACHIRP FILTERBANK

3.1 Implementation with level estimation

Figure 3 shows a block diagram of the gammachirp filterbank. The gammachirp filterbank can be implemented using a gammatone filterbank and a bank of asymmetric compensation filters. The sound pressure level is estimated at the output of the asymmetric compensation filterbank to determine the asymmetric parameter *c*. This kind of feedback control has been introduced in many auditory models [1]. The parameter controller consists of a bank of parameter control units as shown in the right-bottom block. When considering the *k*-th channel, the input signal to this block is rectified, and then, put into a leaky integrator (LI) for smoothing. A weighting function is applied to the LI output of the *k*-th and adjacent channels and they are summed together to obtain the average activity a_{sk} for the *k*-th channel. The estimated sound pressure level P_{sk} is calculated from a straight forward equation,

$$\hat{P}_{sk} = 20 \log_{10}(q \cdot a_{sk}) \tag{22}$$

where q is a constant coefficient. The estimated asymmetric parameter $\hat{c}_k$ for the *k*-th channel of the asymmetric compensation filterbank is then calculated by using Eq. (2).

The value of q can be determined using the data on the masking threshold and the stimulus signals from the notched-noise masking experiments. See [10] for more details. Excitation patterns calculated from this filterbank and roex filters have also been presented in [10].

3.2 Analysis/synthesis filterbank

One of the most important features of the gammachirp filterbank is to establish an analysis/synthesis system as shown in Fig. 4. The output of the linear gammatone filterbank can be recovered from the output of the time-varying, level-dependent gammachirp filterbank using a bank of inverse compensation filters when the control parameters are identical in both of the forward and inverse filterbank. Then, signal resynthesis is possible by using a bank of time-reversal gammatone filters and weighted summation as is usually done in wavelet analysis/synthesis [2]. The error between the original and synthetic signals is determined only by the linear analysis/synthesis gammatone filterbank because the combination of the compensation filter and its inverse filter is a unit impulse when their control parameters are the same at any sampling point. Although Fig. 4 shows the direct analysis/synthesis filterbank, it is also possible to modify time-varying representations at the output of the gammachirp filterbank to resynthesize modified signals.

4. SUMMARY

The gammachirp filter is shown to be excellently approximated by the combination of a gammatone filter and an IIR asymmetric compensation filter. The new implementation reduces the computational cost for time-varying filtering because both of the filters can be implemented with a few filter coefficients. The inverse filter of the asymmetric compensation filter is shown to be stable. The gammachirp filterbank is designed using a time-invariant, linear gammatone filterbank and a bank of time-varying, asymmetric compensation filters controlled by the estimated sound pressure level at the output of the filterbank. This filterbank is shown to provide a procedure for signal resynthesis with a guaranteed precision. The time-varying, analysis/synthesis gammachirp filterbank is a new step beyond conventional auditory models and is applicable to various signal processing tasks.

ACKNOWLEDGMENTS

The authors wish to thank Roy D. Patterson of MRC-APU, UK for his continuous advice, and Minoru Tsuzaki and Hani Yehia of ATR-HIP for their valuable comments. A part of this work was performed while the second author was a visiting student at ATR-HIP. The authors also wish to thank Masato Akagi of JAIST, and Yoh'ichi Tohkura, Hideki Kawahara and Shigeru Katagiri of ATR-HIP for the arrangements.

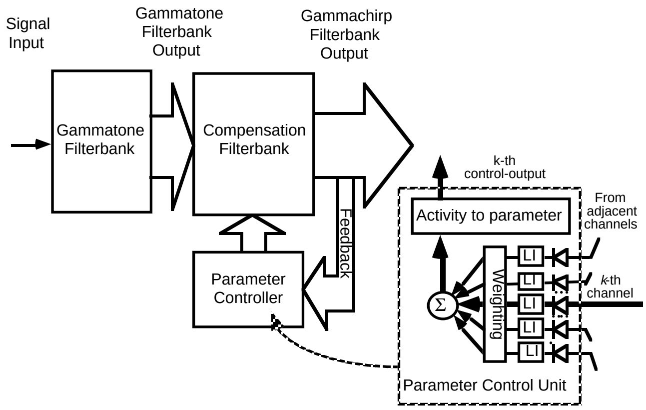

Figure 3. Block diagram of the gammachirp filterbank.

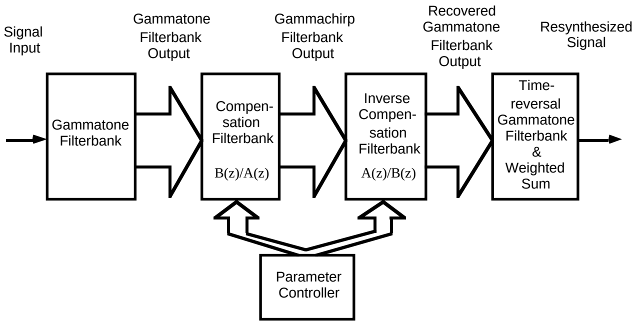

Figure 4. Block diagram of the time-varying, level-dependent analysis/synthesis gammachirp filterbank.

REFERENCES

[1] Giguere, C. and Woodland, P. C. ,"A computational model of the auditory periphery for speech and hearing research. I. Ascending path","--,II Descending path," J. Acoust. Soc. Am., 95, 331-342, 343-349, 1994 and references herein.

[2] Combes, J. M., Grossmann, A. and Tchamitchian, Ph. Eds. , "Wavelets," Springer-Verlag, Berlin, 1989.

[3] Yang, X, Wang, K. and Shamma, S. A., "Auditory representations of acoustic signals," IEEE Trans. Information Theory, 38, pp. 824-839, 1992..

[4] Irino, T. and Kawahara, H., "Signal reconstruction from modified auditory wavelet transform," IEEE Trans. Signal Processing, 41, pp. 3549-3554, 1993.

[5] Irino, T., "An optimal auditory filter," in IEEE Signal Processing Society, 1995 Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY., 1995. See also Proc. ICASSP96, 2, pp. 981-984, 1996 and http://www.hip.atr.co.jp/~irino/.

[6] Irino, T. and Patterson, R.D., "A time-domain, level-dependent auditory filter: The gammachirp", J. Acoust. Soc. Am. 101, pp. 412-419, 1997.

[7] Cohen, L., "The scale transform," IEEE Trans. Signal Processing, 41, pp. 3275-3292, 1993.

[8] Patterson, R. D., Allerhand, M. and Giguere, C., "Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform," J. Acoust. Soc. Am., 98, 1890-1894, 1995.

[9] de Boer, E. and Nuttall, A. L., "The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions," J. Acoust. Soc. Am., 101, 3583-3592, 1997.

[10] Irino, T. and Unoki, M., "An efficient implementation of the gammachirp filter and its filterbank design," ATR Technical Report, TR-H-225, July 14, 1997. See also Trans. Tech, Com. Psycho. Physio. Acoust. of Acoust. Soc. Japan, H-97-69, Oct. 24, 1997.

[11] Slaney, M., "An efficient implementation of the Patterson-Holdsworth auditory filter bank," Apple Computer Technical Report #35, 1993 and references herein.