DYNAMIC, COMPRESSIVE GAMMACHIRP AUDITORY FILTERBANK FOR PERCEPTUAL SIGNAL PROCESSING

Toshio Irino [*] *and Roy D. Patterson* [+]

[*] Faculty of Systems Engineering, Wakayama University, Japan. irino@sys.wakayama-u.ac.jp
[+] Centre for the Neural Basis of Hearing, Cambridge Univ., U.K. roy.patterson@mrc-cbu.cam.ac.uk

ABSTRACT

A gammachirp auditory filter was developed 1) to extend the domain of the gammatone auditory filter, 2) to simulate the changes in filter shape that occur with changes in stimulus level, 3) to explain a large body of simultaneous masking data, 4) to explain the compressive characteristics of the auditory filter system, and 5) to facilitate the development of a nonlinear, analysis/synthesis framework. What remains is to specify the dynamics of how the stimulus level controls the filter parameters. In this paper, we use psychophysical data involving compression to derive the details of the level control circuit for the dynamic version of the cGC (dcGC) filter and filterbank. The dcGC filterbank enhances spectral contrasts and reduces the dynamic range. This property with the analysis/synthesis framework should be useful in various forms of perceptual signal processing.

1. INTRODUCTION

It is now common to use psychophysical and physiological knowledge about the auditory system to optimize audio signal processors. For example, popular audio coders (e.g. MP3 and AAC) use knowledge of auditory masking in human listeners to provide 'perceptual coding' that matches the coding resolution to the limits of human perception[1]. It is also the case for signal processors to enhance and segregate target speech sounds presented in noise[2]. Some of the processors are already based on linear auditory filterbanks but the performance is not necessarily better than with processors based on the short-time Fourier transform (STFT). It appears that further progress may require the introduction of a level-dependent, asymmetric auditory filter [3] and a nonlinear auditory filterbank [e.g.,4]. This enable the model to explain perceptual data such as simultaneous masking and compression. The use of a nonlinear filterbank raises another problem, however; there is no general method for resynthesis of data from coders with nonlinear filterbanks. So, although there are a number of dynamic, nonlinear cochlear models based on transmission-line systems and filterbanks [e.g.,4,5], none of them supports the analysis/synthesis framework of the vocoder community.

Recently, we developed such a filterbank based on the gamma*chirp* auditory filter [6,7] which was developed to extend the domain of the gamma*tone* auditory filter. It provides a realistic auditory filterbank for models of auditory perception and it facilitates the development of a nonlinear, analysis/synthesis system. The resulting compressive gammachirp filter (cGC) was fitted to a large body of simultaneous masking data obtained psychophysically [8]. 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 [9] in which an IIR, asymmetric compensation filter (AF) was defined to simulate the asymmetric function. The filter is minimum phase and, thus, invertible. 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.

Thus, all that is actually required is to extend the static version of the compressive gammachirp (cGC) filter into a dynamic, level-dependent filter that can accommodate the nonlinear behavior of the cochlea. In this paper, we show how a new level control circuit designed for the dynamic version of the cGC filter (i.e., the dcGC filter) can help the filterbank explain highly non-linear psychophysical data involving compression. We then go on to describe an analysis/synthesis filterbank based on the cGC that can be used for speech processing.

2. A DYNAMIC COMPRESSIVE GAMMACHIRP FILTER

Figure 1 is a block diagram of the proposed gammachirp analysis/synthesis filterbank. There are a set of linear passive gammachirp filters and a set of asymmetric compensation filters both for analysis and synthesis. Between the analysis and synthesis stages, it is possible to include an arbitrary signal processing algorithm.

2.1. The compressive gammachirp filter function

The complex form of the gammachirp auditory filter is

$$g_c(t) = a t^{n_1-1} \exp(-2\pi b_1 \mathrm{ERB}_N(f_{r1})\, t) \exp(j 2\pi f_{r1} t + j c_1 \ln t + j\phi_1) \quad (1)$$

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 the asymptotic center frequency; $\mathrm{ERB}_N(f_{r1})$ is the equivalent rectangular bandwidth for average normal hearing subjects [10]; ϕ_1 is the initial phase; and $\ln t$ is the natural logarithm of time. Time is restricted to positive values. When $c_1 = 0$, Eq. 1 reduces to the complex impulse response of the gammatone filter.

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

$$|G_C(f)| = a_\Gamma \cdot |G_T(f)| \cdot \exp(c_1 \theta_1(f)), \quad (2)$$

$$\theta_1(f) = \arctan\{(f - f_{r1})/b_1 \mathrm{ERB}_N(f_{r1})\}. \quad (3)$$

$|G_T(f)|$ 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 anti-symmetric function centered at the asymptotic frequency, f_{r1}. a_Γ is a constant.

Irino and Patterson [7] 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 associated with the outer hair cells. The resulting 'compressive' gammachirp filter, $|G_{cc}(f)|$, is

$$|G_{cc}(f)| = \{a_r |G_T(f)| \cdot \exp(c_1\theta_1(f))\} \cdot \exp(c_2\theta_2(f)) = |G_{cp}(f)| \cdot \exp(c_2\theta_2(f)) \quad (4)$$

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 represents 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.

Figure 2 illustrates how a level-dependent set of compressive gammachirp filters (cGC; upper set of 5 solid lines; left ordinate) can be produced by cascading a fixed passive gammachirp filter (pGC; lower solid line; right ordinate) with a set of highpass asymmetric functions (HP-AF; set of 5 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 [7]. 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 Eq. 3,

$$\theta_1(f) = \arctan\{(f - f_{r1})/b_1 \mathrm{ERB}_N(f_{r1})\} \text{ and}$$
$$\theta_2(f) = \arctan\{(f - f_{r2})/b_1 \mathrm{ERB}_N(f_{r2})\}. \quad (5)$$

The peak frequency, f_{p1}, of pGC , is

$$f_{p1} = f_{r1} + c_1 b_1 \mathrm{ERB}_N(f_{r1})/n_1, \quad (6)$$

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

$$f_{r2} = f_{rat} \cdot f_{p1}. \quad (7)$$

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 [7,8]. The total level of the probe plus the two masker bands at the output of the passive GC, P_{gcp}, was used to control the position of the HP-AF. Specifically,

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

The superscripts 0 and 1 designate the intercept and slope of the line. A detailed description of the steps in the fitting procedure is presented in Appendix B of [8].

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 parameters were each represented by a single coefficient, except for f_{rat} which was represented by two

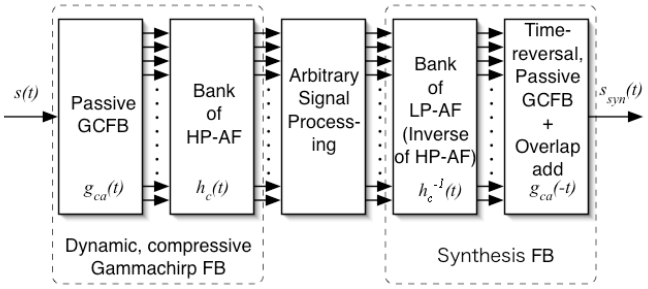

Figure 1. Block diagram of analysis/synthesis filterbank based on dynamic, compressive gammachirp (dcGC).

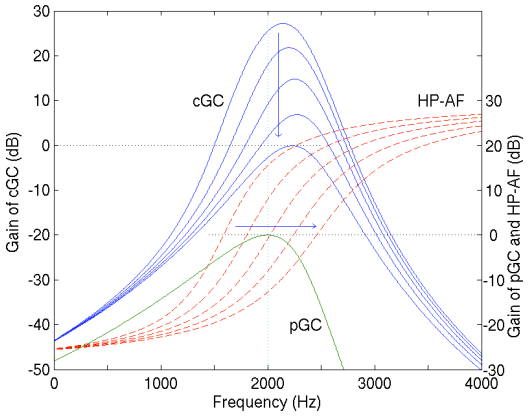

Figure 2. Compressive gammachirp(cGC) as combination of passive gammachirp (pGC) and high-pass asymmetric function (HP-AF). The arrows shows the level-dependence from low to high SPL.

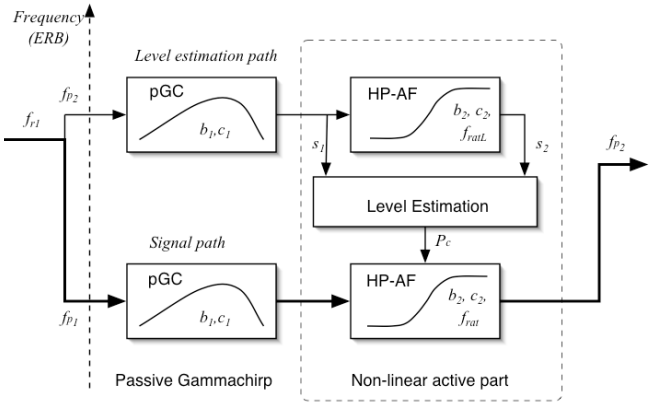

Figure 3. Block diagram of the dynamic, compressive gammachirp filter (dcGC).

Table I

Coefficient values for the compressive gammachirp filter. P_{gcp}, P_c, and P_{RL} are in dB. τ_L is in ms.

n_1	b_1	c_1	f_{rat}		b_2	c_2
4	1.81	-2.96	0.466+0.109 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

coefficients. 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 [8]). As a result, it is possible to construct the gammachirp filterbank with just six coefficients [8]; for

humans, the coefficient values are as listed in the second row in Table I.

2.2. IIR implementation of the filter

The asymmetric function (AF) was implemented as IIR filters [9]. Since this is a minimum phase filter, it is possible to define the inverse filter (see the fourth block of Fig. 1). Using the inverse filterbank, linear representations are recovered from nonlinear, the compressive representation to resynthesize sounds almost identical to the input sounds when there is no modification between the analysis and synthesis filterbanks. See [9] for details of the IIR implementation and the analysis/synthesis procedure.

2.3. Filter architecture for dynamics

Figure 3 shows the architecture of the dcGC filter. As in the previous compressive gammachirp [7], 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, $f_{r2}\,(=f_{rat}\cdot f_{p1})$. This combination of pGC and HP-AF results in the compressive gammachirp (cGC) defined in Eq. 4 with peak frequency f_{p2}. The parameter values are the same as in the previous study and are listed in the second 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, $f_{r2L}\,(=f_{ratL}\cdot f_{p1L})$. 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

$$ERB_N rate(f_{p1L}) = ERB_N rate(f_{p1}) + r_{EL}, \quad (9)$$

where $ERB_N rate(f)$ is the ERB rate at frequency f [10] and r_{EL} is a parameter that represents the frequency separation between the two pGC filters on the ERB_N-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. The level, P_c, was estimated in dB 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:

$$\bar{s}_1(t) = \max\{\bar{s}_1(t-\Delta t)\cdot e^{-\ln 2\cdot(\Delta t/\tau_L)}, \max(s_1(t),0)\} \text{ and}$$
$$\bar{s}_2(t) = \max\{\bar{s}_2(t-\Delta t)\cdot e^{-\ln 2\cdot(\Delta t/\tau_L)}, \max(s_2(t),0)\}, \quad (10)$$

where Δt is the sampling time and τ_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 control level, $P_c(t)$, is calculated as a weighted sum of these linear levels in dB.

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

where $a_{RL} = 10^{P_{RL}/20}$, w_L, v_{1L} and v_{2L} are weighting parameters and P_{RL} is a parameter for the reference level in dB.

In the filterbank, the asymptotic frequencies, f_{r1}, of the pGC filters are uniformly spaced along the ERB scale. The peak frequencies, f_{p1}, of the pGC filters are also uniformly spaced and lower than the asymptotic frequencies, f_{r1}, since $c_1 < 0$ in Eq. 6. 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 dynamic, compressive gammachirp (dcGC) auditory filter. The parameter values in Table I were determined from the results of a simulation of two-tone suppression data [11].

3. EXPERIMENTS

3.1. Compression

Compressive nonlinearity is an important factor in cochlear models. Plack and Oxenham [12] measured the compression characteristics for humans using a forward masking paradigm. This section shows how the dcGC filter can also explain the compression data.

3.1.1. Method

The experiment [12] 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 Hz 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. Then, the NAP 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 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 dB. 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 listed in Table I.

3.1.2. Results

Figure 4 shows the experimental results [12] as thick dashed line. The simulation was preformed for seven half-lives ranging from 0 to 5 ms (Eq. 10), and the results are presented by thin solid lines. The solid lines that run parallel to, and just below, the dotted diagonal show simulated threshold when the probe and masker have the same frequency, namely, 6000 Hz. They are essentially linear input-output functions. The solid lines above the dotted diagonal show 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 threshold for the condition where the probe and masker have the same frequency is located a few dB below the dotted diagonal line. The lines are almost the same despite relatively large half-life differences. This is consistent with the psychophysical data, at least, for one subject [12]. 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.

3.2. Compression for speech processing

It is common to introduce compressive functions, such as the log of amplitude, into STFT processors and filterbank processors to improve performance. Many conventional auditory filterbanks also use an extra stage of instantaneous compression after filtering. It is important, however, to try and reduce the dynamic range without reducing the important spectral contrasts in the speech, and this is where dynamic compression has an advantage.

Figure 5a shows the neural activity pattern (NAP) of the dynamic, compressive gammachirp (dcGC) filterbank for a syllable /ka/. The plosive starts at around 17 ms and there is activity above channel 40 (a center frequency of about 1500 Hz). The vowel starts around 67 ms and it has distinctive formant resonances centered around channels 40, 55, and 75 (center frequencies of about 800, 1500, and 2700 Hz). Figure 5b shows the NAP of the fixed-level version of the compressive gammachirp filterbank (cGC). It is a linear filterbank since the compression is not dynamic. The activity in the plosive region is much less than in the vowel region. Moreover, the second and third formants are weaker than for the output of the dcGC filterbank. Figure 5c shows the NAP after applying instantaneous compression to the output of a fixed-level, linear cGC filterbank with a power function having an exponent of 0.5 (relatively mild compression). It is clear that the activity in the plosive is widely spread in frequency and the resonance of the second formant is not clearly resolved. This is because uniform compression reduces spectral contrasts indiscriminately. In contrast, the dcGC filterbank automatically enhances the formant structure and reduces the dynamic range.

4. CONCLUSIONS

We have developed a dynamic version of the compressive gammachirp (dcGC) filter and filterbank. The dcGC filterbank effectively enhances the spectral contrasts and reduces the dynamic range. Together with the analysis/synthesis framework, this property is important for manipulating peripheral representations of sounds and resynthesize the corresponding sounds properly. The dcGC should be useful for various applications such as perceptual coding, speech enhancement and segregation, and hearing aids.

Acknowledgments This work was partially supported by a grant from the faculty of systems engineering of Wakayama University, by Grant-in-Aid for Scientific Research (B)(2),15300061 of JSPS, and by the UK Medical Research Council (G0500221, G9900369).

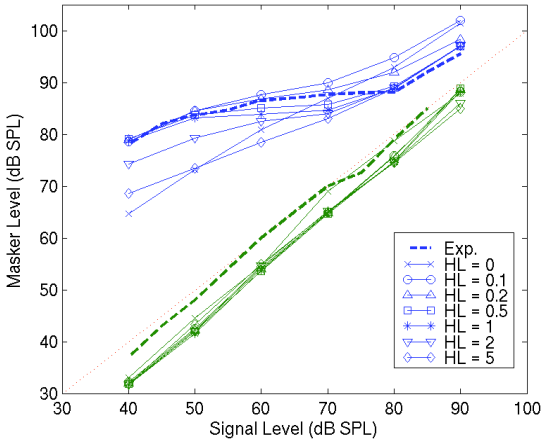

Figure 4. Compression data from [12] (thick dashed lines) and simulations of the data with dcGC filters in which the half-life for level estimation varies from 0 to 5 ms (thin solid lines).

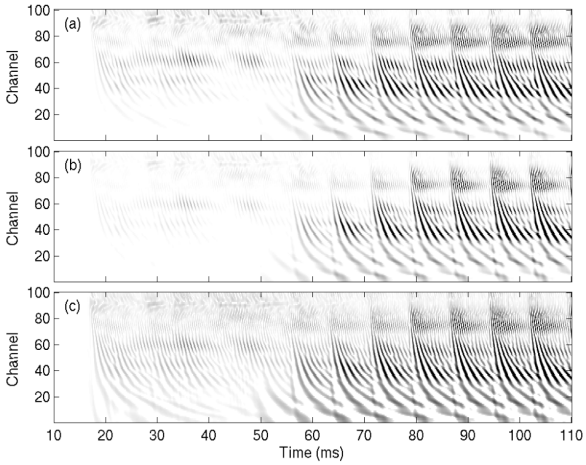

Figure 5. Simulated neural activity pattern (NAP) for a syllable /ka/ using dynamic, cGC filterbank (a) , fixed-level, linear cGC filterbank (b), and power compressed version of linear cGC filterbank (c).

REFERENCES

[1] Painter, T. and Spanias, A., "Perceptual coding of digital audio," Proc. IEEE, 88, pp.451-513, 2000.
[2] Divenyi, P., (Ed). "Speech separation by humans and machines," Kluwer academic publisher, Norwell, USA, 2004.
[3] Lutfi, R.A. and Patterson, R.D., "On the growth of masking asymmetry with stimulus intensity," J. Acoust. Soc. Am. , 76(3), pp.739-745, 1984.
[4] Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A., "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. 109, 2852–2861, 2001.
[5] Zwicker, E. and Fastl, H. (1990), "Psychoacoustics - Facts and Models -," Springer-Verlag, New York.
[6] Irino, T. and Patterson, R.D., "A time-domain, level-dependent auditory filter: the gammachirp," J. Acoust. Soc. Am. , 101 (1), pp.412-419, 1997.
[7] Irino, T. and Patterson, R.D., "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. , 109 (5), pp.2008-2022, 2001.
[8] Patterson, R.D. , Unoki, M. , and Irino, T. "Extending the domain of center frequencies for the compressive gammachirp auditory filter," J. Acoust. Soc. Am. 114 , 1529-1542, 2003.
[9] Irino, T. and Unoki, M. , "An analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp," J. Acoust. Soc. Japan (E) , 20 (6), 397-406, 1999.
[10] Moore B.C.J.,"An Introduction to the Psychology of Hearing (fifth edition)," Academic Press, London, 2003.
[11] Irino, T. and Patterson, R.D., "A dynamic, compressive gammachirp auditory," submitted to IEEE Trans. SAP, 2005.
[12] Oxenham, A. J. , and Plack, C. J. (1997). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired listening," J. Acoust. Soc. Am. 101, 3666–3675.