

TR-H-225

**An Efficient Implementation of the
Gammachirp Filter and its Filterbank Design.**

**Toshio IRINO
Masashi UNOKI (ATR-HIP/JAIST)**

1997.7.14

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 TEL: 0774-95-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

Telephone: +81-774-95-1011

Fax : +81-774-95-1008

An efficient implementation of the gammachirp filter and its filterbank design

Toshio IRINO (ATR Human Information Processing Res. Labs.)
Masashi UNOKI (ATR HIP/JAIST)

irino@hip.atr.co.jp
http://www.hip.atr.co.jp/~irino

Abstract

A new implementation of a gammachirp filter is proposed to reduce computational costs that mainly preclude the conventional FIR gammachirp filter from time-variant filtering. The new gammachirp filter consists of a gammatone filter and an IIR asymmetric compensation filter with a few parameters. The gammachirp filterbank is designed to introduce time-variant, level-dependent characteristics into an auditory filterbank. This filterbank is applicable to various types of signal processing necessary to model characteristics of the human peripheral auditory filter.

1. Introduction

Recently, the "gammachirp" filter (Irino, 1995, 1996) was proposed as an auditory filter in time-domain auditory models. The gammachirp filter is an extension of the popular gammatone filter (e.g. Patterson, R. D., Allerhand, M., and Giguere, C., 1995); 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 is found to provide an excellent fit to 12 sets of notched-noise masking data from three different studies (Irino and Patterson (1997)). It has a well-defined impulse response whereas the roex auditory filter (Patterson et. al. 1982), conventionally used for fitting masking data, does not. The gammachirp is much simpler than a filter response obtained from current physiological models of cochlear mechanics (e.g. Giguere, C. and Woodland, P. C. (1994)) which have not provided a good fit to human masking data. The frequency-modulation in the gammachirp is consistent with recent physiological observations of frequency-modulations in mechanical waveforms of the basilar membrane (de Boer and Nuttall (1997)). Moreover, from a theoretical signal processing point of

view, the gammachirp can analytically be derived as a function satisfying "minimal uncertainty" in joint time-scale representations (Cohen, 1993). With these merits, the gammachirp is an excellent candidate for an asymmetric, level-dependent auditory filterbank in time-domain models of auditory processing.

There is, however, a problem in designing a filterbank with the gammachirp filter, involving the processing speed. The gammachirp filter has been implemented with a finite impulse response (FIR) filter since it is defined with an impulse response and its Laplace transform is not simple enough for direct implementation with an infinite impulse response (IIR) filter. The number of coefficients is large for calculation especially at low frequencies; the FIR gammachirp filterbank involves a computationally heavy process even when constant filter shape asymmetry. In addition, a bank of time-variant filters is required for simulation of realistic auditory filtering. Let us consider signal processing with a time-variant FIR filter. First, at a certain sample point, the impulse response is calculated in correspondence to the signal level. Second, the impulse response is convoluted with the previous signal to obtain the value at the next sample point. These steps are duplicated for the number of filters in the filterbank structure. It is almost impossible to perform this procedure in a reasonable duration even with recent fast computers. At the same time, this procedure introduces differences from realistic auditory processing since the characteristics of the filter continuously vary and the impulse response is never calculated at each time slice. These problems could be resolved if an approximation of the gammachirp filter were developed with sufficiently small number of filter coefficients.

One potential solution is the use of an all-pole gammatone filter (APGF) (Slaney, 1993) or an one-zero gammatone filter (OZGF) (Lyon, 1996) both of which are efficient IIR implementations of the gammatone filter to introduce asymmetric amplitude spectra. They are similar to but still different from the gammachirp. The APGF only has poles. Since the locations of the poles of the APGF are control parameters for the filter asymmetry as well as the peak frequency, the filter asymmetry are dependent on the sampling rate f_s in filter signal processing. For example, when four pairs of poles in a fourth-order APGF locate at the same place and when the peak frequency $f_c = f_s/4$, the amplitude spectrum is completely symmetric relative to the peak frequency since all poles locate at $\pm\pi/2$ in the unit circle. The skirt of the filter spectrum below the peak frequency broadens when $f_c < f_s/4$ and the skirt of the filter above the peak frequency broadens when $f_c > f_s/4$. When the sampling rate is 16 kHz, a frequency commonly used in speech analysis, $f_s/4$ is 4000 Hz. Although it would be possible to avoid this problem by increasing the sampling rate, re-locating the poles properly, and using a one-zero

gammatone filter, a sufficient solution is still unclear with a few reports (Pflueger, Hoeldrich and Reidler, 1997).

This report provides an efficient implementation of the gammachirp filter. The implementation is simply controllable, reasonably accurate, and independent of the sampling rate and the peak frequency. The next section shows a new concept for achieving an approximation of the gammachirp and an IIR filter design for this purpose. Here, sound level estimation is essential for controlling the filter shape. Section 3 shows a procedure for level estimation and an implementation of the gammachirp filterbank that uses it. Since the proposed gammachirp filterbank consists of a gammatone filterbank and a IIR filterbank with few coefficients, high-speed processing is possible and thus, it is applicable to various types of signal processing which requires a spectrum analysis capable of simulating human auditory peripheral processing.

2. Asymmetric Compensation Gammachirp

2.1 Definition of the gammachirp

The complex impulse response of the gammatone is

$$g_t(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_c) t) \exp(j2\pi f_c t + \phi) \quad (1)$$

where time $t > 0$, a is the amplitude, n and b are parameters defining the distribution, f_c is the carrier frequency, ϕ is the initial phase, $\text{ERB}(f_c)$ is the equivalent rectangular bandwidth of the filter, and at moderate levels, $\text{ERB}(f_c) = 24.7 + 0.108 f_c$ in Hz (Glasberg and Moore, 1990).

The complex impulse response of the gammachirp is

$$g_c(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_r) t) \exp(j2\pi f_r t + jc \ln t + \phi) \quad (2)$$

where time $t > 0$, f_r is the asymptotic frequency, and c is a parameter for the frequency-modulation term $\ln t$ which is the natural logarithm of time. When $c = 0$, Eq. (2) is identical to Eq. (1). The result of fitting to masking data is that c is a level-dependent parameter whereas n and b are invariant parameters (Irino and Patterson, 1997).

2.2 Amplitude spectrum of the gammachirp

The amplitude spectrum of the gammachirp in Eq. (2) (Irino and Patterson, 1997) is

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

$$\theta = \arctan \frac{f - f_r}{b \text{ERB}(f_r)} \quad (4)$$

When $c = 0$, Eq. (3) is identical to the amplitude spectrum of the gammatone in Eq. (1) since $e^{c\theta} = 1$. The peak frequency is obtained as

$$f_{peak} = f_r + \frac{c \cdot b \text{ERB}(f_r)}{n} \quad (5)$$

The term $e^{c\theta}$ produces a shift in the peak frequency as Eq. (5) shows and introduces asymmetry in the filter shape. When the amplitude is normalized, Eq. (3) is rewritten as

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

$$|H_A(f)| = e^{c\theta} = \exp\left\{c \cdot \arctan \frac{f - f_r}{b \text{ERB}(f_r)}\right\}, \quad (7)$$

where $|G_T(f)|$ is the amplitude spectrum of the gammatone. Thus, the gammachirp consists of two cascaded filters: the invariant gammatone filter and the asymmetric, level-dependent filter. The gammachirp can be simulated if a filter corresponding to $|H_A(f)|$ is designed with a few parameters, in a reasonable accuracy since an efficient implementation of the gammatone is already known (e.g. Slaney, 1994). Although it is possible to simulate $|H_A(f)|$ with a linear-phase FIR filter when using Remez algorithm (McClellan, J.H., Parks, T.W. and Rabiner, L.R., 1979), this is not effective since the number of coefficients for reasonable accuracy is comparable to the original impulse response and, moreover, these coefficients seem to be inconsistently dependent on parameters $b \cdot c \cdot f_r$. We need to define a filter in which the coefficients are explicit functions of the parameters $b \cdot c \cdot f_r$, through more detailed observations of the characteristics of $|H_A(f)|$.

The amplitude characteristics of $|H_A(f)|$ are shown with the solid lines in Fig. 1 when f_r is 2000 Hz. The characteristics are the necessary conditions for designing approximation filters.

(a) When $c = 0$, the amplitude is unity, i.e. all-pass. When c is positive, the characteristics are high-pass. When c is negative, they are low-pass. 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

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

(c) They change monotonically and there is no peak nor dip

2.3 Implementation of an asymmetric compensation filter

One capable of approximation is an "asymmetric compensation (AC)" filter. First, we consider whether the filter satisfies condition (b). FIR filters cannot satisfy Eq. (8) in the strict sense since they only have zeros. Well-known IIR Butterworth and Chebyshev filters cannot satisfy the equation either. IIR filters having equal numbers of poles and

zeros can satisfy Eq. (8) but only when the poles and zeros are symmetrically located at $f_r + \Delta f$ and $f_r - \Delta f$ for a proper frequency Δf . The absolute values r of the corresponding poles and zeros are equal and are less than unity for the convergence condition of IIR filters. Since bandwidth gets narrower when r gets closer to unity, r might be negatively correlated to the bandwidth parameter $b\text{ERB}(f_r)$. Condition (a) implies that Δf is proportional to c and is positively correlated to $b\text{ERB}(f_r)$. We define a second-order digital filter satisfying these properties as

$$H_{C0}(z) = \frac{(1 - re^{j\phi} z^{-1})(1 - re^{-j\phi} z^{-1})}{(1 - re^{j\phi} z^{-1})(1 - re^{-j\phi} z^{-1})} \quad (9)$$

$$r = \exp\{-p_1 \cdot 2\pi b\text{ERB}(f_r) / f_s\} \quad (10)$$

$$\phi = 2\pi\{f_r + p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (11)$$

$$\varphi = 2\pi\{f_r - p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (12)$$

where p_1 and p_2 are positive coefficients and f_s is the sampling rate. The amplitude spectrum $|H_{C0}(f)|$ satisfies conditions (a) and (b). It does not, however, satisfy condition (c). When r is close to unity, there is a peak around $f_r + p_2 \cdot c \cdot b\text{ERB}(f_r)$ and there is a dip around $f_r - p_2 \cdot c \cdot b\text{ERB}(f_r)$. When r is small, the slope of $|H_{C0}(f)|$ around frequency f_r is not sufficient for the approximation of $|H_A(f)|$. To avoid this problem, several second-order filters with gradually located poles and zeros are cascaded. Then, an asymmetric compensation filter $H_C(z)$ is defined as

$$H_C(z) = \prod_k H_{Ck}(z) \quad (13)$$

$$H_{Ck}(z) = \frac{(1 - r_k e^{j\phi_k} z^{-1})(1 - r_k e^{-j\phi_k} z^{-1})}{(1 - r_k e^{j\phi_k} z^{-1})(1 - r_k e^{-j\phi_k} z^{-1})} \quad (14)$$

$$r_k = \exp\{-k \cdot p_1 \cdot 2\pi b\text{ERB}(f_r) / f_s\} \quad (15)$$

$$\phi_k = 2\pi\{f_r + 2^{k-1} \cdot p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (16)$$

$$\varphi_k = 2\pi\{f_r - 2^{k-1} \cdot p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (17)$$

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. Compensation is possible by increasing the number of cascaded filters. Another purpose of this report, however, is to reduce the number of filter coefficients. Thus, four seems reasonable to be a number as a result of filter evaluation described in the following subsections. Then, the numbers of poles and zeros are 16 in total.

2.4 Approximation error in the amplitude spectrum

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

$$|G_{AC}(f)| = |G_T(f)| \cdot |H_C(f)| \quad (18)$$

This filter is referred to as the "Asymmetric Compensation - gammachirp" or "AC-gammachirp" hereafter to be distinguishable from the original gammachirp defined by Eq. (2). Figure 2 shows the amplitude spectra of the gammachirp $|G_c(f)|$ in Eq. (6) (solid lines), the AC-gammachirp $|G_{AC}(f)|$ in Eq. (18) (dashed lines), and the gammatone $|G_T(f)|$ (dotted line). To improve the fitness, the amplitude $|G_{AC}(f)|$ is 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,

$$f = f_r + k \cdot p_3 \cdot c \cdot b \text{ERB}(f_r) / n \quad (19)$$

The coefficients p_1 , p_2 , and p_3 are set as

$$p_1 = 1.35 - 0.19 \cdot |c|, \quad (20)$$

$$p_2 = 0.29 - 0.0040 \cdot |c|, \quad (21)$$

$$p_3 = 0.23 + 0.0072 \cdot |c|, \quad (22)$$

where $|c|$ is the absolute value of c . 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 $\{ n = 4; b = 1.0, 1.35, \text{ and } 1.7; c = 1.0, 0, -1.0, -2.0, \text{ and } -3.0; f_r = 250, 500, 1000, 2000, 4000, \text{ and } 8000 \text{ (Hz)} \}$, the average rms error is 0.63 dB. The rms error exceeds 2 dB only for three sets when $f_r = 8000$ Hz and $c = -3.0$. As a result, the AC-gammachirp excellently approximates the gammachirp in the amplitude spectrum. The rms errors improve little when the coefficients in Eqs. (20), (21), and (22) are optimized using a least squared-error method. It would be possible to improve the fitness further by changing the locations of poles and zeros defined in Eqs. (15), (16), and (17), but this is beyond the scope of this report.

2.5 Impulse response and phase spectrum

Figure 3(a) shows impulse responses of the AC-gammachirp (dashed line), the original gammachirp (solid line), and the gammatone (dotted line) which is used as the first stage of the AC-gammachirp filter. The AC-gammachirp and the original gammachirp are almost identical when the initial phase is set properly. The rms error between them is 0.0038 (= -48 dB). The initial phase for good fitness correlates to the value of the asymmetry parameter c . This does not affect the filtered output if the variability of c is sufficiently small during the length of the impulse response; this is the case with the

current model as described in the next section.

Figure 3(b) shows phase spectra of the AC-gammachirp (dashed line), the original gammachirp (solid line), and the gammatone (dotted line), where the phase is set to zero when $f = f_r$. The phase characteristics of the AC-gammachirp and the gammachirp are very close except at low frequencies.

2.6 Fast gammachirp filter

From the previous sections, it can be said that the AC-gammachirp excellently approximates the original gammachirp. For time-variant systems, the AC-gammachirp is advantageous as described in the Introduction. The AC-gammachirp filter consists of the linear gammatone filter and the asymmetric compensation filter. Since the asymmetric compensation filter is a four-stage second-order IIR filter and enables a sufficiently fast implementation, it is easy to implement a fast gammachirp filter using a fast gammatone filter in AIM (Patterson, R. D., Allerhand, M., and Giguere, C., 1995) or an IIR gammatone filter (Slaney (1993)).

3. Gammachirp filterbank

This section describes a design of the gammachirp filterbank using the AC-gammachirp filter.

3.1 Design concept

The auditory filter shape depends on the sound pressure level (Lufti and Patterson, 1984; Patterson and Moore, 1966; Moore and Glasberg, 1987). The sound level should be estimated at a certain stage in any auditory filter model. When using a level-dependent roex filter, there are two hypotheses for the level estimation: at the input using a linear Fourier transform (Moore and Glasberg, 1990) and at the output of the level-dependent model (Rosen and Baker, 1994, see also Moore, 1995). The Fourier transform is not taken prior to the auditory filter. When the level estimation is performed at the output of the level-dependent model, there is a problem determining the initial degree of asymmetry. It is possible to use an iterative convergence algorithm started from a certain amount of asymmetry. However, this is computationally ineffective and requires consideration of the convergence conditions. Thus, a feed-forward controller is preferable to a feed-back controller in the model.

Since the proposed AC-gammachirp filter consists of two stages, the sound level is estimated at the output of the fixed gammatone filter to control the time-variant asymmetric compensation filter. However, because the peak frequency of the gammachirp depends on the value of c as shown in Eq. (5), the output value of one gammatone filter does not

correctly reflect the estimated sound level for the desired gammachirp filter. The level estimation should be performed using the output values from several adjacent filters. This implementation is intermediate between the input and output hypotheses.

3.2 Signal processing

We describe an implementation of the AC-gammachirp filterbank (evaluated in the following subsections). The set of parameters used below is obtained from results of fitting to psychological filter shape data (Rosen and Baker, 1994) with the gammachirp (Irino and Patterson, 1997). This set can easily be replaced by another set of values obtained from another gammachirp fitting.

Figure 4 shows a block diagram of the gammachirp filterbank. A signal is first processed by an IIR gammatone filterbank proposed by Slaney (1993) for convenience in programming with MATLAB. The filtered outputs are distributed to the asymmetric compensation filterbank and the parameter estimation block. The input signal to this block is rectified, and then, put into a leaky integrator (LI) with a time constant of 30 ms. A weighting function is applied to the LI output and they are summed together to obtain the average activity for the channel (indicated by the thick line). The weighting function covers about a 3-ERB range around this channel. To introduce asymmetry in this stage, it is the shape of the amplitude spectrum of the gammachirp function when $n = 4$, $b = 1.68$, and $c = -1$. The estimated sound pressure level \hat{P}_s is calculated from the average activity

$$a_a \quad \hat{P}_s = 20 \log_{10}(q \cdot a_a) \quad (23)$$

where q is a constant coefficient. The estimated asymmetric parameter \hat{c} is then calculated by the equation (Irino and Patterson, 1997):

$$\hat{c} = 3.38 - 0.107 \hat{P}_s, \quad (24)$$

under the condition of the notched noise experiments (Rosen and Baker, 1994). The value of \hat{c} is used for the asymmetric compensation filter (the thick line on the final block). The output of this filter depends on asymmetric parameter \hat{c} . When \hat{c} is fixed, this filter approximates the original gammachirp filter with a fixed asymmetry as described in the previous section.

3.3 Evaluation of the filterbank

The parameters of the gammachirp filter above were determined using a power spectrum model of masking (Fletcher, 1940) applied to notched noise masking data for a 2000 Hz probe tone (Rosen and Baker, 1994). There were 78 stimulus conditions, i.e. combinations of the probe level, the noise level, and the noise bandwidth. The error

between the stimulus probe level P_s and the estimated probe level \hat{P}_s could be estimated when the stimuli were processed by the gammachirp filterbank. The stimulus sound was filtered in advance with an ELC correction filter implemented by an FIR linear phase filter. For each stimulus, the auditory filter shape could be determined only at one channel with a maximum signal-to-noise ratio, i.e. an assumption with "off-frequency listening" (Patterson, R. D. and Nimmo-Smith, I., 1980).

The off-frequency can be calculated independently of this gammachirp filterbank using a part of the fitting procedure when a pair of a stimulus and fitted parameters is given. The output activity of the gammatone channel at which the peak frequency is closest to the off-frequency is assumed to determine the shape of the gammachirp filter. The activity is calculated by averaging sufficiently stable responses and is used to calculate the estimated sound level \hat{P}_s with Eq. (23). The rms error between P_s and \hat{P}_s was minimized to about 1.9 dB when $q = 14.5$. This value is slightly larger than the rms error obtained when fitting the gammachirp to the notched noise masking data (1.3 dB). The shape of the weighting function and the rms error between the IIR gammatone filter and the original gammatone contribute toward a part of this difference. The rms error in c is only 0.2 when converting a P_s error using Eq. (24) and does not affect the filter shape much. The rms error between the AC-gammachirp filter and the original gammachirp filter is about 0.63 dB (see subsection 2.4) and is also negligible. As a consequence, this filterbank design is reasonable for level estimation and for filter shape estimation at least under notched noise conditions.

3.4 Comparison in the excitation pattern

The filterbank was also evaluated by comparing excitation patterns. Figure 5(a) shows the excitation pattern of a pure tone at 2000 Hz calculated from the procedure proposed by Moore and Glasberg (1987). Figure 5(b) shows the excitation pattern calculated from the output activity of the gammachirp filterbank when $b = 1.019$ and c was limited between -3.5 and 1. Although the peak location seems to shift in accordance with the level increment in this figure, this appearance is because the abscissa is fixed to f_r of the gammachirp in Eq. (2), whereas the peak frequency of the gammachirp in the corresponding channel is shifted with a value of c . The excitation level above the peak frequency in Fig. 5(b) shows a characteristic corresponding to the "upward spread of masking" (Moore and Glasberg, 1987) as well as in Fig. 5(a). The excitation level below the peak falls around 1700 Hz and then spreads again at a low level of less than 30 dB.

This phenomena can be explained using Fig. 5(c) which shows the average estimated \hat{c} value across frequencies. The estimated level \hat{P}_s is maximized and thus \hat{c} is minimized

at 2000 Hz because of using the output of the fixed gammatone filter and using Eqs. (23) and (24). The skirt of the filter shape below 2000 Hz broadens maximally. The value of \hat{c} gradually increases when channel frequency f_r increases above 2000 Hz. The skirt sharpens gradually but still passes the activity concentrated around 2000 Hz. When channel frequency f_r is just below 2000 Hz, the skirt above f_r is sharpened because of the negative \hat{c} . The filter largely reduces the activity. When f_r becomes slightly apart from 2000 Hz, \hat{c} becomes positive and the skirt above f_r broadens. Then, the filter passes the activity around 2000 Hz. Similar excitation pattern is observed even c is limited to negative. As a consequence, this phenomenon occurs when Eqs. (23) and (24) are applied to every gammachirp filter. It is necessary to modify the relationship if this type of excitation pattern is not observed in realistic basilar membrane motion.

4. Conclusion

The gammachirp filter has been found to be excellently approximated by the combination of the gammatone filter and the asymmetric compensation filter. Since the gammachirp filter has conventionally been implemented only by an FIR filter whereas the asymmetric compensation filter has been implemented by an IIR filter, it has become possible to significantly reduce computational costs when combining a fast implementation of the gammatone filter (e.g. Patterson, R. D., Allerhand, M., and Giguere, C., 1995; Slaney, 1993).

The gammachirp filterbank is designed to introduce time-variant, level-dependent asymmetry into the auditory filterbank. This is applicable to various types of signal processing necessary to model the characteristics of the human auditory filter. The parameters of the filter shape are set to the same values across frequencies in this simulation and the parameters for cochlear dynamics, e.g. the time constant of the leaky integrator, continue to remain as temporal values. It would be possible to get these values after future measurements. The gammachirp filterbank is advantageous to current physiological models of cochlear mechanics in its simplicity, inherent from the original gammachirp defined with a few parameters.

Acknowledgments

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

of ATR-HIP for valuable comments.

References

- Allen, J. B., Hall, J. L., Hubbard, A., Neely, T. S., and Tubis, A. (Eds.). (1985). "Peripheral Auditory Mechanisms," (Springer-Verlag, Berlin).
- Carney, L. H. and Yin, T. C. T. (1988). "Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model," *J. Neurophysiology*, 60, 1653-1677.
- de Boer, E. and Nuttall, A.L. (1997). "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.
- Fletcher, H. (1940). "Auditory patterns," *Rev. Mod. Phys.*, 12, 47-65.
- Giguere, C. and Woodland, P. C. (1994). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," *J. Acoust. Soc. Am.*, 95, 331-342.
- Glasberg, B. R. and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, 47, 103-138.
- Irino, T. (1995). "An optimal auditory filter," in *IEEE Signal Processing Society, 1995 Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.
- Irino, T. (1996). "A 'gammachirp' function as an optimal auditory filter with the Mellin transform," in *IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-96)*, 2, 981-984, Atlanta, GA.
- Irino, T. and Patterson, R.D. (1997). "A time-domain level-dependent auditory filter: The gammachirp", *J. Acoust. Soc. Am.* 101, pp. 412-419.
- Johannesma, P. I. M. (1972). "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in *Symposium on Hearing Theory*, 58-69, IPO, Eindhoven, Holland.
- Lyon, R. F. (1996). "The all-pole gammatone filter and auditory models," in *Forum Acusticum '96*, Antwerp, Belgium.
- McClellan, J.H., Parks, T.W. and Rabiner, L.R. (1979). "FIR linear phase filter design program" in "Programs for Digital Signal Processing" Ed. IEEE, IEEE press, ch. 5.1.
- Moore, B. C. J. and Glasberg, B. R. (1987). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," *Hear. Res.*, 28, 209-225.

- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1990). "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.*, 88, 132-140.
- Moore, B.C.J. (1995). "Frequency analysis and masking" in *Hearing*, edited by B. C. J. Moore (Academic Press, London).
- Patterson, R. D., Allerhand, M., and Giguere, C. (1995). "Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform," *J. Acoust. Soc. Am.*, 98, 1890-1894.
- Patterson, R. D. and Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic Press, London).
- Patterson, R. D. and Nimmo-Smith, I. (1980). "Off-frequency listening and auditory filter asymmetry," *J. Acoust. Soc. Am.*, 67, 229-245.
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.*, 72, 1788-1803.
- Pflueger, M., Hoeldrich, R. and Reidler, W. (1997), Personal communications.
- Rosen, S. and Baker, R. J. (1993). "Characterising auditory filter nonlinearity," *Hear. Res.*, 73, 231-243.
- Slaney, M. (1993). "An efficient implementation of the Patterson-Holdsworth auditory filter bank," *Apple Computer Technical Report #35*.

Figure captions

Fig. 1 Amplitude spectra of $e^{c\theta}$ (solid lines) and an asymmetric compensation filters $|H_c(f)|$ (dashed lines), where $n=4$, $b=1.68$, $f_r=2000$ Hz and the amplitude is normalized when $f=f_r$.

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

Fig. 3 (a) Impulse responses and (b) phase spectra of the original gammachirp filter (solid line), an AC-gammachirp filter (dashed line), and a gammatone filter (dotted line). The base-line is shifted for 0.1 each in (a). The parameters are: $n=4$, $b=1.68$, $c=-1$, and $f_r=2000$ Hz.

Fig. 4 Block diagram of a gammachirp filterbank.

Fig. 5 Excitation patterns of a 2000 Hz sinusoid derived from (a) Moore and Glasberg (1987) and (b) the output of a gammachirp filterbank when $b=1.019$. (c) Estimated parameter c in the gammachirp filterbank. The abscissa in (b) and (c) is the resonance frequency f_r of the gammachirp filter in Eq. (2).

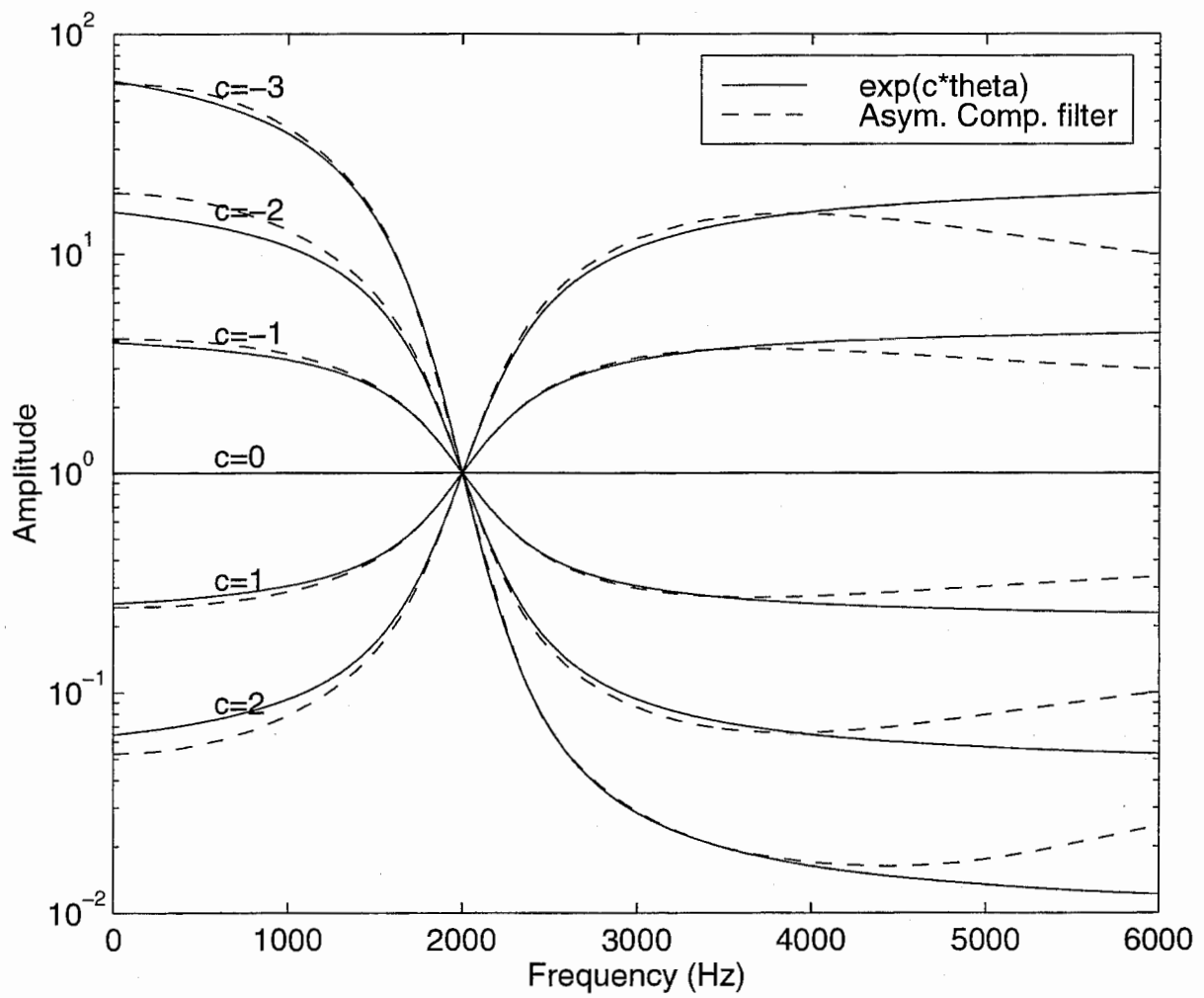


Fig. 1

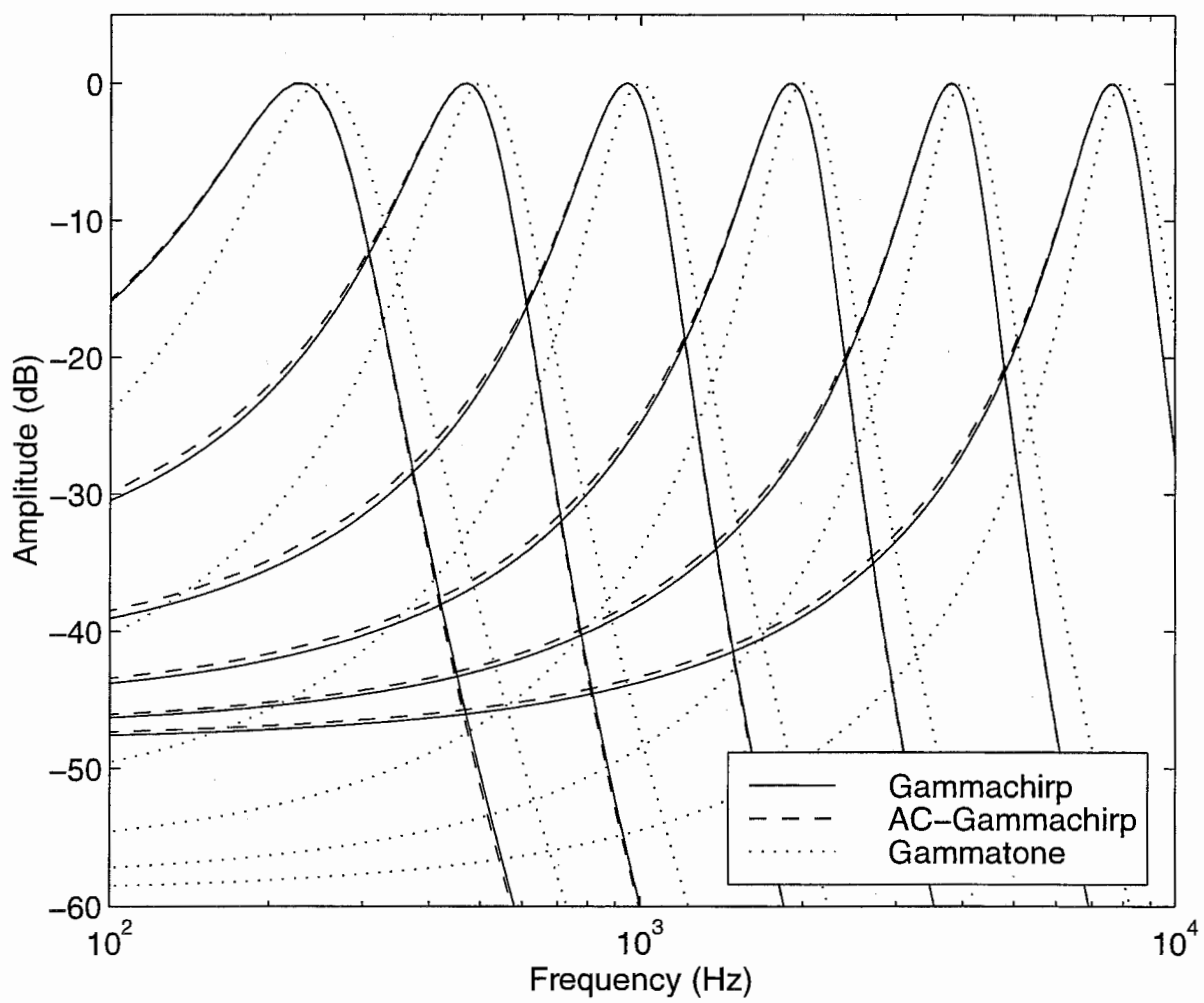


Fig. 2

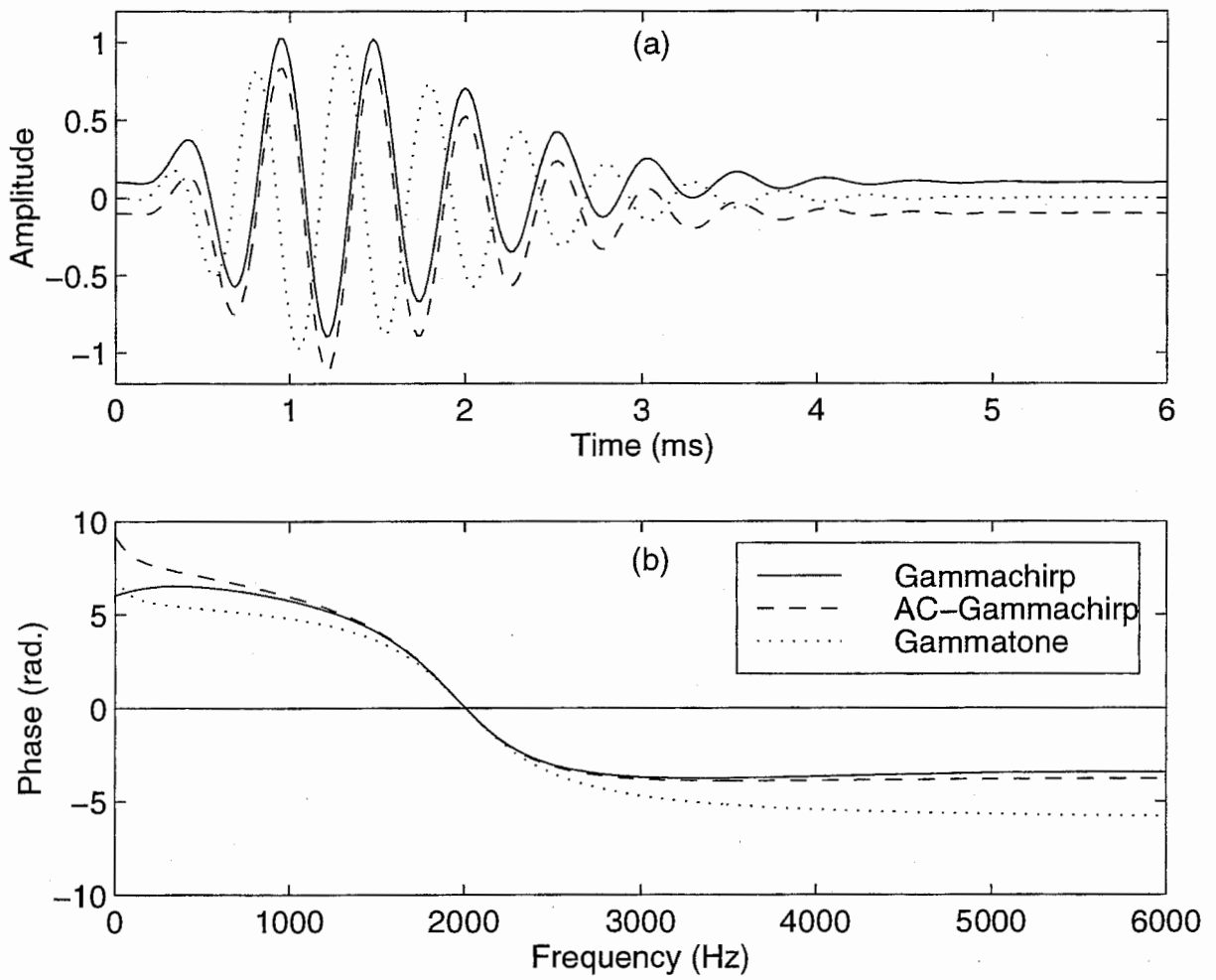


Fig. 3

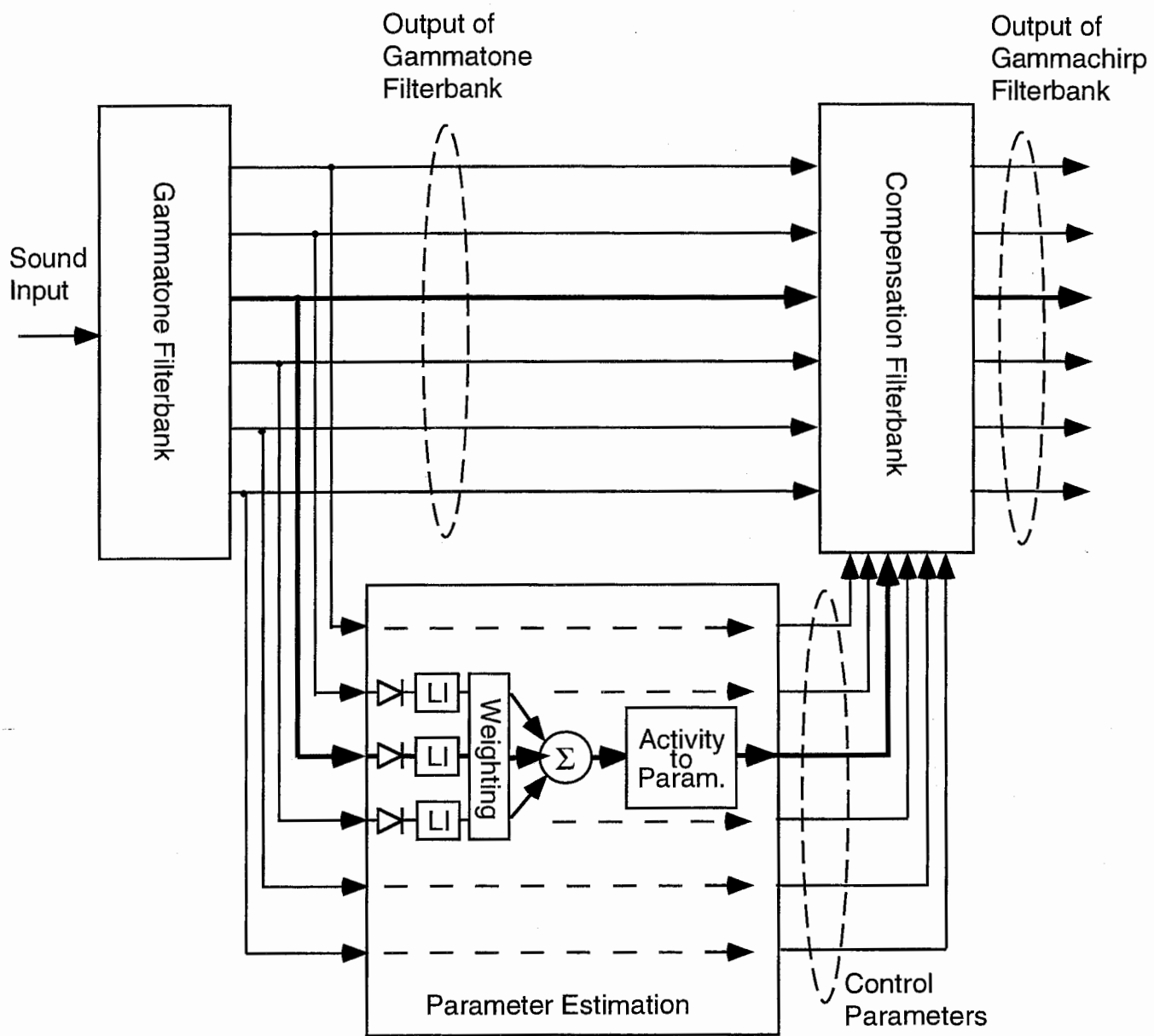


Fig. 4

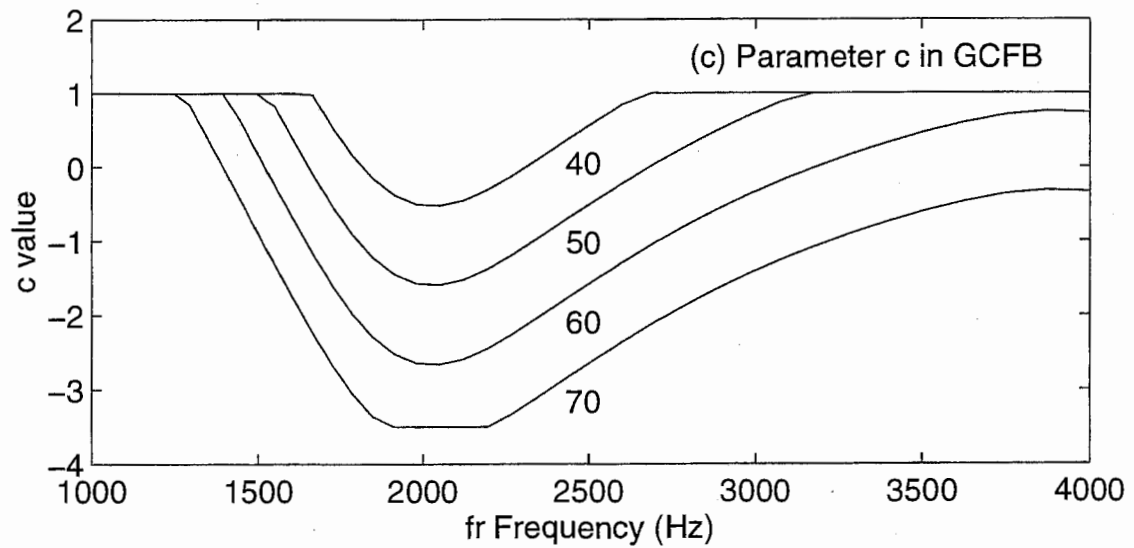
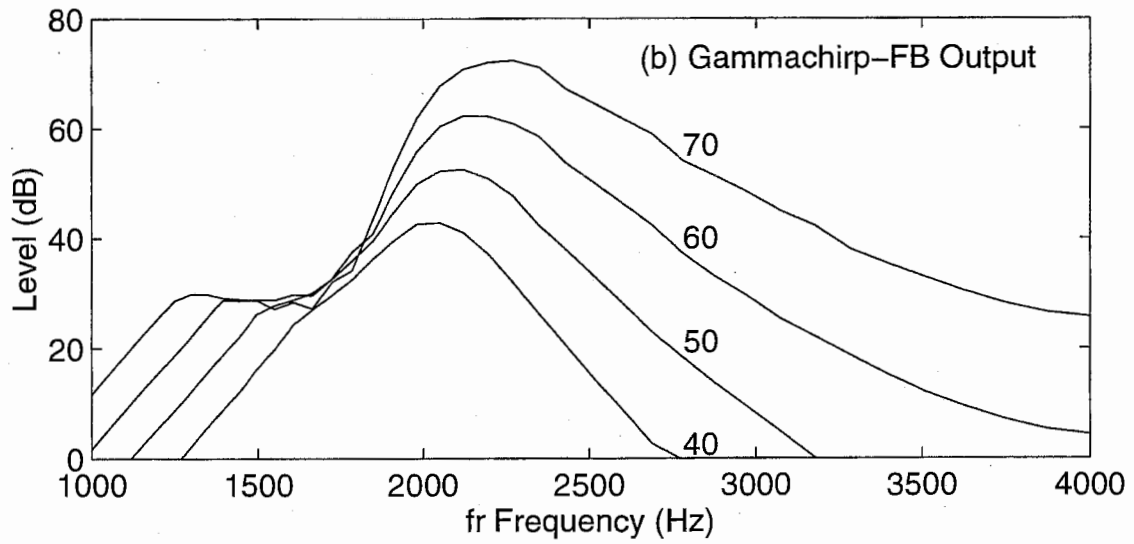
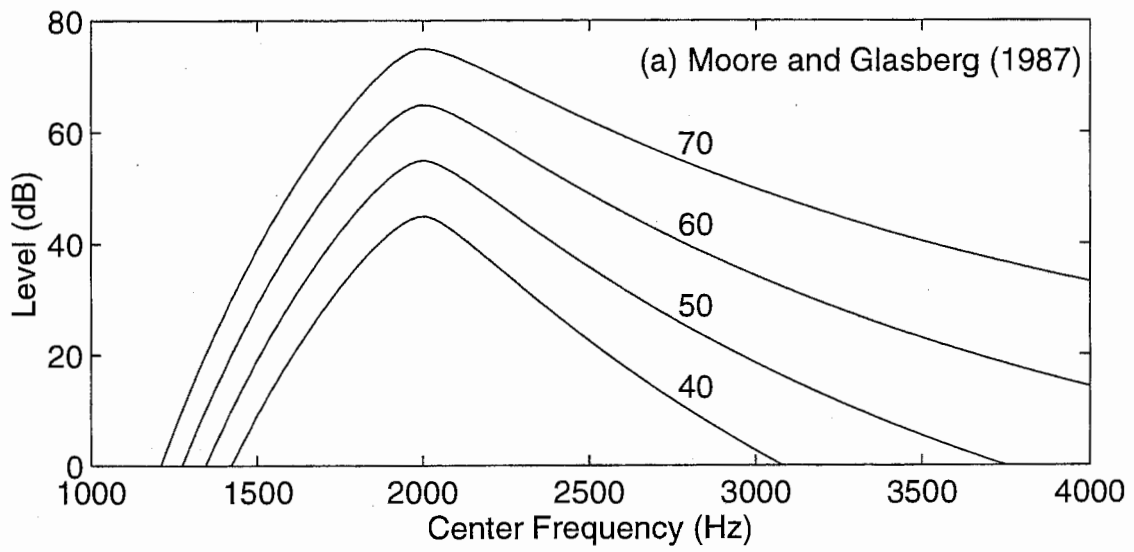


Fig. 5

ガンマチャープフィルタとフィルタバンクの効率的な構成
An efficient implementation of the gammachirp filter
and its filterbank design

入野俊夫 (ATR 人間情報通信研究所)

鷗木祐史 (ATR 人間情報通信研究所 / 北陸先端科学技術大学院大学)

概要

本報告では、ガンマトーンフィルタと IIR フィルタの非対称性補償フィルタを組み合わせるにより、従来 FIR フィルタでしか構成できなかったガンマチャープフィルタを精度良く実現できることを示した。ガンマトーンフィルタに高速演算可能なものを用いれば、全体での演算時間をきわめて短くすることができる。また、時変系の聴覚フィルタバンクを模擬するため、非対称性補償フィルタフィルタバンクの特性をガンマトーンフィルタバンクの出力レベルで制御する形のガンマチャープフィルタバンクの構成法も示した。時変系の聴覚フィルタバンクで、パラメータ数が少なく制御が比較的簡単で演算時間が短いのは利点である。前段階処理として人間の聴覚特性を考慮した周波数分析が必要となるさまざまな信号処理に応用可能である。

1. はじめに

最近、聴覚フィルタモデルとして広く使われている線形のガンマトーンフィルタ (Patterson et.al, 1996) に、周波数変調項を加えたガンマチャープフィルタが提案された (Irino, 1995, 1996; 入野 1995)。このフィルタのインパルス応答を表わす「ガンマチャープ関数」は、時間スケール表現における最小不確定性の意味で最適な関数として理論的に導出されたものである。このフィルタは、roex フィルタ (Patterson et.al. 1982) と異なり、時間応答で定義されていてかつ周波数領域での非対称性をもっている。この非対称性の度合を信号レベルに依存させることにより、線形で対称なガンマトーンフィルタやパラメータ数の多い基底膜振動解析モデル (たとえば Allen et al (1985) や Giguere, C. and Woodland, P. C. (1994)) では無理であった、心理物理的なデータに適合できるようになった。これについて、3つの異なる研究機関の 12 組のノッチ雑音法によるマスキングデータに良く適合させられることを報告した (入野・パターソン(1996); Irino and Patterson (1997))。また、基底膜振動での周波数変調が近い最近の計測でわかって

きた(de Boer and Nuttall (1997))が、これと整合性がある。

このようにガンマチャープフィルタは優れた特性を持っているが、時間波形で与えられておりその周波数応答が単純な形でなく、有限長インパルス応答(FIR)形でしか実現できていなかった。フィルタバンク処理でフィルタ数がある程度多くなると時間がかかる。非対称性の度合が時間によって変化しないのであれば、特性には問題は生じない。しかし、実際の聴覚末梢系の動特性は時々刻々変化しているであろう。このような信号分析を FIR フィルタで近似させようとする、まず、音圧レベルに合った非対称性を持つインパルス応答を求め、これをさらにそれまでの入力波形と掛け合わせて積分をすることによりその時点での値を得る、という手順を各サンプル点ごと信号の長さ分だけ行う必要があり非常に時間がかかる。ある程度変化がゆるやかであれば更新のサンプル点間隔を広げることができるが、このようにあるサンプル点ごとにインパルス応答を先に計算することは明らかに聴覚末梢系では行われていない。そこで、このような FIR 形のままでは本来の動特性を完全には反映しきれないおそれが生ずる。もし、ガンマチャープフィルタを無限インパルス応答(IIR)フィルタを用いて少ないパラメータで近似実現することができれば、これらの問題点を解消できる。

一つの方法として、ガンマチャープと少し異なるが、IIR フィルタでガンマトーンフィルタに周波数非対称性を与えたものに、全極形ガンマトーンフィルタ(APGF, All-pole gammatone filter)や一零形ガンマトーンフィルタ(OZGF, One-zero gammatone filter)が提案されている(Slaney, 1993; Lyon, 1996)。APGFは極のみを持つIIRフィルタで、その極の位置だけによって非対称性の度合を制御するので、サンプリング周波数 f_s への依存性がある。この4次の極のペアの位置がすべて等しい場合、ピーク周波数 $f_c = f_s/4$ の時、極が $\pm\pi/2$ の位置に置かれて、振幅周波数特性は左右対称になる。それ以下の周波数では低周波数側のすそ引きがなだらかで、それ以上の周波数では高周波数側のすそ引きがなだらかになる。たとえば、音声分析でよく使われる f_s が16 kHzの場合、 f_c が4 kHzでこの周波数となる。サンプリング周波数を十分高くしたり、極の再配置をすることやOZGFを使えばある程度克服も可能かもしれないが、制御パラメータの設定が難しそう、部分的な議論しか報告されていない(Pflueger, Hoeldrich and Reidler, 1997)。

これに対し、本報告では、サンプリング周波数や設計する周波数に依存せず、基本的にフィルタの帯域幅を決める2つのパラメータと非対称性を決める1つのパラメータのみで特性を特定できるガンマチャープフィルタ自体を近似度高く実現する方法を提案する。第2節では、IIRフィルタを用いてガンマチャープフィルタを近似実現する方法を述べる。また、第3節ではフィルタバンクを構

成し、ガンマチャープフィルタの振幅周波数特性の非対称性を音圧レベル依存させるための方略も示す。このフィルタバンクは、計算量が少なくすむので、実時間に近い速度での処理が期待できる。人間の聴覚特性を考慮した周波数分析が前段階処理として必要となるさまざまな信号処理への応用が期待できる。

2. 非対称性補償形ガンマチャープフィルタ

2.1 ガンマチャープの定義

まず、ガンマトーンフィルタの複素表示のインパルス応答は

$$g_c(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_c) t) \exp(j2\pi f_c t + \phi), \quad (1)$$

で与えられる。ここで時間 $t > 0$ 、 a は振幅、 n と b はガンマ分布関数のパラメータで f_c は搬送波周波数、 ϕ は初期位相、 $\text{ERB}(f_c)$ は等価矩形帯域幅で $\text{ERB}(f_c) = 24.7 + 0.108f_c$ (Hz) で表わされる (Glasberg and Moore, 1990)。

これに対し、ガンマチャープフィルタの複素表示のインパルス応答は、

$$g_c(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_r) t) \exp(j2\pi f_r t + jc \ln t + \phi). \quad (2)$$

ここで時間 $t > 0$ 、 f_r は最終的に収束する周波数、 c は周波数変調項の係数で非対称性の度合を決めるパラメータで $\ln t$ は自然対数を表わす。ガンマトーン関数との違いは $jc \ln t$ だけで、明らかに $c = 0$ の時は、ガンマトーンと一致する。このフィルタは周波数の分数の変化率を持つ FM 搬送波 (チャープ信号) とガンマ分布関数の包絡線を持つことからガンマチャープと名付けられた。等価矩形帯域幅は信号レベルによって変化するが、簡単のためにガンマトーンの時と同様、周波数 f_r だけの関数とした。マスキングデータへの適合の結果によると、 c がレベル依存するパラメータで、 n と b は不変パラメータである (入野・パターソン、1996)

2.2 ガンマチャープフィルタの周波数特性

式(2)のガンマチャープフィルタのフーリエスペクトルの振幅は以下のように求められる (導出は入野・パターソン(1996)付録参照)。

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

$$\theta = \arctan \frac{f-f_r}{b \text{ERB}(f_r)} \quad (4)$$

$c = 0$ の時はガンマトーンと一致する。ここで、式(3)を周波数で微分し零を取る周波数を求めるとガンマチャープ関数のピークの周波数が決定できて、

$$f_{\text{peak}} = f_r + \frac{c \cdot b \text{ERB}(f_r)}{n} \quad (5)$$

と非常に簡単になる。式(3)の分数で表わされている部分が、ガンマトーンの特

性であり、残りの $e^{c\theta}$ の項が、式(5)のようにピーク周波数を変化させて非対称性を決定している。すなわち式(3)は、振幅を正規化するとして、

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

$$|H_A(f)| = e^{c\theta} = \exp\left\{c \cdot \arctan \frac{f - f_r}{b \text{ERB}(f_r)}\right\} \quad (7)$$

という形に書き換えることができる。ここで $|G_T(f)|$ はガンマトーンの周波数振幅特性で、ガンマトーンフィルタに非対称性のフィルタ $H_A(f)$ を縦続接続にした形である。この振幅特性 $|H_A(f)|$ を少ないフィルタ係数で近似的に実現することによって、ガンマチャープの振幅周波数特性を近似することを考える。ガンマトーンフィルタのIIR形での実現方法はすでに確立しているので、これによりガンマチャープフィルタを高速演算できる形で近似実現できることになる。ここで、Remezのアルゴリズム(McClellan, J.H., Parks, T.W. and Rabiner, L.R., 1979)を用いれば、FIR形の直線位相で $|H_A(f)|$ を割合精度高く実現することが出来るが、係数の数が多くなり、また、 $b \cdot c \cdot f_r$ の値ごとに係数を計算するかテーブルとして持つ必要が出て現実的ではない。そこで、特性をもっと詳しく見て、 $b \cdot c \cdot f_r$ の関数としてのフィルタ係数が与えられる形で実現を考える必要がある。

$|H_A(f)|$ の特性を第1図の実線で示す。ここで f_r は2000 Hzである。特徴として以下のことが言えて、そのまま近似フィルタが満たすべき条件となる。

- (a) $c=0$ の時、振幅が全周波数で1となるのに対し、 c が正の場合高域通過形、 c が負の場合低域通過形で、 c の絶対値が大きくなるほど振幅の範囲が広がり f_r での傾きが大きくなる。
- (b) 任意の周波数 f_a に対し以下の関係が成立する。

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

- (c) 変化は単調で、極大値・極小値がない。

2.3 非対称性補償フィルタの実現

$|H_A(f)|$ を近似するフィルタを非対称性補償フィルタ呼ぶことにし、まず、条件(b)を満たすことを考える。FIRフィルタは零点しか持たないので、式(8)を厳密には満たせない。IIRフィルタで比較的良く使われるバターワース形やチェビシェフ形フィルタでも同様に満たせない。IIRフィルタで、極と零点の数が等しく、それらの配置が周波数 f_r に対して対称の適当な Δf に対し $f_r + \Delta f$ と $f_r - \Delta f$ に相当する位置に置かれていれば満たすことができる。これから、極・零点の大きさ r は等しくなくてはならず、かつ発散しないために1より小さくなくてはならない。一般に r は1に近づくほど共振の Q は大きくなるので、係数がかかった

帯域幅 $b\text{ERB}(f_r)$ と負の相関を持てばよい。 Δf は、条件(a)を考慮にいれると非対称性の c に比例し、帯域幅 $b\text{ERB}(f_r)$ にも相関がありそうである。上記のような特性を、デジタルフィルタで直接実現することを考える。一般的に行われるように極・零点を周波数の正負で対称に配置するとし、2次のIIRデジタルフィルタを次のように定義する。

$$H_{c0}(z) = \frac{(1 - re^{j\phi}z^{-1})(1 - re^{-j\phi}z^{-1})}{(1 - re^{j\phi}z^{-1})(1 - re^{-j\phi}z^{-1})} \quad (9)$$

$$r = \exp\{-p_1 \cdot 2\pi b\text{ERB}(f_r) / f_s\} \quad (10)$$

$$\phi = 2\pi\{f_r + p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (11)$$

$$\varphi = 2\pi\{f_r - p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (12)$$

(p_1 と p_2 は正係数、 f_s はサンプリング周波数)。この振幅周波数特性 $|H_{c0}(f)|$ は、条件(a)・(b)を満足させることができる。しかし、 r が1に近い場合、周波数 $f_r + p_2 \cdot c \cdot b\text{ERB}(f_r)$ 周辺に極大点、 $f_r - p_2 \cdot c \cdot b\text{ERB}(f_r)$ 周辺に極小点が生じてしまい条件(c)を満足させることができない。また r が1よりも小さくなると、特性はなだらかに変化するようになるが、今度は周波数 f_r での傾きが十分でなくなってしまう、 $|H_A(f)|$ を近似するには不十分となる。そこで、極と零点の位置を段階的にずらしたものを何段か縦続接続することによって、極大点や極小点をなるべく生じさせず、かつ周波数 f_r で十分傾きがあるように設定する。そこで、フィルタの伝達関数 $H_c(z)$ を

$$H_c(z) = \prod_k H_{ck}(z) \quad (13)$$

$$H_{ck}(z) = \frac{(1 - r_k e^{j\phi_k} z^{-1})(1 - r_k e^{-j\phi_k} z^{-1})}{(1 - r_k e^{j\phi_k} z^{-1})(1 - r_k e^{-j\phi_k} z^{-1})} \quad (14)$$

$$r_k = \exp\{-k \cdot p_1 \cdot 2\pi b\text{ERB}(f_r) / f_s\} \quad (15)$$

$$\phi_k = 2\pi\{f_r + 2^{k-1} \cdot p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (16)$$

$$\varphi_k = 2\pi\{f_r - 2^{k-1} \cdot p_2 \cdot c \cdot b\text{ERB}(f_r)\} / f_s \quad (17)$$

の形で実現することとし、これを非対称性補償フィルタとする。縦続接続数を4段とした場合の振幅周波数特性 $|H_c(f)|$ を、 $f = f_r$ での値で正規化したものを第1図の破線で示す。周波数 f_r の倍(4000 Hz)以下では近似すべき $H_A(f)$ の特性と非常によく一致している。これに対し、中心周波数から離れるにしたがって、ずれが生じることがわかる。明らかにフィルタの段数を増やすことにより近似度を上げることができるが、ここでの目的はなるべくパラメータ数を少なくすることであるので、4段程度が良いと考えられる。極と零で16個で実現できる。

2.4 振幅周波数特性の近似度

この非対称性補償フィルタにガンマトーンフィルタを縦続接続すれば、ガンマチャープフィルタを近似できる。この振幅周波数応答は、式(6)の $|H_A(f)|$ を $|H_C(f)|$ で置き換えた

$$|G_{CA}(f)| = |G_T(f)| \cdot |H_C(f)| \quad (18)$$

と表わすことができる。このようにして作られたガンマチャープを非対称性補償形ガンマチャープ(Asymmetric Compensation - Gammachirp, AC-Gammachirp)と呼び、式(2)の時間応答で定義されるもともとのガンマチャープとは区別する。図2に、この振幅周波数特性(破線)と、比較のために式(6)のガンマチャープフィルタ(実線)とガンマトーンフィルタ $|G_T(f)|$ (点線)の振幅周波数特性を示す。ここで、 $c = -1$ で、特性がよりよく一致するようにするため補償フィルタの振幅特性 $|H_C(f)|$ を式(5)のピークシフトに比例した周波数の所の値で正規化する。この周波数は第 k 番目で

$$f = f_r + k \cdot p_3 \cdot c \cdot b \text{ERB}(f_r) / n \quad (19)$$

で、 $p_1 \cdot p_2 \cdot p_3$ は、

$$p_1 = 1.35 - 0.19 \cdot |c| \quad (20)$$

$$p_2 = 0.29 - 0.0040 \cdot |c| \quad (21)$$

$$p_3 = 0.23 + 0.0072 \cdot |c| \quad (22)$$

と置いている。すると、図2で $|G_C(f)| > -50$ dBの範囲の周波数ごとの実効誤差は最大で $f_r = 500$ Hzの時の0.41 dBである。また、 b が、1.0、1.35、1.7、 c を、1.0、0、-1.0、-2.0、-3.0、周波数 f_r を250、500、1000、2000、4000、8000 Hzとして組み合わせさせた90組の係数に関し、近似される $|G_C(f)|$ が-50 dB以上の範囲での平均実効誤差は約0.63 dBであり、実効誤差が2 dBを超えるのは f_r が8000 Hzで c が-3.0の3組のみであった。このように、非常によく一致させることができる。さらに、これらの係数を最小2乗の意味で最適化を試みたが、あまり改善がみられなかった。しかし、式(15)から(17)の係数の取りかたを変えればさらに改善できる可能性はある。

2.5 時間応答波形と周波數位相

図3(a)に、非対称性補償形ガンマチャープフィルタの時間応答波形(破線)と式(2)のガンマチャープ(実線)の時間応答波形を示す。この図で $c = -1$ で、応答の形はほとんど違いがないことがわかる。この図での差は、実効振幅で0.0038(最大振幅に対して約-48 dB)程度である。ここで、初段のガンマトーンの初期位相が両者が一致するように合わせ込んでいる。一致させるために初期位相は、 c の値によって変化する。しかし、 c の変化速度がインパルス応答の長さのなか

で変化がほとんどない程度とすれば、位相のずれによる応答波形への影響はないと考えられる。

図 3(b)に、非対称性補償形ガンマチャープフィルタの位相周波数応答(破線)をガンマチャープ(実線)の位相周波数応答とともに示す。 $f = f_r$ での位相を零として表示している。振幅周波数応答の場合と同様、中心周波数 f_r の上下 2 ERB (= 480 Hz) 以内はもちろん、周波数零に近い所以外おおむね一致している。

2.6 高速ガンマチャープフィルタ

以上より線形時不変のガンマトーンフィルタと非対称性補償フィルタを組み合わせることにより、ガンマチャープの特性を精度良く実現できるがわかった。非対称性補償フィルタの特性は固定する必要がなく、時間とともに変化するようにできる。これについては、次節で検討する。この非対称性補償フィルタは、IIR フィルタで構成されているので、AIM 中のガンマトーンフィルタ(Patterson, 1995)や IIR のガンマトーンフィルタ(Slaney(1993))を組み合わせれば、全体での演算時間がきわめて短い高速ガンマチャープフィルタが実現できることになる。

3. フィルタバンクとしての構成

本節では、非対称性補償形ガンマチャープフィルタを使ったフィルタバンクを構成することを考える。

3.1 構成の概念

先に触れたとおり、音圧レベルにより聴覚フィルタの形状が変化する (Lufti and Patterson, 1984; Patterson and Moore, 1966; Moore and Glasberg, 1987)。この音圧を、非対称性に反映させるためには、どこかの段階で音圧を推定する必要がある。この推定には、あらかじめフーリエ変換で音圧を決めるフィルタの入力側説(Moore and Glasberg, 1990)と非対称性を入れた roex フィルタの出力で決める出力側説(Rosen and Baker, 1994)の 2 説がある(Moore, 1995)。実際の聴覚フィルタの前にあらかじめフーリエ変換を取っていないことから、信号処理の観点で前者は問題がある。後者の場合は、非対称性の度合によって出力が変化してしまうので、非対称性のパラメータに何らかの初期値を設定して繰り返し演算で決定する必要が生じる。こちらの方が聴覚系としては可能性があるが、計算モデルとして作るには計算量が多くなり、また収束条件を満足させるかどうかも考慮に入れる必要がある。このことからフィードバック形の制御よりフィードフォワード形の制御の方が望ましい。今回提案した非対称性補償形ガンマチャープフィルタでは、固定線形フィルタバンクのガンマトーンフィルタバンクと時

変非線形の非対称性補償フィルタバンクの2段構成になっている。そこで、このガンマトーンフィルタバンクの出力を使って次の段の非対称性補償フィルタバンクを制御することを考える。ここで、式(5)に示したとおり、ガンマチャープフィルタは非対称性の c の値によりピーク周波数がずれる。これに対し初段のガンマトーンは固定であるので、1つのガンマトーンフィルタ出力から音圧を推定しても、実際にガンマチャープで通過させる周波数での音圧とはならない。そこで、隣あった複数のガンマトーンフィルタ出力から音圧を推定し、後段の処理によって影響が出ないようにする。このようにして構成したフィルタバンクでは、フィルタの入力段・出力段説の長所を残しながらそれぞれの欠点を補う可能性が出る。

3.2 信号処理

図4にここでの信号処理のブロック図を示す。以下 Irino and Patterson (1997)のパラメータを例に従って記述するが、別のパラメータ値に容易に置き換えることができる。入力信号は、ガンマトーンフィルタバンクに入力される。ここでは MATLAB での使いやすさから Slaney(1993)による IIR ガンマトーンフィルタバンクを用いる。今、太線で示された k 番目のチャンネルの処理を考える。ガンマトーンフィルタバンクの出力は、非対称性補償フィルタバンクとパラメータ推定回路群に分けられる。後者では、整流され漏洩積分器に入り平均振幅を計算される。さらに太線のチャンネルを中心に隣接するチャンネルの平均振幅は荷重平均されて、そのチャンネルでの音圧推定のための平均活性度となる。ここで、この荷重関数は、おおよそ平均的な非対称度を持つ関数として、 $n=4$ 、 $b=1.68$ 、 $c=-1$ の時のガンマチャープの振幅周波数特性を使う。この出力の平均活性度 a_{sk} から、推定音圧 \hat{P}_{sk} を、

$$\hat{P}_{sk} = 20 \log_{10}(q \cdot a_{sk}) \quad (23)$$

で求める。ここで、 q は定数である。さらに非対称性のパラメータの推定値 \hat{c}_k は Irino and Patterson (1997)のレベル依存性の式、

$$\hat{c}_k = 3.38 - 0.107 \hat{P}_{sk} \quad (24)$$

から求める。これは最終段の太線で示された非対称性補償フィルタの値として用いられる。このフィルタ出力は、 \hat{c}_k の値に依存した形になる。もし \hat{c}_k がある一定の値 c に固定された場合、元々の c 固定のガンマチャープフィルタ出力を良く近似することは前節で述べたとおりである。

3.3 音圧推定の誤差評価

もともとガンマチャープフィルタのパラメータは、周波数 2000 Hz のプロー

ブ音を用いたノッチ雑音マスキングによって測定した心理物理的なフィルタ形状のデータ(Rosen and Baker, 1994)を、マスキングのパワースペクトルモデルをもとにしてガンマチャープで適合させた場合の結果をもとにして決定した。この実験の刺激音のプローブ音圧、雑音音圧とその帯域幅わかっている。そこで、実際に実験条件の78個のプローブ音+雑音の刺激音を作成して、このフィルタバンクに入力して、その時に得られる推定音圧 \hat{P}_s と入力したプローブ音圧 P_s の誤差を調べる。刺激音を、まず等ラウドネス曲線(ELC)の振幅特性を直線位相FIRフィルタで近似したフィルタに通す。ノッチ雑音法においても、離聴聴取(off-frequency listening)が起きて(Patterson, R. D. and Nimmo-Smith, I., 1980)、フィルタ形状が決定できるのは、SNRが最も大きいチャンネルでのフィルタだけである。この離聴周波数は、形状を決めたときのフィッティングのプログラムでこのフィルタバンクとは無関係に決定できる。得られた離聴周波数にもっとも近いピーク周波数を持つガンマトーンフィルタのチャンネルの出力が、このピーク周波数のフィルタ形状を決定していると仮定する。十分安定した所の平均活性度から式(23)に基づいて \hat{P}_s を計算した。このようにすると、 $q=14.5$ の時、 P_s と \hat{P}_s の自乗誤差が最小となり、実効値で約1.9 dBとなった。この値は、もともとのガンマチャープをデータに適合した時の実効誤差約1.3 dBよりわずかに大きいだけである。用いた荷重関数野形や、IIRガンマトーンフィルタと元々のガンマトーンフィルタの誤差がこの誤差の一部に関与しているであろう。この実効誤差約1.9 dBは式(24)を用いて c の誤差になおすと約0.2であり、フィルタ形状としてはあまり影響がない。最終段の非対称補償形ガンマチャープフィルタと元々のガンマチャープフィルタの振幅特性の実効誤差約0.63 dBは c の誤差になおすと約0.06でこれもほとんど影響を与えない。このことから、このフィルタバンクの構成で、ノッチ雑音法における音圧推定についてはフィルタ形状推定については妥当であるということを示している。

3.4 エキサイテーションパターンの比較

エキサイテーションパターンの比較によりフィルタバンクの評価をする。図5(a)に Moore and Glasberg (1987)の方法で算出したエキサイテーションパターン、図5(b)に上記のガンマチャープフィルタバンクで b の値だけ1.019とし、 c の値を1から-3.5の間に制限した場合のエキサイテーションパターンを示す。図5(b)のエキサイテーションのピークの周波数は、レベルとともに上昇していように見えるが、これは横軸が式(2)の f_r の値を用いていて、 c の値によってガンマチャープフィルタのピーク周波数がずれるためである。このピーク周波数より高周波側では、図5(a)と同様に'upward spread of masking' (Moore and Glasberg, 1987)

に対応する特性が実現されている。これに対し、ピーク周波数より下に、大きな窪みができて、30 dB 以下値の活性度の広がりが多い。これは、推定された \hat{P}_s をもとに式(24)から非対称性パラメータ \hat{c} を求めた分布(図 5(c))から説明できる。音圧推定にはガンマトーンの出力的を使っているので f_r が 2000 Hz の所で \hat{P}_s が一番大きくなり、 \hat{c} の値が一番小さくなる。これより上では、なだらかに \hat{c} が増加し、これにともない低周波側の裾が 2000 Hz の信号になだらかににかからなくなる。これに対し、 f_r が 2000 Hz よりすぐ下の場合、 \hat{c} が負で値が大きいので非対称性が大きく、高周波側の裾が急峻で 2000 Hz の信号を通さないように働く。これに対し、周波数が下がるにつれて、 \hat{c} が 0 から正へと変わって、高周波側に裾がなだらかになるので、2000 Hz の信号を通しやすくなる。結論として、式(23)・(24)をどの周波数のフィルタでも適用しているためにこの窪みが生じる。もしこの現象が実際の基底膜振動で観察されないのであれば、この関係を修正しなくてはならない。

4. まとめ

本報告では、線形時不変のガンマトーンフィルタと IIR フィルタの非対称性補償フィルタを組み合わせることにより、従来 FIR フィルタでしか構成できなかったガンマチャープフィルタの特性を精度良く実現できることを示した。ガンマトーンフィルタに高速演算可能なものを用いれば、全体での演算時間をきわめて短くすることができる。また、時変系の聴覚フィルタバンクを模擬するため、非対称性補償フィルタフィルタバンクの特性をガンマトーンフィルタバンクの出力レベルで制御する形のガンマチャープフィルタバンクの構成法も示した。時変系の聴覚フィルタバンクで、パラメータ数が少なく、演算時間が短いのは利点である。前段階処理として人間の聴覚特性を考慮した周波数分析が必要となるさまざまな信号処理に応用可能である。

今回は、音圧推定値と非対称性パラメータの関係やその他のパラメータをフィルタの周波数にかかわらず一定として求めたが、これについてはさらに心理物理的な実験データを用いて改善する余地がある。また、ノッチ雑音法による同時マスキングのデータやエキサイテーションパターンしか比較の対称がなかったため、動特性を模擬できていない。たとえば、漏洩積分器の時定数は非対称性の変化速度はあまり大きくないという仮定での値で、実際の値ではない。しかし、生理学的な測定で、聴覚末梢系の基底膜の動特性について知見が得られれば、伝送線路モデルよりパラメータ数が少ない分、より明確な形で与えて模擬することが可能になった。聴覚末梢系自体を考察する際にも重要なモデルであると考えられる。

謝辞

日頃より継続的にアドバイスいただき、MRC-APU の Patterson 博士に感謝する。また、第2著者が ATR に滞在し共同研究を行えるようにしていただいた北陸先端大赤木助教授、ATR 人間情報通信研究所東倉元社長・河原元室長・片桐室長に感謝する。有益なコメントをいただいた同研究所、津崎主任研究員・Yehia 研究員に感謝する。

参考文献

- Allen, J. B., Hall, J. L., Hubbard, A., Neely, T. S., and Tubis, A. (Eds.). (1985). "Peripheral Auditory Mechanisms," (Springer-Verlag, Berlin).
- Carney, L. H. and Yin, T. C. T. (1988). "Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model," *J. Neurophysiology*, 60, 1653-1677.
- de Boer, E. and Nuttall, A.L. (1997). "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.
- Fletcher, H. (1940). "Auditory patterns," *Rev. Mod. Phys.*, 12, 47-65.
- Giguere, C. and Woodland, P. C. (1994). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," *J. Acoust. Soc. Am.*, 95, 331-342.
- Glasberg, B. R. and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, 47, 103-138.
- Irino, T. (1995). "An optimal auditory filter," in *IEEE Signal Processing Society, 1995 Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.
- Irino, T. (1996). "A 'gammachirp' function as an optimal auditory filter with the Mellin transform," in *IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-96)*, 2, 981-984, Atlanta, GA.
- Irino, T. and Patterson, R.D. (1997). "A time-domain level-dependent auditory filter: The gammachirp", *J. Acoust. Soc. Am.* 101, pp. 412-419.
- Johannesma, P. I. M. (1972). "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in *Symposium on Hearing Theory*, 58-69, IPO, Eindhoven, Holland.
- Lyon, R. F. (1996). "The all-pole gammatone filter and auditory models," in *Forum Acusticum '96*, Antwerp, Belgium.
- McClellan, J.H., Parks, T.W. and Rabiner, L.R. (1979). "FIR linear phase filter design

- program" in "Programs for Digital Signal Processing" Ed. IEEE, IEEE press, ch. 5.1.
- Moore, B. C. J. and Glasberg, B. R. (1987). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," *Hear. Res.*, 28, 209-225.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1990). "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.*, 88, 132-140.
- Moore, B.C.J.(1995). "Frequency analysis and masking" in *Hearing*, edited by B. C. J. Moore (Academic Press, London).
- Patterson, R. D., Allerhand, M., and Giguere, C. (1995). "Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform," *J. Acoust. Soc. Am.*, 98, 1890-1894.
- Patterson, R. D. and Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic Press, London).
- Patterson, R. D. and Nimmo-Smith, I. (1980). "Off-frequency listening and auditory filter asymmetry," *J. Acoust. Soc. Am.*, 67, 229-245.
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.*, 72, 1788-1803.
- Pflueger, M., Hoeldrich, R. and Reidler, W. (1997), Personal communication.
- Rosen, S. and Baker, R. J. (1993). "Characterising auditory filter nonlinearity," *Hear. Res.*, 73, 231-243.
- Slaney, M.(1993). "An efficient implementation of the Patterson-Holdsworth auditory filter bank" Apple Computer Technical Report #35.
- 入野俊夫 (1995). "聴覚末梢系の計算理論," 日本音響学会聴覚研究会・電子情報通信学会音声研究会, H-95-44, SP95-40, pp.23-30.
- 入野俊夫, Roy D. Patterson(1996). "レベル依存聴覚フィルタとしてのガンマチャープ," 日本音響学会聴覚研究会資料, H-96-73.