Published in final edited form as:
IEEE Trans Audio Speech Lang Processing. 2006 November ; 14(6): 2222–2232. doi:10.1109/TASL.2006.874669.

A Dynamic Compressive Gammachirp Auditory Filterbank

Toshio Irino [Senior Member, IEEE] and
Faculty of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan (e-mail: irino@sys.wakayama-u.ac.jp).

Roy D. Patterson
Centre for Neural Basis of Hearing, Department of Physiology, University of Cambridge, Cambridge CB2 3EG, U.K. (e-mail: roy. patterson@mrc-cbu.cam.ac.uk).

Abstract

It is now common to use knowledge about human auditory processing in the development of audio signal processors. Until recently, however, such systems were limited by their linearity. The auditory filter system is known to be level-dependent as evidenced by psychophysical data on masking, compression, and two-tone suppression. However, there were no analysis/synthesis schemes with nonlinear filterbanks. This paper describe18300060s such a scheme based on the compressive gammachirp (cGC) auditory filter. It was developed to extend the gammatone filter concept to accommodate the changes in psychophysical filter shape that are observed to occur with changes in stimulus level in simultaneous, tone-in-noise masking. In models of simultaneous noise masking, the temporal dynamics of the filtering can be ignored. Analysis/synthesis systems, however, are intended for use with speech sounds where the glottal cycle can be long with respect to auditory time constants, and so they require specification of the temporal dynamics of auditory filter. In this paper, we describe a fast-acting level control circuit for the cGC filter and show how psychophysical data involving two-tone suppression and compression can be used to estimate the parameter values for this dynamic version of the cGC filter (referred to as the “dcGC” filter). One important advantage of analysis/synthesis systems with a dcGC filterbank is that they can inherit previously refined signal processing algorithms developed with conventional short-time Fourier transforms (STFTs) and linear filterbanks.

Keywords

Compression; nonlinear analysis/synthesis auditory filterbank; simultaneous masking; speech processing; two-tone suppression

I. Introduction

It is now common to use psychophysical and physiological knowledge about the auditory system in audio signal processors. For example, in the field of computational auditory scene analysis (CASA) (e.g., [1]), models based on auditory processing [2]-[6] are recommended to enhance and segregate the speech sounds of a target speaker in a multisource environment. It is also the case that popular audio coders (e.g., MP3 and AAC) use human masking data in their “perceptual coding,” to match the coding resolution to the limits of human perception on a moment-to-moment basis [7]-[11]. Nevertheless, most speech segregation systems and audio coders still use nonauditory forms of spectral analysis like the short-time Fourier transform (STFT) and its relatives. One of the major reasons is their computational efficiency. It is also the case that simple auditory models with linear auditory filterbanks do not necessarily improve the performance of audio processors. Research over the past two decades shows that the auditory filter is highly nonlinear and it is dynamic;

Europe PMC Funders Author Manuscripts

specifically, the frequency response of the auditory filter exhibits level-dependent asymmetry [12]-[14] and a compressive input/output function [15]-[17], and both of these characteristics are fundamentally dynamic; that is, the filter adapts to signal amplitude with a time constant on the order of 1 ms. It seems likely that these nonlinear characteristics are partly responsible for the robustness of human speech recognition, and that their inclusion in perceptual processors would make them more robust in noisy environments. In this paper, we introduce a dynamic version of the compressive gammachirp filter with a new level-control path that enables the filter to explain "two-tone suppression," a prominent nonlinear feature of human masking data. Dynamic auditory filterbanks with these properties should also be useful as preprocessors for hearing aids [18].

The use of a nonlinear filterbank raises a problem for analysis/synthesis processors, because there is no general method for resynthesizing sounds from auditory representations produced with nonlinear filterbanks. So, although there are a number of dynamic nonlinear cochlear models based on transmission-line systems (e.g., [19], [20]) and filterbanks (e.g., [21]), none of them supports the analysis/synthesis framework. The reason is that they were developed to simulate auditory peripheral filtering, and the brain does not resynthesize directly from the encoded representation. This is a serious constraint for CASA systems, where the resynthesized version of the target speaker is used to evaluate the performance of the system. The filter structures in cochlear models are complex and, typically, the specification of the impulse response is not sufficiently precise to support high-quality resynthesis. Recently, we developed a linear auditory filterbank with the aim of eventually developing a nonlinear analysis/synthesis system [22]. In this paper, we demonstrate how the linear system was extended to produce a dynamic nonlinear auditory filterbank that can explain a substantial range of nonlinear behavior observed in psychophysical experiments. We also demonstrate how it can be used as the basis for an analysis/synthesis, perceptual processor for CASA and speech research.

Theoretically, within the framework of wavelet (e.g., [23]), inversion is straightforward when the amplitude and phase information is preserved. It can be accomplished using filterbank summation techniques after compensation for the group delay and phase lag of the analysis filter. The same is not true, however, for nonlinear filterbanks. There were a limited number of studies of inversion with auditory filterbanks where part of the phase information was missing [25]-[27]. The resynthesis technique involved an iterative process which had local minima problems and which precluded establishing a one-to-one correspondence between the representation and the resynthesized signal. Moreover, the resynthesized sounds were distorted even when there was no manipulation of the coded representation because these systems can never guarantee high-quality reconstruction. Thus, what is required is a nonlinear filterbank that enables properly defined resynthesis, at least when the amplitude and phase information are preserved. A nonlinear dynamic filterbank that can guarantee the fidelity of a processor would enable us to manipulate the encoded representation of a sound and then resynthesize the corresponding sound appropriately. Such a system could inherit the many excellent signal-processing algorithms developed previously in the linear domain (e.g., [28]), while avoiding the problems of the STFT and the linear filterbank. Thus, the framework should be useful for a range of applications from coding and speech enhancement to speech segregation [1]-[6] and hearing aids [18].

The gamma*chirp* auditory filter [22], [29]-[31] was developed to extend the domain of the gamma*tone* auditory filter [32], to provide a realistic auditory filterbank for models of auditory perception and to facilitate the development of a nonlinear analysis/synthesis system. A brief summary of the development of the gammatone and gammachirp filterbanks over the past 20 years is provided in[31, Appendix A]. The resultant compressive gammachirp filter (cGC) was fitted to a large body of simultaneous masking data obtained

psychophysically [31]. The cGC consists of a passive gammachirp filter (pGC) and an asymmetric function which shifts in frequency with stimulus level as dictated by data on the compression of basilar membrane motion. The fitting of the psychophysical data in these studies was performed in the frequency domain without temporal dynamics.

A time-varying version of the gammachirp filterbank was proposed [22], [33] in which an infinite impulse response (IIR) asymmetric compensation filter (AF) was defined to simulate the asymmetric function. The filter is minimum phase and, thus, invertible. Moreover, since it is a time-varying linear filter, it is possible to invert the signal even when the filter coefficients are time-varying if the history of the coefficients from the analysis stage is preserved and applied properly in the synthesis stage. (Indeed, it is only necessary to preserve the history of the estimated signal level, since the filter coefficients are entirely determined by the signal level.) This enables us to resynthesize sound from the output of the dynamic filterbank. The resynthesized sound is very similar to the original input sound; the fidelity is limited simply by the frequency characteristics and the density of the filters, and the total bandwidth of the linear analysis/synthesis filterbank. When the coefficients of the IIR asymmetric compensation filter are controlled by the estimated level of the input signal, the system has nonlinear characteristics that enable it to explain psychophysical suppression and compression data.

Thus, all that is actually required is to extend the static version of the cGC filter into a dynamic level-dependent filter that can accommodate the nonlinear behavior observed in human psychophysics. In this paper, we use psychophysical data involving two-tone suppression [34], [35] and compression [15], [16] to derive the details of the level control circuit for a dynamic version of the cGC. We then go on to describe an analysis/synthesis filterbank based on the cGC that can resynthesize compressed speech.

II. Gammachirp Auditory Filters

Fig. 1 is a block diagram of the proposed gammachirp analysis/synthesis filterbank. The system consists of a set of linear passive gammachirp filters, a set of asymmetric compensation filters both for analysis and synthesis, and a level estimation circuit. Between the analysis and synthesis stages, it is possible to include a very wide range of signal processing algorithms including ones previously developed with linear systems. This section explains the dynamic, compressive gammachirp (dcGC) filterbank in terms of A) the mathematical background of the compressive gammachirp (cGC) filter [29]-[31] and the method used to fit it to psychophysical masking data [12]-[14], B) a time-domain implementation of the cGC filter [22], [33], C) the incorporation of a new level estimation circuit, in a channel somewhat higher in frequency than the signal channel, that enables the system to accommodate two-tone suppression data [34], [35] and compression data [15], [16], and D) a discussion of the computational costs.

A. Compressive Gammachirp Filter Function

The complex analytic form of the gammachirp auditory filter [29] is

$$g_c(t) = at^{n_1-1}\exp(-2\pi b_1 \mathrm{ERB}_N(f_{r1})t) \times \exp(j2\pi f_{r1}t + jc_1 \ln t + j\varphi_1) \quad (1)$$

where a is amplitude; n_1 and b_1 are parameters defining the envelope of the gamma distribution; c_1 is the chirp factor; f_{r1} is a frequency referred to as the asymptotic frequency since the instantaneous frequency of the carrier converses to it when t is infinity; $\mathrm{ERB}_N(f_{r1})$ is the equivalent rectangular bandwidth of average normal hearing subjects [13], [14]; φ_1 is the initial phase; and $\ln t$ is the natural logarithm of time. Time is restricted to positive values. When $c_1 = 0$, (1) reduces to the complex impulse response of the gammatone filter.

$$g_t(t) = at^{n-1}\exp(-2\pi b\mathrm{ERB}_N(f_r)t)\exp(j2\pi f_r t + j\varphi). \quad (2)$$

The Fourier magnitude spectrum of the gamma*chirp* filter is

$$|G_C(f)| = a_\Gamma \cdot |G_T(f)| \cdot \exp(c_1\theta_1(f)) \quad (3)$$

$$\theta_1(f) = \arctan\left(\frac{f - f_{r1}}{b_1\mathrm{ERB}_N(f_{r1})}\right). \quad (4)$$

$|G_T(f)|$ is the Fourier magnitude spectrum of the gamma*tone* filter, and exp ($c_1\theta_1(f)$) is an asymmetric function since θ_1 is an antisymmetric function centered at the asymptotic frequency, f_{r1} (4). a_Γ is a constant.

Irino and Patterson [30] decomposed the asymmetric function exp ($c_1\theta_1(f)$) into separate low-pass and high-pass asymmetric functions in order to represent the passive basilar membrane component of the filter separately from the subsequent level-dependent component of the filter to account for compressive nonlinearity observed psychophysically. The resulting "compressive" gammachirp filter $|G_{cc}(f)|$ is

$$\begin{aligned} |G_{cc}(f)| &= [a_\Gamma |G_T(f)| \cdot \exp(c_1\theta_1(f))] \cdot \exp(c_2\theta_2(f)) \\ |G_{cc}(f)| &= |G_{cp}(f)| \cdot \exp(c_2\theta_2(f)). \end{aligned} \quad (5)$$

Conceptually, this compressive gammachirp is composed of a level-*in*dependent, "passive" gammachirp filter (pGC) $|G_{CP}(f)|$ that represents the passive basilar membrane, and a level-dependent, high-pass asymmetric function (HP-AF) exp($c_2\theta_2(f)$) that simulates the active mechanism in the cochlea. The filter is referred to as a "compressive" gammachirp (cGC) because the compression around the peak frequency is incorporated into the filtering process itself. The HP-AF makes the passband of the composite gammachirp more symmetric at lower levels.

Fig. 2 illustrates how a level-dependent set of compressive gammachirp filters (cGC; upper set of five solid lines; left ordinate) can be produced by cascading a fixed passive gammachirp filter (pGC; lower solid line; right ordinate) with a set of high-pass asymmetric functions (HP-AF; set of five dashed lines; right ordinate). When the leftmost HP-AF is cascaded with the pGC, it produces the uppermost cGC filter with most gain. The HP-AF shifts up in frequency as stimulus level increases and, as a result, at the peak of the cGC, gain *decreases* as stimulus level increases [30]. The filter gain is normalized to the peak value of the filter associated with the highest probe level, which in this case is 70 dB.

The angular variables are rewritten in terms of the center frequency and bandwidth of the passive gammachirp filter and the level-dependent asymmetric function to accommodate the shifting of the asymmetric function relative to the basilar membrane function with level. If the filter center frequencies are f_{r1} and f_{r2}, respectively, then from (4)

$$\theta_1(f) = \arctan\left(\frac{f - f_{r1}}{b_1\mathrm{ERB}_N(f_{r1})}\right)$$

and

$$\theta_2(f) = \arctan\left(\frac{f - f_{r2}}{b_2 \mathrm{ERB}_N(f_{r2})}\right). \quad (6)$$

The peak frequency f_{p1} of pGC is

$$f_{p1} = f_{r1} + c_1 b_1 \mathrm{ERB}_N(f_{r1}) / n_1 \quad (7)$$

and the center frequency f_{r2} of HP-AF is defined as

$$f_{r2} = f_{\mathrm{rat}} \cdot f_{p1}.$$

In this form, the chirp parameters, c_1 and c_2, can be fixed, and the level dependency can be associated with the frequency ratio f_{rat}. The peak frequency f_{p2} of the cGC is derived from f_{r2} numerically. The frequency ratio f_{rat} is the main level-dependent variable when fitting the cGC to the simultaneous masking data merically [30], [31]. The total level at the output of the passive GC P_{gcp} was used to control the position of the HP-AF. Specifically

$$f_{\mathrm{rat}} = f_{\mathrm{rat}}^{(0)} + f_{\mathrm{rat}}^{(1)} \cdot P_{\mathrm{gcp}}. \quad (8)$$

The superscripts 0 and 1 designate the intercept and slope of the line.

In Fig. 2, as the signal level increases, the peak frequency of the cGC filter first increases slightly and then decreases slightly, because the pGC filter is not level independent in the current model. It would be relatively easy to include monotonic level-dependency in the peak frequency f_{p2} of the cGC filter by introducing a level-dependency in the asymptotic frequency f_n of the pGC filter. In this case, the pGC filters would not necessarily be equally spaced along the ERB_N rate axis. It is, however, beyond the scope of this paper because 1) the level-dependent peak frequency cannot be estimated from the notched noise masking data used to determine the coefficients of the current cGC filter, 2) a small amount of peak fluctuation does not affect the output of the filterbank much since adjacent filters tend to shift together in the same direction, and 3) it is simpler to use a linear pGC filter for the discussion of analysis/synthesis filterbanks.

A detailed description of the procedure for fitting the gammachirp to the psychophysical masking data is presented in [31, Appendix B]. Briefly, the five gammachirp filter parameters b_1, c_1, b_2, c_2 and f_{rat} were allowed to vary in the fitting process; n_1 was fixed at 4. The filter coefficients were found to be largely independent of peak frequency provided they were written in terms of the critical band function (specifically, the ERB_N rate function [14], [31]). So, each filter parameter can be represented by a single coefficient. The f_{rat} parameter has to change with level and so it requires two coefficients. This means that a dynamic, compressive gammachirp filterbank that explains masking and two-tone suppression data for a very wide range of center frequencies and stimulus levels can be described with just six coefficients [31], whose values are as listed in the second row of Table I.

B. Time Domain Implementation

The description above is based on the frequency-domain response of the gammachirp filter. For realistic applications, it is essential to define the impulse response. The following is a brief summary of implementation; the details are presented in [22], [30], and [33].

The high-pass asymmetric function $\exp(c_2\theta_2)$ does not have an analytic impulse response. So, an asymmetric compensation filter was developed to enable simulation of the cGC impulse response, in the form

$$g_{cc}(t) = a_c \cdot g_{ca}(t) * h_c(t). \quad (9)$$

Here, a_c is a constant, $g_{ca}(t)$ is the gammachirp impulse response from (1), and $h_c(t)$ is the impulse response of the asymmetric compensation filter $H_c(f)$ that simulates the asymmetric function such that

$$|H_C(f)| \cong \exp(c \cdot \theta). \quad (10)$$

The asymmetric compensation filter [22], [33] is defined in the z-plane as

$$H_C(z) = \prod_{k=1}^{N} H_{Ck}(z) \quad (11)$$

$$H_{Ck}(z) = \frac{\left(1 - r_k e^{j\phi k} z^{-1}\right)\left(1 - r_k e^{-j\phi k} z^{-1}\right)}{\left(1 - r_k e^{j\varphi k} z^{-1}\right)\left(1 - r_k e^{-j\varphi k} z^{-1}\right)} \quad (12)$$

$$r_k = \exp\left\{-p_1 (p_0/p_4)^{k-1} \cdot 2\pi b \mathrm{ERB}_N(f_r)/f_s\right\} \quad (13)$$

$$\varphi_k = \begin{cases} 2\pi\{f_r + \Delta f_r\}/f_s, & f_r + \Delta f_r \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

$$\phi_k = \begin{cases} 2\pi\{f_r - \Delta f_r\}/f_s, & f_r - \Delta f_r \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

$$\Delta f_r = (p_0 \cdot p_4)^{k-1} \cdot p_2 \cdot c \cdot b\mathrm{ERB}_N(f_r) \quad (16)$$

where p_0, p_1, p_2, and p_4 are positive coefficients; f_s is the sampling rate; and N is the number of filters in the cascade. When $N = 4$ (which is the case throughout this paper)

$$p_0 = 2,\ p_4 = 1.0724$$
$$p_1 = 1.7818 \cdot (1 - 0.0791 \cdot b) \cdot (1 - 0.1655 \cdot |c|)$$

and

$$p_2 = 0.5689 \cdot (1 - 0.1620 \cdot b) \cdot (1 - 0.0857 \cdot |c|).$$

With these values, the discrepancy between $|H_c(f)|$ and $\exp(c \cdot \theta)$ is small in the critical region near the asymptotic frequency f_r [33]. Since the asymmetric compensation filter is always accompanied by the bandpass filter of the gammatone or gammachirp filter, the error in the combined filter is reliably reduced to less than 1 dB within the wide range required by parameters b and c. It is also the case that the impulse responses are in excellent agreement. The coefficients p_2 and p_4 are functions of the parameters b and c. So, it is also possible to

derive the values on a sample-by-sample bases even when b and c are time-varying and level-dependent, although it is not the case of the current simulation.

Since the asymmetric compensation filter is a minimum phase filter, it is possible to define the inverse filter which is

$$H_c^{-1}(b, c, f_r; z) = H_c(b, -c, f_r; z) \quad (17)$$

since the numerator and denominator in (12) are invertible depending on the sign of c. The inverse filter is a low-pass filter when the analysis filter is a high-pass filter, so that their product is unity. The crucial condition is to ensure that it is possible to invert the filtered signal, even when the parameters b, c, and f_r vary with stimulus level [22], [33]; the coefficients used in the analysis are preserved and precisely applied in the synthesis. In the current study, it is sufficient to preserve the temporal sequences of the estimated levels since the gammachirp parameters are level-independent except for f_{rat}, which is a linear function of the level as in (8).

Fig. 1 shows the block diagram of the cGC analysis/synthesis filterbank. The initial block is a bank of linear pGC filters; the second block is a bank of HP-AF filters which simulate the high-pass asymmetric function in (9) and (10). We refer to both the high-pass filter and the high-pass function as “HP-AF” for simplicity, since there is a one-to-one correspondence between them. Together, this cascade of filterbanks represent the dcGC filterbank; the architecture of the dcGC filter itself is described in the next section. After arbitrary signal processing of the dcGC output, it is possible to resynthesize the sound: 1) The outputs of filterbank are applied to a bank of low-pass asymmetric compensation filters (LP-AFs) that is the inverse of the HP-AF filterbank as in (17) and has level-dependent coefficients based on the estimated level at the analysis filterbank. (2) The linearized filterbank outputs are applied to a time-reversed pGC filterbank and then summed up across the channel. When there is no signal processing between the analysis and resynthesis stages, the resynthesized sound is almost indistinguishable from the input sound. The degree of precision is determined by the passband of the linear pGC filterbank and the density of the filters. There are many possible variations of the architecture, depending on the purpose of the signal processing. For example, in Section III-C, we demonstrate resynthesis from compressed speech by removing the LP-AF filterbank; under normal circumstances, the original, noncompressed speech is recovered as described above.

C. Filter Architecture

Preliminary simulations had shown that the previous cGC filterbank with six coefficients (second row in Table I) could not explain two-tone suppression data (e.g., [34], [35]). So, we had to modify the filterbank architecture. Since the cGC has a precise frequency response, it is possible to simulate two-tone suppression in the frequency domain just as we did when fitting the simultaneous masking data. This greatly reduces the simulation time required to find a reasonable candidate for the filter architecture from the enormous number of possible variations. The result was the filter architecture shown in Fig. 3.

As in the previous compressive gammachirp [31], there are two paths which have the same basic elements; one path is for level-estimation and the other is for the main signal flow. The signal path (bottom blocks) has a pGC filter with parameters, b_1, c_1, f_{p1}, and a HP-AF with parameters b_2, c_2, f_{r2} (= $f_{\mathrm{rat}} \cdot f_{p1}$. This combination of pGC and HP-AF results in the compressive gammachirp (cGC) defined in (5) with peak frequency f_{p2}. The parameter values are the same as in the previous study and are listed in the fourth row of Table I. The level-estimation path (upper blocks) has a pGC with parameters, b_1, c_1, f_{p1L} and an HP-AF with parameters b_2, c_2, f_{r2L} (= $f_{\mathrm{ratL}} \cdot f_{p1L}$). The components of the level-estimation path are

essentially the same as those of the signal path; the difference is the level-*independent* frequency ratio, f_{ratL}. The peak frequency f_{p1L} of the pGC in the level-estimation path is required to satisfy the relationship

$$\mathrm{ERB_N\,rate}\left(f_{p1L}\right)=\mathrm{ERB_N\,rate}\left(f_{p1}\right)+r_{EL} \quad (18)$$

where $\mathrm{ERB_N rate}(f)$ is the $\mathrm{ERB_N}$ rate at frequency f [13], [14], and r_{EL} is a parameter that represents the frequency separation between the two pGC filters on the ERB rate axis.

The output of the level-estimation path is used to control the level-dependent parameters of the HP-AF in the signal path. In order to account for the different rates of growth of suppression in the upper and lower suppression regions [35], it was necessary to use not only the level at the output of the pGC as in the previous cGC [31], but also the level of the output of the HP-AF. The level P_c was estimated in decibels on a sample-by-sample basis and used to control the level in the signal path.

If the outputs of the pGC and HP-AF in the level-estimation path are s_1 and s_2, then the estimated linear levels $\bar{s}_1$ and $\bar{s}_2$ are given by

$$\bar{s}_1(t)=\max\left\{\bar{s}_1(t-\Delta t)\cdot e^{-\ln 2\cdot(\Delta t/\tau_L)},\max(s_1(t),0)\right\}$$

and

$$\bar{s}_2(t)=\max\left\{\bar{s}_2(t-\Delta t)\cdot e^{-\ln 2\cdot(\Delta t/\tau_L)},\max(s_2(t),0)\right\} \quad (19)$$

where Δt is the sampling time, and T_L is the half-life of the exponential decay. It is a form of "fast-acting slow-decaying" level estimation. The estimated level tracks the positive output of the filter as it rises in level, but after a peak, the estimate departs from the signal and decays in accordance with the half-life. The effect of the half-life on the simulation of compression is illustrated in Section III-B. The control level $P_c(t)$ is calculated as a weighted sum of these linear levels in decibels.

$$P_c(t)=20\log_{10}\left\{w_L\cdot a_{RL}\left(\frac{\bar{s}_1(t)}{a_{RL}}\right)^{v_{1L}}+(1-w_L)\cdot a_{RL}\left(\frac{\bar{s}_2(t)}{a_{RL}}\right)^{v_{2L}}\right\} \quad (20)$$

and

$$a_{RL}=10^{p_{RL}/20}$$

where w_L, v_{1L}, and v_{2L} are weighting parameters, P_{RL} and is a parameter for the reference level in decibels.

In the filterbank, the asymptotic frequencies f_{n} of the pGC filters are uniformly spaced along the $\mathrm{ERB_N}$ scale. The peak frequencies f_{p1} of the pGC filters are also uniformly spaced and lower than the asymptotic frequencies f_{p1}, since $c_1 < 0$ in (7). The peak frequencies f_{p2} of the dcGC filters are, of course, level-dependent and closer to the asymptotic frequencies f_{r1} of the pGC filters. The resultant filterbank is referred to as a dcGC auditory filter.

We used an equal-loudness contour (ELC) correction to simulate the outer and middle-ear transfer functions [13], [14] in the following simulations. The ELC filter is implemented with an FIR filter, and it is possible to define an inverse filter for resynthesis.

D. Computational Cost

The computational cost of a filterbank is one of important properties, particularly in realtime applications. We estimated the computational cost in terms of the total number of filters in the system. The cGC filter consists of a gammatone filter (GT), a lowpass asymmetric compensation filter (LP-AF), and a highpass asymmetric compensation filter (HP-AF) as in (5). The GT filter is implemented with a cascade of four second-order IIR filters [36]. The LP-AF and HP-AF filters are also implemented with a cascade of four second-order IIR filters. So, there are a total of 12 second-order IIR filters for one channel of the signal path. Since the pGC filter in the level-estimation path of one cGC filter is identical to the pGC in the signal path of a cGC filter with a higher peak frequency, it is not necessary to calculate the output of the pGC filter in the level-estimation path twice. The HP-AF in the level estimation path is necessary and is also implemented as a cascade of four second-order IIR filters. So, in total, one channel in the analysis filterbank requires calculation of 16 second-order IIR filters.

For the synthesis filterbank, it is necessary to use a cascade of four second-order IIR filters per channel for the LP-AF filter (inverse of HP-AF) to linearlize the nonlinear representation. The temporally-reversed gammachirp filterbank is not essential when considering the cost because the synthesis is accomplished with a filtebank summation technique after compensating for the group delay and phase lag of the analysis filter. The maximum group delay is defined as the group delay of the gammachirp auditory filter with the lowest center frequency; it is just under10 ms when the lowest center frequency is 100 Hz.

The computational cost increases linearly with the number of channels. It is, however, possible to reduce the cost considerably by down sampling. It should now be possible to produce a real time version of the analysis and synthesis components. So, the total computational cost would largely depend on the cost of the signal processing implemented between the analysis and synthesis filterbanks.

In the current study, we used two filterbanks—one for the two-tone suppression data and one for the compression data. The suppression filterbank had 100 channels covering the frequency range from 100 to 4000 Hz (i.e., ERBNrates from 3.4 to 27) The compression filterbank also had 100 channels with a frequency range from 100 to 15 000 Hz (i.e., ERB_N rates from 3.4 to 39). The filter densities were 4.2 and 2.8 filters per ERB_N rate, respectively, which was sufficient to obtain reasonbly accurate paramater values. The sampling rate was 48 000 Hz, and no down sampling was used since the fitting procedure does not need to run in real time. The maximum center frequency of the auditory filter needs to be less than one quarter of the sampling rate in order to define the filter impulse response properly. In the simulation of compression, however, there was no problem since the maximum frequency of the signal components was 6000 Hz and the sampling rate was 48 000 Hz.

III. Results

This section illustrates the use of the dcGC filterbank to simulate two-tone suppression and compression, and the potential of the filterbank in speech processing. The dcGC filter parameters b_1, c_1, f_{rat}, b_2 and c_2 (Table I) are essentially the same values as for the previous cGC filter used to fit the simultaneous masking data [31]. These specific values were

determined with a fitting procedure that was constrained to minimize the number of free parameters as well as the rms error of the fit. The frequency ratio parameters, f_{ratL}, in the level-estimation path is 1.08 so that the peak gain of the cGC is 0 dB when the peak gain of the pGC is 0 dB, as it is in this simulation. The other level-estimation parametes r_{EL}, w_L, v_{1L}, v_{2L} and P_{RL} were set to the values listed in the bottom row of Table I which were derived from preliminary simulations.

A. Two-Tone Suppression

Two-tone suppression [34], [35] is one of the important characteristics for constructing an auditory filterbank. The amplitude of the basilar membrane in response to a "probe" tone at a given frequency is *reduced* when a second "suppressor" tone is presented at a nearby frequency at a higher level. The suppressor dominates the level-estimation path of the dcGC (Fig. 3) where it increases the compression of the probe tone by shifting the HP-AF of the signal path.

The method for simulating suppression is simple. A probe tone about 100 ms in duration and 1000 Hz in frequency is presented to the filterbank, and the output level of the filter with the peak at the probe frequency is calculated, in decibels, for various suppressor tones. Fig. 4 shows the suppression regions (crosses) and the probe tone (triangle). They show combinations of suppressor-tone frequency and level where the suppressor-tone reduces the level of the filter output at the probe frequency by more than 3 dB. There are regions both above and below the probe frequency. The solid curve shows the "excitatory" filter, that is, the inverted frequency response of the dcGC with a peak frequency of 1000 Hz, when the probe tone level is 40 dB. The dashed lines centered at about 1100 and 1300 Hz show the "suppressive" filters, that is, the inverted frequency response curves of the pGC and cGC in the level estimation path, respectively. When the estimated level of an input signal increases, the HP-AF in the signal path moves upward in the frequency and reduces or "suppresses" the output level of the signal path. The two-tone suppression is produced by the relationship between these excitatory and suppressive filters.

The dashed and dotted lines show the suppression regions observed psychophysically with the pulsation threshold technique [35]; the simulated suppression regions are quite similar to the observed regions except for the upper-left corner of the high-frequency region. The discrepancy arises partially because the upper skirt of the dcGC filter is shallower than what is usually observed in physiological measurements. The current parameters were derived from two large databases of human data on simultaneous masking without any constraints on the upper slope. The simulated suppression areas could be manipulated to produce a better fit by changing the filter parameters if and when the correspondence between the physiological and psychophysical data becomes more precise. The current example serves to demonstrate that the dcGC filter produces suppression naturally and it is of roughly the correct form.

At this point, it is more important to account for the asymmetry in the growth of suppression with stimulus level in the lower and upper suppression regions [35]. Plack *et al.* [16] reported that the current dual resonance nonlinear (DRNL) model [21] could not account for the asymmetry in growth rate even when the parameters were carefully selected. Fig. 5 shows the relative output level of the dcGC filter for a 1000-Hz probe tone, as a function of suppressor level, when the suppressor frequency is either 400 Hz (left panel) or 1400 Hz (right panel). It is clear that the absolute growth rate of the suppression for the lower suppressor frequencies is greater than for the upper suppressor frequencies. It is also the case that the suppressor levels are different for the "bend points" (or "break points" in [35, Fig. 11]), where the output level starts to decrease as the suppressor level increases. The bend-point levels for a 40-dB probe tone are about 60 dB for 400 Hz and 40 dB for 1400 Hz. This

difference it appears to be largely due to the difference in the curvature of the suppression curve; it is more acute in the lower region and more gradual in the upper region.

The maximum absolute growth rate is about 0.4 dB/dB when the suppressor frequency is 400 Hz. In contrast, the maximum slope is about 0.3 dB/dB when the suppressor frequency is 1400 Hz. Note that the output level is compressed by the very nature of the dcGC architecture, and the degree of compression increases as the probe level increases. The observed decrement in the depth for the 60-dB tone does not necessarily mean the actual suppression slope decreases. To avoid the effect of compression, the degree of suppression was measured in terms of the input signal level so that the output level at the probe frequency was unchanged before and after the suppressor was introduced. Using this criterion, the growth rates in the model data increase slightly to about 0.5 and 0.3 dB/dB, respectively, when the probe is 40-dB sound pressure level (SPL). The suppression levels in psychophysical data vary considerably with listener and level [35]; the rates are 0.5-3 dB/dB for a 400-Hz suppressor as in [35, Fig. 4], and less than 0.2 dB/dB for one subject (no data for other subjects) for a 1400-Hz suppressor as in [35, Fig. 10]. The reason for the variability across listeners and levels is unclear. The growth rates in the lower frequency suppressor are generally much larger than the rates in the current simulation. We could change the level-estimation parameter values or modify the level estimation function in (20) to accommodate the data. It is, however, not currently clear which set of data is the most appropriate or reliable, and so we will not pursue the fitting further in this paper. We did, however, confirm that we were able to change the depth of suppression for 400- and 1400-Hz suppressors by changing the weight parameters w_L, v_{1L} and v_{2L}. For current purposes, it is sufficient to note that the dcGC filter produces two-tone suppression, the growth rate is greater on the low-frequency side of the probe tone, and qualitatively, at least, the model is consistent with psychophysical data unlike the DRNL filter model [16], [21].

B. Compression

Compressive nonlinearity is also an important factor in the auditory filterbanks. Oxenham and Plack [15] estimated the compression characteristics for humans using a forward-masking paradigm. They also explained the data using a DRNL filter model [21]. This section shows how the dcGC filter can also explain the compression data.

1) Method—The experiment in question [15] was performed as follows: a brief, 6000-Hz, sinusoidal probe was presented at the end of a masker tone whose carrier frequency was either 3000 or 6000 Hz, depending on the condition. The probe envelope was a 2-ms Hanning window to restrict spectral splatter; the duration of the masker was 100 ms. In addition, a low-level noise was added to the stimulus to preclude listening to low-level, off-frequency components. Threshold for the probe was measured using a two-alterative, forced choice (2AFC) procedure in which the listener was required to select the interval containing the probe tone. The level of the masker was varied over trials to determine the intensity required for a criterion level of masking.

The dcGC filter was used to simulate the experiment as follows: The output of each channel of the dcGC filterbank was rectified and low-pass filtered to simulate the phase-locked neural activity pattern (NAP) in each frequency channel, and then the activation was averaged using a bank of temporal windows to simulate the internal auditory level of the stimulus. The window was rectangular in shape, 20-ms in duration, and located to include the NAPs of the end of the masker and the probe. The shape of the temporal window does not affect the results because it is a linear averaging filter and the temporal location of the probe tone is fixed. The output levels for all channels were calculated for the masker alone and the masker with probe, and the array was scanned to find the channel with the maximum

difference, in decibels. The calculation was performed as a function of masker level in 1-dB steps. Threshold was taken to be the masker level required to reduce the difference in level between the two intervals to 2 dB in the channel with the maximum difference. The half-life of the level estimation was varied to minimize the masker level at threshold; the remaining parameter values were exactly the same as in the simulation of the two-tone suppression data (Table I).

2) Results—Fig. 6 shows the experimental results [15] as thick dashed lines. The simulation was performed for seven half-lives ranging from 0 to 5 ms (19), and the results are presented by thin solid lines. The solid lines above the dotted diagonal show the simulated threshold when the probe and masker have different frequencies, namely, 6000 and 3000 Hz. It is clear that the half-life affects the growth of masked threshold. When the half-life is 0.5 or 1 ms, the change in the growth rate is very similar to that in the experimental data (thick dashed line). The average growth rate is larger in other conditions; it is about 0.5 dB/dB when the half-life is 5 ms and it is more than 0.3 dB/dB when the half-life is 0.1 ms. When the half-life is 0 ms, the average slope is close to 0.8 dB/dB which means almost no compression. So, the level-estimation process must be quick, but not instantaneous, with a half-life on the order of 0.5-1.0 ms.

The best fit would appear to be for a half-life of 0.5 ms. In this case, the simulation error is less than 3 dB, since we set the threshold criterion to 2.0 dB to minimize this error. Threshold for the condition where the probe and masker have the same frequency (namely, 6000 Hz) is located a few decibels below the dotted diagonal line. The threshold functions are almost the same, despite relatively large half-life differences, and they are essentially linear input-output functions. This is consistent with the psychophysical data, at least, for one subject [23]. When the threshold criterion decreases, the lines for both conditions shift up in the same way, that is, both when the probe and masker have the same frequency and when they have different frequencies. We would still need to explain the subject variability which can be more than 5 dB when the probe and masker have the same frequency. We would also need to estimate the half-life for frequencies other than 6000 Hz, which is not possible currently because there are no psychophysical data for other frequencies.

In summary, the current model provides a reasonable account of the compression data; with the exception of the time constant, the parameters values were identical to those used to explain two-tone suppression and simultaneous masking.

C. Speech Processing

It appears that the dcGC analysis/synthesis filterbank can be used to enhance the plosive consonants in speech and the high-frequency formants of back vowels. The effects are illustrated in Fig. 7 which shows three “cochlear” spectrograms, or “cochleograms,” for the Japanese word “aikyaku”; the three segments of each cochleogram correspond to “ai,” “kya,” and “ku.” The cochleograms were produced by the pGC filterbank on its own (a), the linear cGC filterbank without dynamic level-estimation and when the control level P_c was fixed at 50 dB (b), and the dcGC filterbank with dynamic level-estimation (c). The output of each filterbank was rectified, averaged for 2 ms with a frame shift of 1 ms, normalized by the rms value of the whole signal, and plotted on a linear scale. The smearing of the formants in (a) arises from the fact that the pGC filter has a much wider passband than either the cGC or dcGC filter. Compare the representations of the plosives around 350 and 570 ms, and the representation of the high-frequency formants of the vowel in “ku” in the region beyond 600 ms. The comparisons show that the dcGC filter compresses the dynamic range of the speech which emphasizes the plosive consonants and the higher formants of back

vowels, and do so without the need of a separate compression stage like those typically used with linear auditory filterbanks or short-time Fourier transforms.

Fig. 8 shows excitation patterns (or frequency distributions) derived from the same speech segment at points centered on 60 ms (a) and 630 ms (b) in the sustained portions of the /a/ and /u/ vowels, respectively. The solid curve was derived by averaging the output of the dcGC filterbank [Fig. 7(c)] for 21 ms (1024 sample points). The dashed curve was derived from the output of the linear cGC filterbank [Fig. 7(b)] and the total rms level was set to the same level as the output of the dcGC filterbank. The excitation patterns of the nonlinear dcGC and linear cGC filterbanks are similar but in both cases the dcGC filterbank increases the relative size of the upper formants, and the effect is stronger for the /u/ which has the weaker upper formants [Fig. 8(b)]. The dashed and dotted curve is a level-dependent excitation pattern derived with a roex filterbank [13], which is provided for reference. The pattern was calculated from the signal level produced by a STFT with a hanning window of 1024 points.

The speech can be resynthesized from the cochleograms using the time-reversed pGC filterbank in which the peak frequencies are almost the same as those of the cGC and dcGC filterbanks. The synthesis LP-AF is not required in this case. The original speech wave is shown in Fig. 9(a); the resynthesized speech from the linear cGC and dcGC filterbanks are shown in Fig. 9(b) and (c), respectively. These sounds are normalized to the rms value of the whole signal. The resynthesized cGC wave [Fig. 9(b)] is essentially the same as the original [Fig. 9(a)]. It is clear that the peak factor of the resynthesized dcGC wave [Fig. 9(c)] is reduced and the relative level of the plosives has been increased. The sound quality of the compressed speech is not quite as good as the original, but it has the advantage of sounding louder for a given rms value.

Fig. 10 shows the compression characteristics (input-output functions) for the linear cGC and dcGC filterbanks. The sound pressure level, in decibels, is derived from the rms value of a entire word. The average and standard deviation of the SPL were calculated from fifty word segments of speech in a phonetically-balanced Japanese database. The dashed line with error bars on the dotted diagonal is for the analysis/synthesis signal produced with the linear cGC filterbank. The solid line with error bars is for speech compressed by the dcGC filterbank; the output level is set to 100-dB SPL for an input level of 90-dB SPL. The solid line with circles shows the compression characteristic for the forward-masking condition where the half-life is 0.5 ms, as shown in Fig. 6. The linear analysis/synthesis signal has variability because the filterbank restricts the passband between about 100 and 6000 Hz and, thus, the low- and high-frequency components drop off. The variability of the compressed speech is less than about 2 dB.

The slope of the input/output (I/O) function is about 0.6 dB/dB which is greater than that for the masking of short probe tones, where it is about 0.2 dB/dB at minimum. This moderate slope is reasonable for speech signals because speech consists of a range of frequency components which interact with each other; at one moment a component acts like a suppressor and at another it acts like a suppressee. This is an important observation for the design of compressors like those in hearing aids because the degree of compression is different for the simple tone sounds used to define the compression, and the speech sounds that the user wants to hear.

The compression of the dcGC filterbank is reminiscent of the compression in the much simpler wide dynamic range compression (WDRC) hearing aids [18]. However, both of these compression processes have a serious drawback. When there is background noise or concurrent speech, small noise components are effectively enhanced, and they interfere with

Europe PMC Funders Author Manuscripts

the speech components. It will be essential to introduce noise reduction [28] and speech segregation (e.g., [1]) in future speech processors. The analysis/synthesis, dcGC filterbank provides a framework for the design and testing of advanced auditory signal processors of this sort.

IV. Conclusion

We have developed a dynamic version of the compressive gammachirp filter with separate paths for level-estimation and signal processing. We have also developed a complete, analysis/synthesis filterbank based on the dynamic, compressive gammachirp auditory filter. We have demonstrated that the filterbank can simulate the asymmetric growth of two-tone suppression and the compression observed in nonsimultaneous masking experiments. The dcGC filterbank provides a framework for the development of signal processing algorithms within a nonlinear analysis/synthesis auditory filterbank. The system enables one to manipulate peripheral representations of sounds and resynthesize the corresponding sounds properly. Thus, it provides an important alternative to the conventional STFTs and linear auditory filterbanks commonly used in audio signal processing. The new analysis/synthesis framework can readily inherit refined signal processing algorithms developed previously in the linear domain. This framework should be useful for various applications such as speech enhancement and segregation [1]-[6], [28], speech coding [7]-[11], and hearing aids [18].

Acknowledgments

This work was supported in part by a project grant from the Faculty of Systems Engineering of Wakayama University, in part by the Japan Society of the Promotion of Science under Grant-in-Aid for Scientific Research (B) (2),15300061, 18300060, and in part by the U.K. Medical Research Council under Grant G9900369, Grant G9901257, and Grant G0500221. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gerald Schuller.

Biography

Toshio Irino (SM’04) was born in Yokohama, Japan, in 1960. He received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1982, 1984, and 1987, respectively.

From 1987 to 1997, he was a Research Scientist at NTT Basic Research Laboratories, Tokyo Japan. From 1993 to 1994, he was a Visiting Researcher at the Medical Research Council—Applied Psychology Unit (MRC-APU, currently CBU), Cambridge, U.K. From 1997 to 2000, he was a Senior Researcher in ATR Human Information Processing Research Laboratories (ATR HIP). From 2000 to 2002, he was a Senior Research Scientist in NTT Communication Science Laboratories. Since 2002, he has been a Professor of the Faculty of Systems Engineering, Wakayama University, Wakayama, Japan. He is also a Visiting

Professor at the Institute of Statistical Mathematics. The focus of his current research is a computational theory of the auditory system.

Dr. Irino is a member of the Acoustical Society of America (ASA), the Acoustical Society of Japan (ASJ), and the Institute of Electronics, Information and Communication Engineers (IEICE), Japan.

Roy D. Patterson was born in Boston, MA, on May 24, 1944. He received the B.A. degree from the University of Toronto, Toronto, ON, Canada, in 1967 and the Ph.D. degree in residue pitch perception from the University of California, San Diego, in 1971.

From 1975 to 1995, he was a Research Scientist for the U.K. Medical Research Council, at their Applied Psychology Unit, Cambridge, U.K., focusing on the measurement of frequency resolution in the human auditory system, and computational models of the auditory perception. He also designed and helped implement auditory warning systems for civil and military aircraft, railway maintenance equipment, the operating theaters and intensive care wards of hospitals, and most recently, fire stations of the London Fire Brigade. Since 1996, he has been the Head of the Centre for the Neural Basis of Hearing, Department of Physiology, Development, and Neuroscience, University of Cambridge, Cambridge, U.K. The focus of his current research is an “Auditory Image Model” of auditory perception and how it can be used to: 1) normalize communication sounds for glottal pulse rate and vocal tract length and 2) produce a size-invariant representation of the message in communication sounds at the syllable level. He has published over 100 articles in *JASA* and other international journals.

Dr. Patterson is a Fellow of the Acoustical Society of America.

REFERENCES

[1]. Divenyi, P., editor. Speech Separation by Humans and Machines. Norwell, MA: Kluwer; 2004.

[2]. Brown GJ, Cooke MP. Computational auditory scene analysis. Comput. Speech Lang. 1994; 8:297–336.

[3]. Slaney M, Naar D, Lyon RF. Auditory model inversion for sound separation. Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP). 1994; II:77–80.

[4]. Ellis, DPW. Cambridge: Dept. Elec. Eng Comp. Sci., Mass. Inst. Technol.; 1996. Prediction-driven computational auditory scene analysis. Ph.D. dissertation

[5]. Wang DL, Brown GJ. Separation of speech from interfering sounds based on oscillatory correlation. IEEE Trans. Neural Netw. May; 1999 10(3):684–697. [PubMed: 18252568]

[6]. Irino T, Patterson RD, Kawahara H. Speech segregation using an auditory vocoder with event-synchronous enhancements. IEEE Trans. Audio, Speech, Lang. Process. Nov.2006 14(6):2212–2221.

[7]. ISO/IEC JTC1/SC29, Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1,5 Mbit/s—Part 3: Audio, ISO/IEC 11172-3, Int. Std. Org. Geneva, Switzerland, 1993.

[8]. ISO/IEC JTC1/SC29, Generic Coding of Moving Pictures and Associated Audio Information—Part 7: Advanced Audio Coding (AAC), ISO/IEC 13 818-7, Int. Std. Org. Geneva, Switzerland, 2004.

[9]. Painter T, Spanias A. Perceptual coding of digital audio. Proc. IEEE. Apr.2000 88(4):451–513.

[10]. Baumgarte F. Improved audio coding using a psychoacoustic model based on a cochlear filter bank. IEEE Trans. Speech Audio Process. Oct.2002 10(7):495–503.

[11]. Baumgarte F. Application of a physiological ear model to irrelevance reduction in audio coding. Proc. AES 17th Int. Conf. High Quality Audio Coding. 1999:171–181.

[12]. Lutfi RA, Patterson RD. On the growth of masking asymmetry with stimulus intensity. J. Acoust. Soc. Amer. 1984; 76(3):739–745. [PubMed: 6491046]

[13]. Glasberg BR, Moore BCJ. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 1990; 47:103–138. [PubMed: 2228789]

[14]. Moore, BCJ. An Introduction of the Psychology of Hearing. 5th ed.. Oxford, U.K.: Academic; 2003.

[15]. Oxenham AJ, Plack CJ. A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired listening. J. Acoust. Soc. Amer. 1997; 101:3666–3675. [PubMed: 9193054]

[16]. Plack CJ, Oxenham AJ, Drga V. Linear and nonlinear processes in temporal masking. Acta Acust. 2002; 88:348–358.

[17]. Plack, CJ. The Sense of Hearing. London, U.K.: Lawrence Erlbaum Associates; 2005.

[18]. Dillon, H. Hearing Aids. New York: Thieme Medical Publishers; 2001.

[19]. Zwicker, E.; Fastl, H. Psychoacoustics—Facts and Models—. New York: Springer-Verlag; 1990.

[20]. Giguère C, Woodland PC. A computational model of the auditory periphery for speech and hearing research. I. Ascending path. J. Acoust. Soc. Amer. 1994; 95:331–342. [PubMed: 8120244]

[21]. Meddis R, O'Mard LP, Lopez-Poveda EA. A computational algorithm for computing nonlinear auditory frequency selectivity. J. Acoust. Soc. Amer. 2001; 109:2852–2861. [PubMed: 11425128]

[22]. Irino T, Unoki M. An analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp. J. Acoust. Soc. Japan (E). 1999; 20(6):397–406.

[23]. Combes, JM.; Grossmann, A.; Tchamitchian, P. Wavelets. Berlin, Germany: Springer-Verlag; 1989.

[24]. Rabiner, LR.; Schafer, RW. Digital Processing of Speech Signals. Englewood Cliffs, NJ: Prentice-Hall; 1978.

[25]. Yang T, Wang K, Shamma S. Auditory representations of acoustic signals. IEEE Trans. Inf. Theory. Mar.1992 38(2):824–839.

[26]. Irino T, Kawahara H. Signal reconstruction from modified auditory wavelet transform. IEEE Trans. Signal Process. Dec.1993 41(12):3549–3554.

[27]. Slaney, M. Proc. IEEE Systems, Man, Cybernetics Conf. Canada: Vancouver, BC; 1995. Pattern playback from 1950 to 1995; p. 3519-3524.

[28]. Lim JS. Speech enhancement. Proc. ICASSP. 1986:3135–3142.

[29]. Irino T, Patterson RD. A time-domain, level-dependent auditory filter: the gammachirp. J. Acoust. Soc. Amer. 1997; 101(1):412–419.

[30]. Irino T, Patterson RD. A compressive gammachirp auditory filter for both physiological and psychophysical data. J. Acoust. Soc. Amer. 2001; 109(5):2008–2022. [PubMed: 11386554]

[31]. Patterson RD, Unoki M, Irino T. Extending the domain of center frequencies for the compressive gammachirp auditory filter. J. Acoust. Soc. Amer. 2003; 114:1529–1542. [PubMed: 14514206]

[32]. Patterson RD, Allerhand M, Giguere C. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Amer. 1995; 98:1890–1894. [PubMed: 7593913]

[33]. Unoki M, Irino T, Patterson RD. Improvement of an IIR asymmetric compensation gammachirp filter. Acoust. Sci. Tech. 2001; 22(6):426–430.

[34]. Houtgast T. Psychophysical evidence for lateral inhibition in hearing. J. Acoust. Soc. Amer. 1972; 51:1885–1894. [PubMed: 4339849]

[35]. Duifhuis H. Level effects in psychophysical two-tone suppression. J. Acoust. Soc. Amer. 1980; 67:914–927. [PubMed: 7358916]

[36]. Slaney, M. Apple Computer Technical Rep. #35. 1993. An Efficient Implementation of the Patterson-Holdsworth Auditory Filterbank.

Europe PMC Funders Author Manuscripts

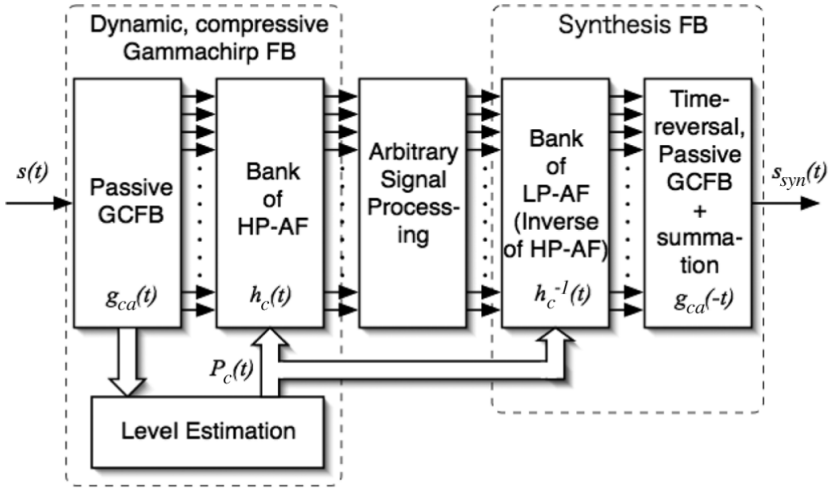

Fig. 1.

Block diagram of an analysis/synthesis filterbank based on the dynamic, compressive gammachirp auditory filter. The first two blocks produce a peripheral representation of sound whose features can be manipulated with standard signal processing algorithms. Then, the sound can be resynthesized to evaluate its quality.

Europe PMC Funders Author Manuscripts

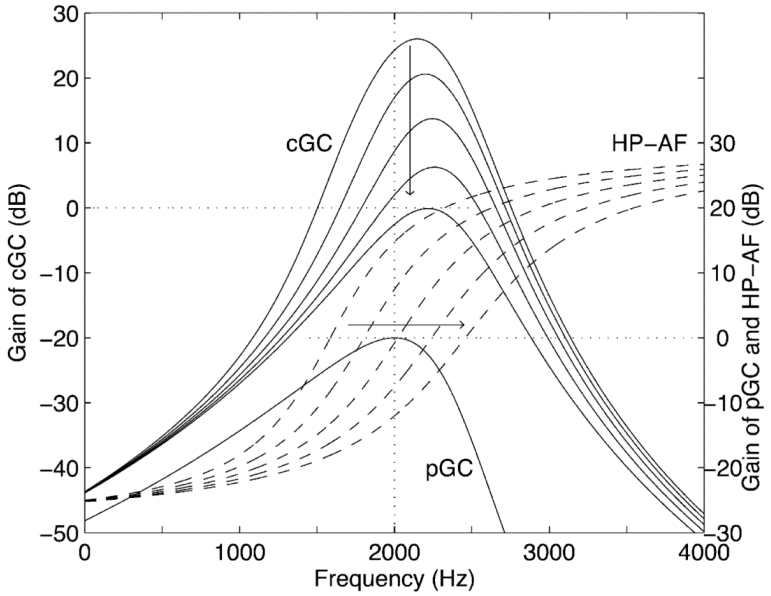

Fig. 2.

Set of compressive gammachirp filters (cGC, with peak frequency f_{p2}) which are constructed from one passive gammachirp filter (pGC, with peak frequency f_{p1}) and a high-pass asymmetric function (HP-AF) whose center frequency f_{r2} shifts up as stimulus level increases, as indicated by the horizontal arrow [30]. The gain of the cGC filter reduces as level increases, as indicated by the vertical arrow. The five filter shapes were calculated for probe levels of 30, 40, 50, 60, and 70 dB using the parameter values listed in the second row of Table I.

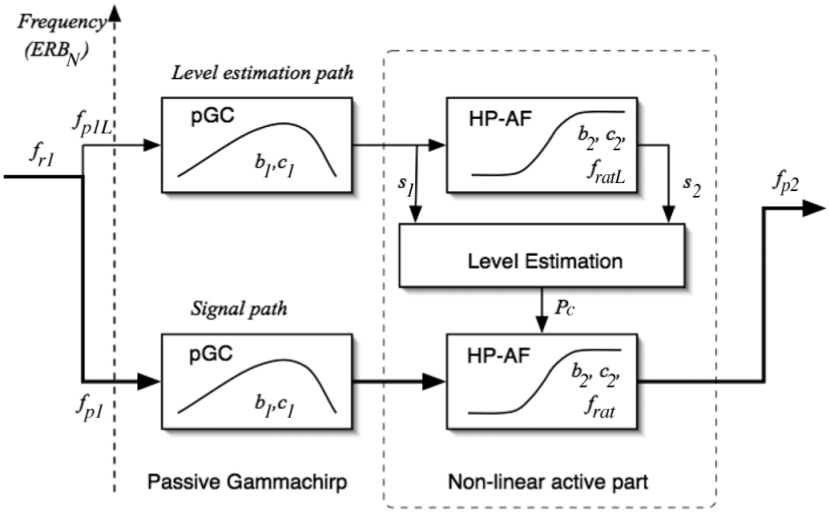

Fig. 3.

Block diagram of the dcGC filter illustrating how the pGC and HP-AF in a higher frequency channel (f_{p1L}) are used to estimate the level for the HP-AF in the signal path of the dcGC filter with channel frequency f_{p1}.

Europe PMC Funders Author Manuscripts

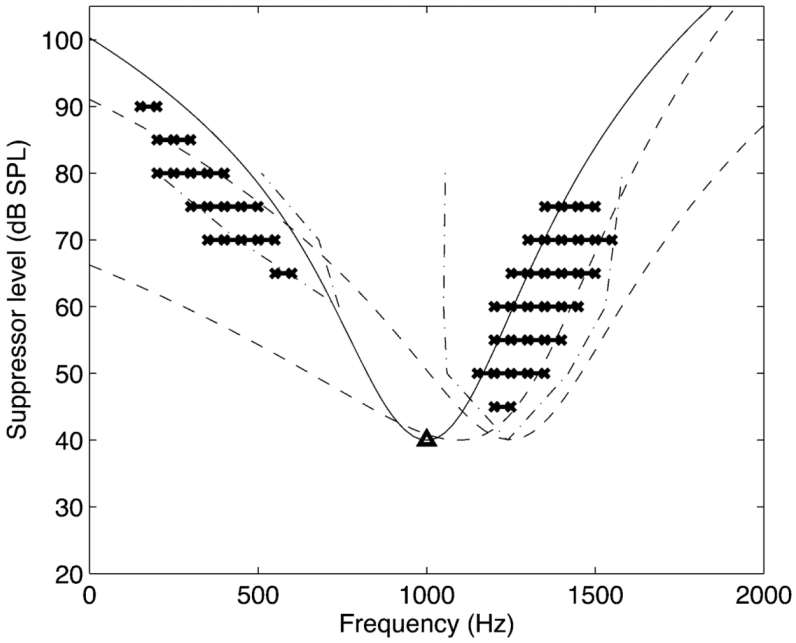

Fig. 4.

Simulation of two-tone suppression data. The probe tone is shown by the triangle. The suppression regions are shown with crosses. The dashed and dotted lines show the suppression regions observed psychophysically with the pulsation threshold technique [34]. The solid curve shows the filter shape of the cGC for the probe tone on its own. The dashed curves show the inverted frequency response curves of the pGC and cGC in the level estimation path, respectively.

Europe PMC Funders Author Manuscripts

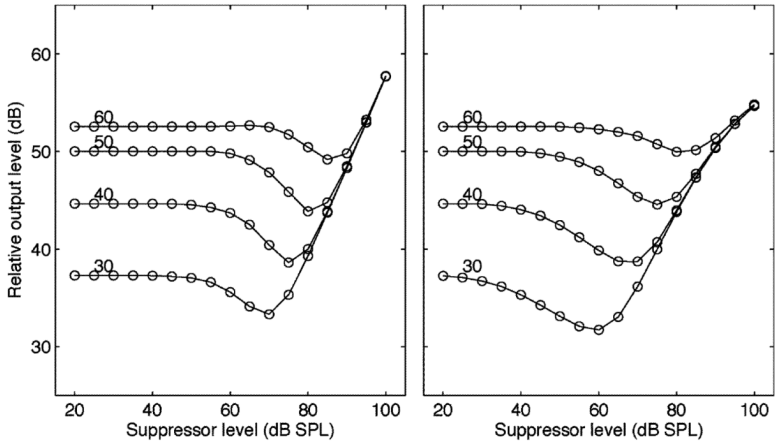

Fig. 5.

Relative level of the output of the dcGC for a 1000-Hz probe tone, as a function of suppressor level, when the suppressor frequency is either 400 Hz (left panel) or 1400 Hz (right panel). The numbers in the left-hand side show the probe level in decibels SPL. The output level is normalized to 50-dB SPL by shifting a constant decibel value. There is suppression whenever the probe level drops below its starting value where the suppressor is 20-dB SPL.

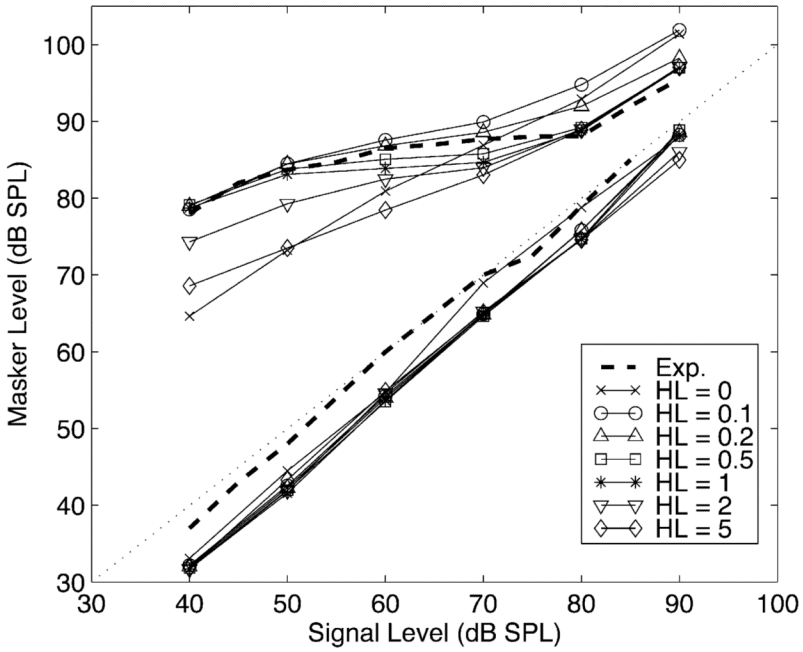

Fig. 6.
Compression data from [15] (thick dashed lines) and simulations of the data with dcGC filters in which the half-life for level estimation varies from 0 to 5 ms (thin solid lines).

Europe PMC Funders Author Manuscripts

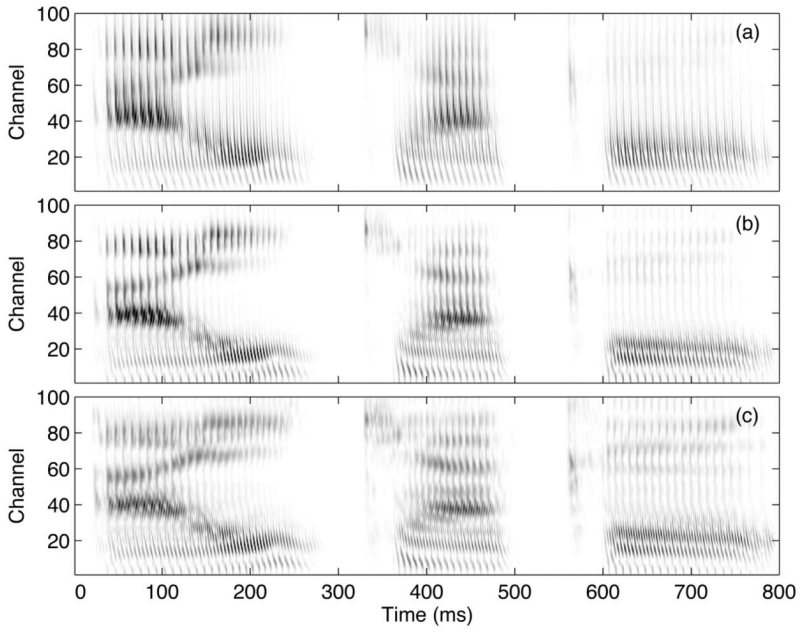

Fig. 7.
"Cochlear" spectrograms, or cochleograms, for the Japanese word "aikyaku," plotted on a linear scale to reveal level differences. (a) pGC filter. (b) Linear cGC filter. (c) dcGC filter.

Europe PMC Funders Author Manuscripts

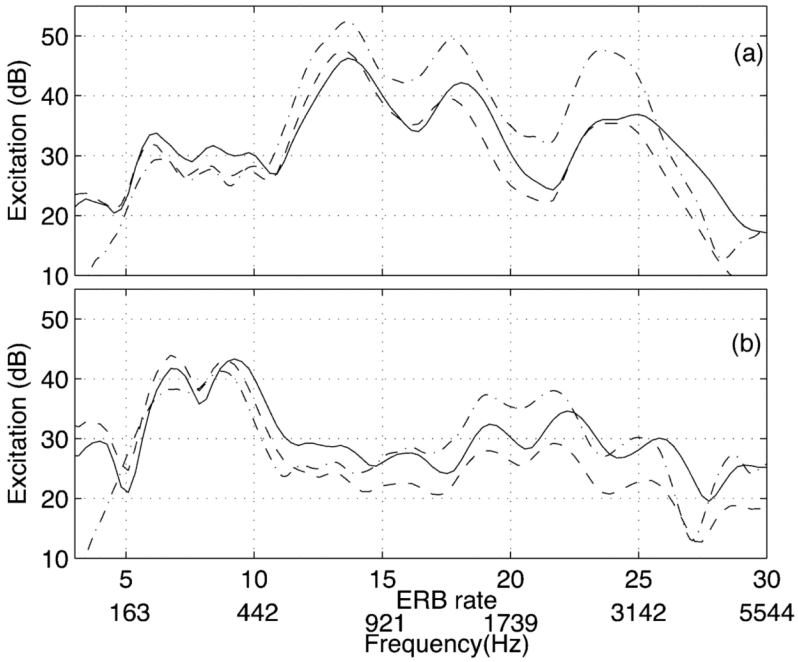

Fig. 8.

Excitation patterns calculated from the dcGC filterbank (solid line) and a linear cGC filterbank (dashed line). The time is (a) 60 ms and (b) 630 ms. A rectangular window with 1024 points was used for averaging the filter output. The dashed and dotted curve is a level-dependent excitation pattern derived with a roex filterbank [13].

Europe PMC Funders Author Manuscripts

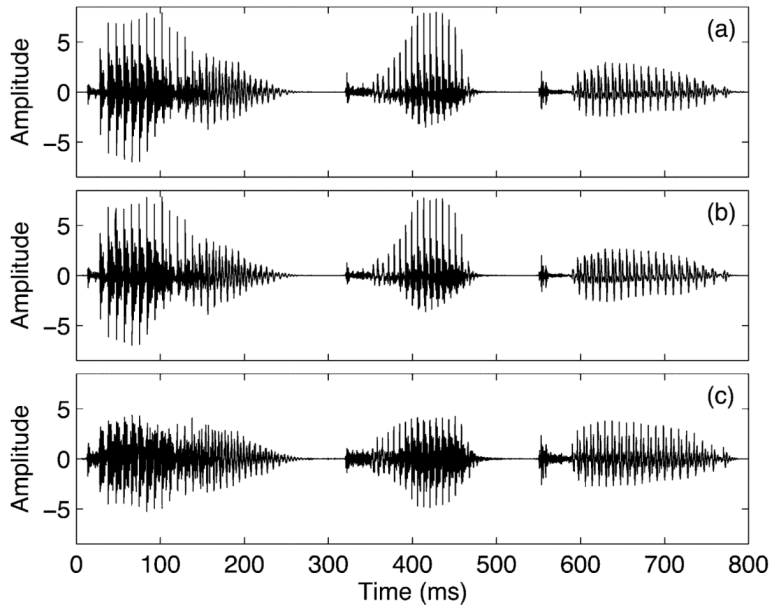

Fig. 9.

(a) Original speech wave. (b) Resynthesized versions from the linear cGC analysis/synthesis filterbank. (c) dcGC analysis filterbank with the linear pGC synthesis filterbank.

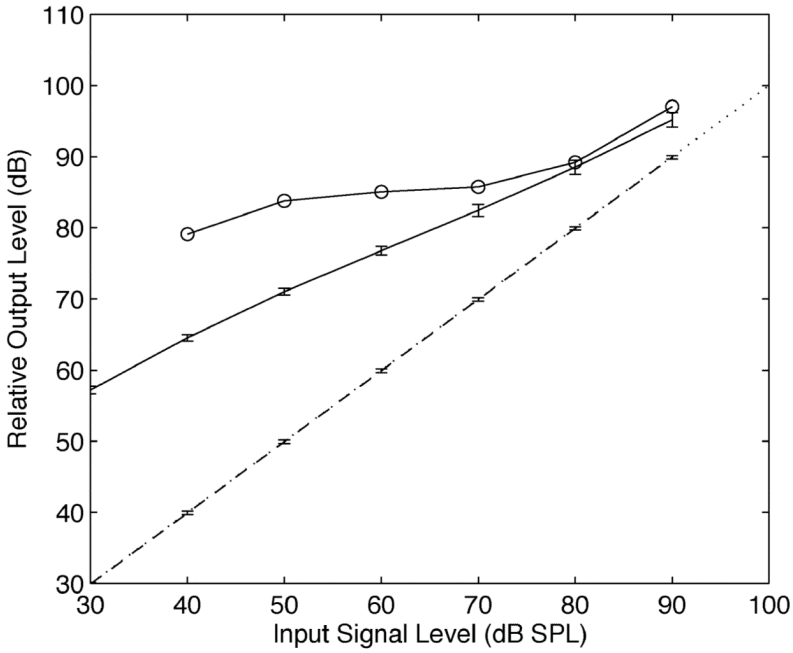

Fig. 10.

Compression characteristics (input-output functions) of the resynthesized speech sounds. The solid line with error bars shows the compressed speech from the dcGC filterbank; the dashed line with error bars shows the analysis/synthesis signal from the linear cGC filterbank; the solid line with circles shows the compression characteristic for the forward-masking condition where the half life is 1 ms, as shown in Fig. 5.

Europe PMC Funders Author Manuscripts

TABLE I

Coefficient Values for the Compressive Gammachirp Filter in Patterson *et al.* [31] and the Current Study. P_{gcp}, P_c, and P_{RL} are in Decibels. τ_L is in Milliseconds and was Varied Between 0 and 5 ms When We Calculated the Effect of the Half-Life in Section III-B

Patterson et al. [31]	n_1	b_1	c_1	f_{rat}		b_2	c_2
	4	1.81	-2.96	0.466+0.0109 P_{gcp}		2.17	2.20
Current study	n_1	b_1	c_1	f_{rat}		b_2	c_2
	4	1.81	-2.96	0.466+0.0109 P_c		2.17	2.20
	r_{EL}	f_{ratL}	τ_L	w_L	v_{1L}	v_{2L}	P_{RL}
	1.5	1.08	0.5	0.5	1.5	0.5	50