

A time-domain, level-dependent auditory filter: The gammachirp

Toshio Irino^{a)}

NTT Basic Research Laboratories, 3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa, 243-01 Japan

Roy D. Patterson^{b)}

MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

(Received 28 June 1996; accepted for publication 4 September 1996)

A frequency-modulation term has been added to the gammatone auditory filter to produce a filter with an asymmetric amplitude spectrum. When the degree of asymmetry in this “gammachirp” auditory filter is associated with stimulus level, the gammachirp is found to provide an excellent fit to 12 sets of notched-noise masking data from three different studies. The gammachirp has a well-defined impulse response, unlike the conventional roex auditory filter, and so it is an excellent candidate for an asymmetric, level-dependent auditory filterbank in time-domain models of auditory processing. © 1997 Acoustical Society of America. [S0001-4966(97)02701-X]

PACS numbers: 43.66.Ba, 43.66.Dc [WJ]

INTRODUCTION

In time-domain auditory models, the spectral analysis performed by the basilar membrane is often simulated by a bank of gammatone auditory filters (see, for example, Patterson *et al.*, 1995). The impulse response of the gammatone is

$$g_i(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_c)t) \cos(2\pi f_c t + \phi) \quad (t > 0), \quad (1)$$

where a , b , n , f_c , and ϕ are parameters. $\text{ERB}(f_c)$ is the equivalent rectangular bandwidth of the filter, and at moderate levels $\text{ERB}(f_c) = 24.7 + 0.108f_c$ in Hz (Glasberg and Moore, 1990). The filter gets its name from the fact that the envelope formed by the power function and the exponential is a gamma distribution function, and the cosine carrier is a tone when it is in the auditory range. The amplitude spectrum of the gammatone filter is essentially symmetric on a linear frequency scale.

The gammatone function was introduced by Johannesma (1972) to characterize impulse-response data gathered physiologically from primary auditory fibers in the cat (see Carney and Yin, 1988, for an overview). The gammatone has also been used to characterize spectral analysis in humans at moderate levels where the amplitude characteristic of the auditory filter is nearly symmetric on a linear frequency scale (see Patterson 1994, for an overview).

The use of the gammatone filter is limited, however, by the repeated demonstration that, below its center frequency, the skirt of the auditory filter broadens substantially with increasing stimulus level, and above its center frequency the skirt sharpens a little with increasing level (Lutfi and Patterson, 1984; Patterson and Moore, 1986; Moore and Glasberg, 1987). The level dependence of the auditory filter has been modeled using the “roex” function (Patterson *et al.*, 1982; Glasberg and Moore, 1990; Rosen and Baker, 1994). But the roex auditory filter does not have a well-defined impulse

response which largely precludes its use in auditory filterbanks. More physiological models of cochlear mechanics (for example, Giguère and Woodland, 1994) do not provide good fits to human masking data; nor do they have sufficiently simple impulse responses for the traditional filterbank architecture.

Irino (1995, 1996) recently demonstrated that an analytic relative of the gammatone function, referred to as the “gammachirp” function, is a theoretically optimum auditory filter, in the sense that it leads to minimal uncertainty in a joint time and scale representation of auditory signal analysis. The derivation of the gammachirp function is based on operator methods (Gabor, 1946; Cohen, 1991, 1993) involving the Mellin transform (Titchmarsh, 1948); it is summarized in Appendix A. The gammachirp auditory filter is the real part of the analytic gammachirp function, Eq. (A20). It has an asymmetric amplitude characteristic, and in the following we show that, when the asymmetry is associated with stimulus level, the gammachirp filter provides an excellent fit to human masking data. The gammachirp has a well-defined impulse response and, with only one parameter more than the gammatone, it would appear to be an excellent candidate for an asymmetric, level-dependent auditory filterbank.

I. METHOD

A. The power spectrum model with a gammachirp filter

The impulse response of the gammachirp auditory filter is

$$g_c(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_r)t) \times \cos(2\pi f_r t + c \ln t + \phi) \quad (t > 0). \quad (2)$$

The only difference between it and the impulse response of the gammatone [Eq. (1)] is the term $c \ln t$; c is an additional parameter, and \ln is the natural logarithmic operator. The filter has a monotonically frequency-modulated carrier (a chirp) with an envelope that is a gamma distribution func-

^{a)}Electronic mail: irino@nttlab.brl.ntt.co.jp; WWW: <http://www.brl.ntt.co.jp/people/irino/index.html>

^{b)}Electronic mail: roy.patterson@mrc-apu.cam.ac.uk

tion, and hence the name “gammachirp.”¹ We use f_r instead of f_c for the frequency parameter because the peak frequency of the amplitude spectrum varies with c , and to a lesser extent, b and n . The equivalent rectangular bandwidth of the filter varies with stimulus level, but for convenience, we associate the parameter b with stimulus level so that the basic formula for filter width, $ERB(f_r) = 24.7 + 0.108f_r$, is the same as in Eq. (1).

The auditory filter shape is derived using the power spectrum model of masking (Fletcher, 1940; Patterson, 1976). In the experiment, the listener is required to detect a brief sinusoidal signal, referred to as a “probe” tone, in the presence of a masker which is a noise with a spectral notch in the frequency region of the probe tone. This “notched noise” has a constant spectrum level N_0 in a band below the tone between $f_{l_{\min}}$ and $f_{l_{\max}}$ and in a band above the tone between $f_{u_{\min}}$ and $f_{u_{\max}}$. The level of the probe tone is varied to determine the power required to make it just audible (probe “threshold”), as a function of the width of the notch in the noise. The details of the experiment and the criterion for threshold are described in Patterson (1976). If the “shape” of the auditory filter (that is, its power spectrum) is represented by the weighting function, $W(f)$, then the power spectrum model is

$$P_s = K + N_0 + 10 \log_{10} \left(\int_{f_{l_{\min}}}^{f_{l_{\max}}} W(f) df + \int_{f_{u_{\min}}}^{f_{u_{\max}}} W(f) df \right), \quad (3)$$

where P_s is the power of the probe tone at threshold in dB, and K is a constant which is related to the efficiency of the detection mechanism following the auditory filter. Following Patterson *et al.* (1982), a parameter r is introduced to limit the dynamic range of the filter. The weighting function is associated with the power spectrum of the gammachirp, $|G_C(f)|^2$, as follows:

$$W(f) = (1 - r) \cdot W_{om}(f) \cdot |G_C(f)|^2 + r. \quad (4)$$

Here, $W_{om}(f)$ is the “ELC” correction recommended by Glasberg and Moore (1990) to simulate the effects of the outer and middle ears. The maximum absolute magnitude of $W(f)$ is normalized to unity (See Appendix B for the analytic form of the amplitude spectrum of the gammachirp.)

B. Parameters and fitting procedure

We characterize the level dependence of the auditory filter shape in terms of the level dependence of the five parameters of the gammachirp: n , b , c , K , and r . The auditory filter becomes broader on the low side and sharper on the high side as stimulus level increases (Moore and Glasberg, 1987). Changes in the parameters n and b have little effect on the asymmetry of the amplitude spectrum, and K and r do not affect asymmetry since they are not filter parameters. Thus, the degree of asymmetry is primarily determined by c . Rosen and Baker (1994) showed, in an analogous fit with the roex auditory filter, that the level dependence can be summa-

rized with a linear function. In our initial fit, then, we provide two coefficients for c and one coefficient for each of n , b , and K .

The parameters n and b affect bandwidth reciprocally; the bandwidth of the filter decreases, either when n increases or when b decreases, and vice versa. We are mainly concerned with the filter shape around the center frequency and in this case we can fix either n or b and let the other vary to match the auditory filter shape. Preliminary simulations applied to several data sets and previous work with the gammatone suggested that we begin by fixing n at the value 4.

The fitting procedure is broadly similar to the PolyFit procedure of Rosen and Baker (1994). Thresholds were calculated for a range of filters with center frequencies around the probe frequency. The value of the filter giving the highest signal-to-noise ratio was chosen as the threshold estimate P_s (Patterson and Nimmo-Smith, 1980). We used the Levenberg–Marquardt method (Press *et al.*, 1988) to minimize the squared error between the data and P_s ; this is a standard procedure for a nonlinear least-mean-square problem. This fitting procedure is referred to as the “gammachirp fit” in the following.

C. Data sets

We began by applying the gammachirp fit to the notched-noise masking data of Rosen and Baker (1994) because the results can be compared directly to their results with the roex filter and the PolyFit procedure. We will specify the data source in the following by the initials of the subject and the probe frequency, for example, “LM at 2000 Hz” for this data set which contains 78 tone-in-noise thresholds. Rosen and Baker used total squared error in dB^2 to evaluate alternative fits. We will use rms (root-mean-squared) error in dB; it is a more intuitive measure and makes it easier to compare fits when the data sets have different numbers of thresholds.

The gammachirp was also fitted to subsets of the notched-noise data reported in Lutfi and Patterson (1984) and Moore *et al.* (1990). The data from Lutfi and Patterson are those of HM, RL, RM, and WW at 1000 and 4000 Hz. Each set contains 39 data points distributed over three different noise levels, except for RM at 4000 Hz, where there are 52 data points over four noise levels. The data of Moore *et al.* (1990) are those of CP at 200, 400, and 800 Hz. Each set contains 75 data points distributed over five noise levels.

II. RESULTS

A. Rosen and Baker (1994)

Rosen and Baker (1994) fitted a wide range of roex filter models to a set of masking data gathered with both probe-fixed and masker-fixed conditions to maximize the range of signal levels represented by the data. They discuss a subset of their fits for probe-dependent models with 24, 15, 10, 8, 7, 6, and 5 variable coefficients. The rms errors for the 78 thresholds are 1.15, 1.16, 1.19, 1.19, 1.19, 1.20, and 1.42 dB.² The rms errors for masker-dependent models are much greater than these values, and this is why they restricted their attention to probe-dependent models. The focus of their dis-

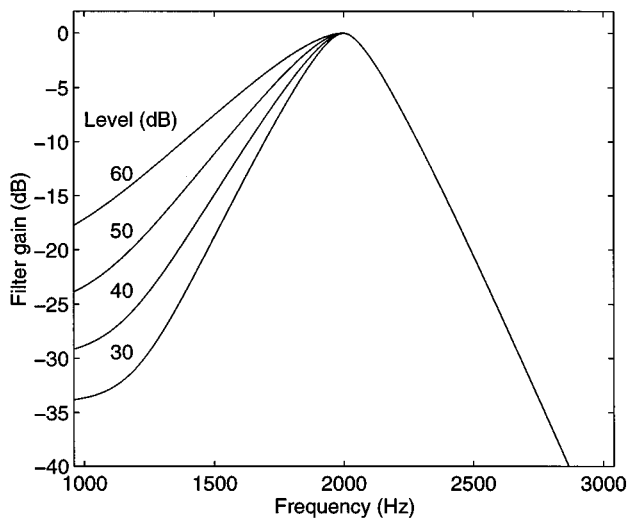


FIG. 1. The roex auditory filter shape as a function of probe level (30–60 dB SPL in 10-dB steps) with six variable coefficients [adapted from Rosen and Baker (1994)].

cussion is the trade-off between number of free parameters and goodness of fit, and they conclude that a roex(p, r) model with six coefficients is the most appropriate. Specifically, their fit employed one parameter for each of p_u and k , and two for each of p_l and r_l . The auditory filter shapes produced by this fit are shown in Fig. 1 as a function of probe level. The lower side of the filter becomes considerably broader as level increases; the upper side is invariant.

Column 7 of Table I shows rms error values (dB) obtained with the gammachirp fit using probe-dependent models with various numbers of coefficients. The integers in columns 2–6 show the number of coefficients used for the gammachirp parameter in that column. Following Rosen and Baker, we used the absolute threshold value (22.7 dB) to limit the minimum value of P_s . We also investigated

masker-dependent models, but found, like Rosen and Baker, that the rms errors were always much greater than those with the probe-dependent models. Consequently, we only consider probe-dependent models in what follows. The fit without the parameter c , i.e., the gammatone fit, is shown in the last row of Table I. The rms error is about 30% greater than that of the gammachirp fits, indicating that the gammatone filter is not suitable as an asymmetric, level-dependent filter.

The rms errors in column 7 are the same for gammachirp models with between 4 and 7 variable coefficients. This indicates that the coefficients converge even with relatively few coefficients. The rms errors with the gammachirp are greater than those for roex models with six, seven, and eight variable coefficients (compare columns 7 and 8), and smaller for models with five variable coefficients. The gammachirp model with four variable coefficients, where n is fixed to 4, produces the same rms error as the model with five variable coefficients, where the estimated value of n is 3.89. Accordingly, the model with four variable coefficients seems sufficient to explain the masking data. Rosen and Baker do not report results with a four-coefficient model. The fixed- n model also has advantages when fitting smaller data sets and when comparing coefficients obtained with different data sets.

The coefficients for the four-coefficient model are listed in row “LM 2000” in Table II. The auditory filter shapes produced by this fit are shown in Fig. 2 as a function of probe level. In the fitting process, the peak frequency of the amplitude spectrum varies with level, as described previously; for clarity, however, the peak frequency is normalized to 2000 Hz in the figure by adjusting the value of f_r in Eq. (2). Below its peak frequency, the skirt of the gammachirp auditory filter broadens substantially with increasing stimulus level; above its peak frequency, the skirt sharpens a little with increasing level. These shapes are quite similar to the roex filter shapes in Fig. 1, although there are small differ-

TABLE I. Relationship between the number of filter coefficients and rms error. Columns 7 and 8 show rms (root-mean-squared) errors in dB obtained with the gammachirp filter and the roex filter when fitting the probe-dependent model with various numbers of coefficients to all 78 data points in Rosen and Baker (1994). The rms errors in column 8 are calculated from the total squared errors in Rosen and Baker (1994). The integers in the first column show the total number of variable coefficients for both the gammachirp and the roex. The integers in other columns show the number of coefficients used for the gammachirp parameter in that column: “1” indicates a filter parameter that is constant across signal level and “2” indicates a linear dependence of the parameter on signal level. The symbol “-” indicates an r value of -100 dB (practically zero in linear terms); “*” indicates an n value of 4; “—” indicates no model fitted at that value. The last row shows the results without parameter c , i.e., the gammatone fit.

Number of coefficients	Gammachirp						roex rms error
	n	b	c	K	r	rms error	
10	2	2	2	2	2	1.18	1.19
9	2	2	2	2	1	1.27	—
8	2	2	2	1	1	1.29	1.19
7	2	2	2	1	-	1.33	1.19
6	1	2	2	1	-	1.33	1.20
6	2	1	2	1	-	1.33	—
5	1	1	2	1	-	1.33	1.42
4	*	1	2	1	-	1.33	—
4	1	2	0	1	-	1.72	—

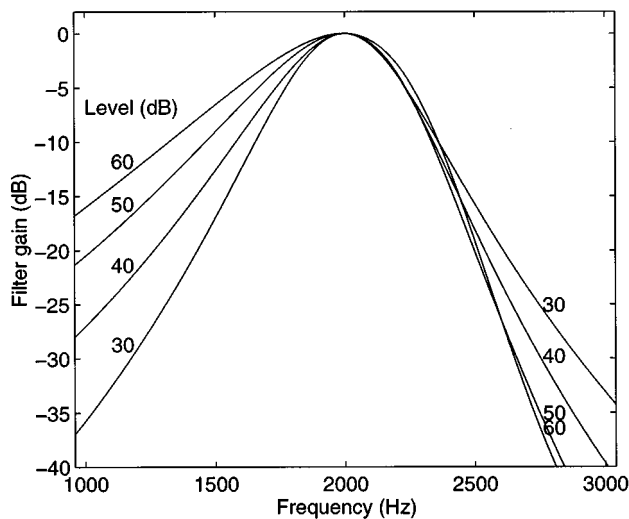


FIG. 2. The gammachirp auditory filter shape as a function of probe level (30–60 dB SPL in 10-dB steps) with four variable coefficients when applied to the masking data of Rosen and Baker (1994). The peak frequency is normalized to 2000 Hz.

ences in the flatness around the peak frequency and in the variability of the upper skirt. Unlike the roex, the upper skirt of the gammachirp has the “backward S” shape observed in the dense threshold functions in Patterson (1976). The gammachirp filter naturally introduces the physical constraints of realistic filters into the estimation of the auditory filter shape. The derived filter shapes are also in agreement with those reported in previous studies (Lutfi and Patterson, 1984; Patterson and Moore, 1986; Moore and Glasberg, 1987).

It also appears that we do not need the parameter r when fitting the data of LM; absolute threshold is sufficient limit to the dynamic range of the fitting process. This is another advantage of using the gammachirp fit.

B. Other data sets

Following the results in the previous subsection, we applied several probe-dependent models to the notched-noise masking data of Lutfi and Patterson (1984) and Moore *et al.* (1990). The models with four, five, and six variable coefficients were fitted to each set of data. As before, the model with four variable coefficients proved most appropriate and so we begin with it. Given their limited size, each set of masking data was fitted with seven different sets of initial values; the set that produced coefficients giving minimum mean squared error is listed with the rms error in Table II. It is clear that the value of b converges between 1 and 2 and that c is always negatively correlated with probe level P_s , as in the previous fits for LM at 2000 Hz. The filter shapes are similar to the shapes in Fig. 2 in terms of change in slope with level, except for four conditions: the filter shape is almost level independent for CP at 200 Hz and RM at 4000 Hz; the upper slope changes as much as the lower slope does for HM at 1000 and 4000 Hz. The last two rows in Table II show the means and standard deviations of the parameter values. Since the mean coefficients are close to those for LM, a “typical” auditory filter set resembles those shown in Fig. 2 when the peak frequencies are normalized to unity. Thus, the gammachirp with four variable coefficients provides a reasonable summary to the masking data in these data sets, although the rms errors are larger than those for the data set of LM.

For completeness, we also performed the gammachirp fit with five variable coefficients (1 n , 1 b , 2 c ’s, and 1 K) and six variable coefficients (1 b , 2 c ’s, 1 K , and 2 r ’s). Only two of the models with five variable coefficients reduced the rms error more than 5%, reductions that are negligible when compared with the variance in the data sets. Since absolute threshold values were not included in these fits, the model with level-dependent r was also applied to each set (i.e., six

TABLE II. Rms errors and coefficients obtained with the gammachirp auditory filter when fitting a probe-dependent model with four variable coefficients. The first column specifies the data source by the initials of the subject. CP represents data from Moore *et al.* (1990); HM, RL, RM, and WW represent data from Lutfi and Patterson (1984); LM represents data from Rosen and Baker (1994). The second column is probe frequency in Hz. The third column shows rms error in dB. The remaining columns show the best coefficients for b , c , and K with $n=4$ and $r=-100$ (dB). The last two rows show the means and standard deviations for b , c , and K .

Subject	Frequency	rms error	b	c	K
CP	200	4.72	1.19	-0.59	-0.0097 P_s
CP	400	2.92	1.43	2.64	-0.082 P_s
CP	800	2.65	1.75	2.16	-0.070 P_s
HM	1000	4.04	1.17	8.43	-0.180 P_s
RL	1000	4.46	1.59	4.98	-0.146 P_s
RM	1000	2.93	1.21	5.27	-0.148 P_s
WW	1000	3.46	1.38	3.56	-0.098 P_s
LM	2000	1.33	1.68	3.38	-0.107 P_s
HM	4000	4.14	1.85	6.31	-0.153 P_s
RL	4000	5.10	1.75	5.17	-0.182 P_s
RM	4000	4.94	1.50	0.61	-0.019 P_s
WW	4000	2.75	1.79	4.18	-0.110 P_s
mean	—	—	1.51	3.88	-0.109 P_s
s.d.	—	—	0.24	2.46	0.057

variable coefficients). All of the derived values for the parameter r were negatively correlated with probe level P_s and were smaller than -40 dB, even when the signal level was 30 dB SPL. That is, the values are less than absolute threshold, and so, absolute threshold is a more suitable limit to the dynamic range when fitting these data. Moreover, the rms error was reduced more than 10% in only three cases. Thus, parameter r does not seem necessary to explain the general form of the masking data, and the model with four variable coefficients seems sufficient to explain these masking data, as well as those of Rosen and Baker.

III. SUMMARY

A “gammachirp” function derived as an optimum auditory filter (Irino, 1995, 1996) is shown to have an asymmetric amplitude characteristic in frequency. Using the power spectrum model of masking, and the assumption that the asymmetry is associated with stimulus level, the amplitude spectrum of the gammachirp was fitted to notched-noise masking data from 12 data sets reported in 3 different studies. A probe-dependent model with four variable coefficients is shown to provide an excellent fit to the masking data. The resultant gammachirp filter shape is similar to that obtained with a six-coefficient roex filter by Rosen and Baker (1994). The gammachirp has a well-defined impulse response unlike the roex auditory filter and, thus, it is an excellent candidate for an asymmetric, level-dependent auditory filterbank in time-domain models of auditory processing.³

ACKNOWLEDGMENTS

The authors wish to thank Brian C. J. Moore and Brian R. Glasberg of Cambridge University, and Stuart Rosen and Richard J. Baker of University College, London for providing their notched-noise masking data, as well as fruitful discussions. The first author wishes to thank Takeshi Okadome of NTT BRL for the suggestion to study optimality of the auditory filter.

APPENDIX A: THE DERIVATION OF THE GAMMACHIRP FUNCTION

The gammachirp function arose from consideration of the contrast between the traditional representation of sound, the spectrogram, and the representation produced by auditory filterbanks designed to mimic the spectral processing of the cochlea. The contrast is set out in Sec. A of this Appendix. It has led to the hypothesis that the time-frequency representation of sound observed at the output of the cochlea is an intervening representation produced by the auditory system to support a subsequent “scale transform,” and that the function that minimizes uncertainty in the time-scale representation is the gammachirp. The scale transform and the gammachirp function are the subjects of Secs. B and C, respectively.

A. The spectrogram and auditory filtering

The spectrogram is a typical example of a joint time-frequency representation of sound. It is produced by converting successive segments of the sound wave into spectral

frames with a Fourier transform. There is a trade-off between the resolution of time and the resolution of frequency in this representation. The trade-off is known as the uncertainty principle, and Gabor (1946) showed that the function which satisfies minimal uncertainty in the joint time-frequency representation is a complex sinusoidal carrier with a Gaussian envelope. This “Gabor function” is symmetric in time and symmetric in frequency; moreover, the frequency bands all have the same width in this time-frequency representation.

The spectral analysis produced by auditory filtering differs significantly from that produced by the Fourier transform: The impulse response of the auditory filter is asymmetric in time with a fast rise and a slow decay (de Boer and de Jongh, 1975; Carney and Yin, 1988); the amplitude spectrum of the auditory filter is definitely not Gaussian (Patterson, 1976), and at high sound levels, it is asymmetric with the lower skirt shallower than the upper skirt (Glasberg and Moore, 1990). The gammatone function [Eq. (1)] provides a much better fit to auditory filtering data than the Gabor function. It is clear, however, that to the extent that the gammatone differs from the Gabor function, it does not satisfy minimum uncertainty in a joint time-frequency representation of sound. Moreover, the bandwidth of the auditory filter increases with center frequency; in the region above about 500 Hz, it is essentially a “constant-Q system,” that is, bandwidth is proportional to center frequency (Greenwood, 1990; Glasberg and Moore, 1990).

It is possible that the auditory system is non-optimal because it has to satisfy some mechanical or physiological constraint that is not compatible with minimal uncertainty, and which restricts the bandwidth to be a proportion of the center frequency. On the other hand, it seemed reasonable, on encountering the discrepancy between optimality and auditory filtering, to explore the possibility that the auditory system is optimal, but optimal for a different representation of sound. It is this hypothesis that led to the “scale transform” and the derivation of the gammachirp function.

B. Scale analysis

Cohen (1991, 1993) has suggested that “scale” is a physical attribute of a signal just like time and frequency, and that a time-scale representation is more appropriate than Fourier analysis for “scaled” signals. A “scaled” signal is simply one that is compressed or extended in time relative to the original, as when a tape recording is replayed at a rate faster or slower than that at which it was recorded. Cohen (1991) introduced a “scale transform” in the form of an orthogonal Mellin transform (Titchmarsh, 1948) to produce the scale representation. It is described in Sec. C. The Mellin transform converts a scaled signal into (a) an invariant absolute distribution in the scale representation, and (b) a value specifying the scale value of the signal. When the speed of a recording fluctuates on playback, it has a pronounced effect on the pitch we hear, but the source of the sound is not perceived to change. This suggests that the “scale” value and the “invariant distribution” of the Mellin transform may be analogous to pitch and timbre in the auditory system. Thus, the time-scale representation of sound could have distinct advantages when analyzing systems where a vibrating

source with variable rate excites a complex resonator. “Source-filter” models of this sort are commonly used to explain the production of sound by the vocal tract (Fant, 1970) and musical instruments (Fletcher and Rossing, 1991).

The scale transform can be applied to a sound wave directly. It is clear, however, that in the auditory system, the scale transform would have to be applied after auditory filtering. The wavelet transform is similar to the auditory filterbank inasmuch as it is a “constant-Q” system; both the envelope and the carrier of the impulse response scale with center frequency in these systems. The Mellin transform converts the individual wavelets into an invariant distribution in the scale representation. Thus, with a wavelet filterbank, when a sound is scaled, its components shift to wavelet filters that have been scaled by the same amount. So, the outputs of the scaled filters are exactly the same as the scaled versions of the outputs of the original filters. Both the scaled and unscaled filter outputs are transformed into the same distribution in the scale representation. Thus, the wavelet filterbank is “transparent” to the scaling of sounds, and in this sense, the wavelet transform is optimal as a preprocessor for the Mellin transform. This implies that the auditory filterbank would be a near optimum preprocessor for the Mellin transform.

The optimal relationship between the Mellin transform and the wavelet transform does not uniquely determine the form of the wavelet that produces minimal uncertainty in a joint time-scale representation; Clearly, the Gabor function does not; After Cohen (1991, 1993) adapted Klauder’s (1980) results on affine variables in quantum mechanics to produce the scale representation for time domain functions, he showed that the optimal function for minimal uncertainty in a time-scale representation has a gamma envelope and a monotonically frequency-modulated carrier. In Cohen’s case, the instantaneous frequency of the carrier starts at infinity and converges on zero as time proceeds. This solution is not suitable for relatively narrow-band applications like auditory filtering. This led Irino (1995, 1996) to introduce a frequency shift term that makes it possible to model the bandwidth/center-frequency function of the auditory system, and produce a narrow-band filter centered on a specific frequency. This, in turn, led to the derivation of the gammachirp function through the optimality constraint. This, then, is the logic for the time-scale representation of sound and the gammachirp function.

C. Mathematical derivation

a. The Mellin transform

The Mellin transform (Titchmarsh, 1948) of a signal, $s(t)$ ($t > 0$), is defined as

$$S(p) = \int_0^\infty s(t) t^{p-1} dt, \quad (A1)$$

where p is a complex argument. One of the important properties is

$$\text{if } s(t) \Rightarrow S(p), \text{ then } s(at) \Rightarrow a^{-p} S(p), \quad (A2)$$

where the arrow (\Rightarrow) indicates “is transformed into” and a is a real dilation constant. That is, the distribution $S(p)$ is just multiplied with a constant a^{-p} when the function $s(t)$ is scaled in time. If p is denoted by $p_r + jp_i$,

$$a^{-p} = a^{-(p_r + jp_i)} = a^{-p_r} a^{-jp_i} = a^{-p_r} \exp(-j \ln a p_i), \quad (A3)$$

where $j = \sqrt{-1}$, and \exp and \ln are the exponential and natural logarithmic operators. Since $|a^{-p} S(p)| = |a^{-p_r}| \cdot |S(p)|$, the absolute distribution $|S(p)|$ is not affected by a scaling of the signal, except for the constant that specifies the scale of the current signal; nor is it affected when the distribution is normalized.

b. Minimal uncertainty and operator methods

With the Mellin transform, questions concerning minimal uncertainty in a joint representation are assessed with operator methods. They were introduced into signal processing from quantum mechanics by Gabor (1946) because of the similarity in mathematical formalism. The following is a tutorial on operator methods based on the derivation of the Gabor function; it is adapted from Cohen (1991, 1993).

Time and frequency operators are defined as $\mathcal{T} = t$ and $\mathcal{W} = -j(d/dt)$ in the time domain. When the operator \mathcal{W} is applied to the function $Ae^{j\omega t}$, the result is

$$\mathcal{W} Ae^{j\omega t} = \left(-j \frac{d}{dt} \right) Ae^{j\omega t} = \omega Ae^{j\omega t}. \quad (A4)$$

Thus, for a complex exponential, the operator \mathcal{W} introduces the frequency term ω . This is the essence of operator methods. The commutator between these operators is again an operator; namely,

$$[\mathcal{T}, \mathcal{W}] = \mathcal{T}\mathcal{W} - \mathcal{W}\mathcal{T} = t \left(-j \frac{d}{dt} \right) - \left(-j \frac{d}{dt} \right) t = j. \quad (A5)$$

It is easy to prove by applying this operator to the function $Ae^{j\omega t}$. Since the commutator is not zero, time and frequency do not commute. Thus, time and frequency cannot be measured independently and there is uncertainty between them, and in this case, it is

$$\Delta t \cdot \Delta \omega \geq \frac{1}{2} |\langle [\mathcal{T}, \mathcal{W}] \rangle| = \frac{1}{2} |\langle j \rangle| = \frac{1}{2}, \quad (A6)$$

where $(\Delta \cdot)$, $|\cdot|$, and $\langle \cdot \rangle$ denote the standard deviation, the absolute value, and the average, respectively. Functions which satisfy minimal uncertainty are solutions to the equation

$$(\mathcal{W} - \langle \mathcal{W} \rangle) s(t) = \lambda (\mathcal{T} - \langle \mathcal{T} \rangle) s(t), \quad (A7)$$

where

$$\lambda = \langle [\mathcal{T}, \mathcal{W}] \rangle / 2(\Delta \mathcal{T})^2 = j / 2(\Delta t)^2. \quad (A8)$$

Using $\mathcal{T} = t$, $\mathcal{W} = -j d/dt$, $\langle \mathcal{T} \rangle = \langle t \rangle$, and $\langle \mathcal{W} \rangle = \langle \omega \rangle$, Eq. (A7) is expanded as follows:

$$\left(-j \frac{d}{dt} - \langle \omega \rangle \right) s(t) = \lambda (t - \langle t \rangle) s(t);$$

$$j \frac{d}{dt} s(t) + \lambda t s(t) + (\langle \omega \rangle - \lambda \langle t \rangle) s(t) = 0;$$

$$\frac{d}{dt} s(t) + \frac{1}{2(\Delta t)^2} t s(t) + \left(-j\langle\omega\rangle - \frac{\langle t \rangle}{2(\Delta t)^2} \right) s(t) = 0. \quad (\text{A9})$$

The nontrivial solution is

$$\begin{aligned} s(t) &= a \exp\left(-\frac{1}{4(\Delta t)^2} t^2 + \frac{\langle t \rangle}{2(\Delta t)^2} t + j\langle\omega\rangle t \right) \\ &= a' \exp\left\{ -\frac{1}{4(\Delta t)^2} (t - \langle t \rangle)^2 \right\} \exp(j\langle\omega\rangle t), \end{aligned} \quad (\text{A10})$$

where a and a' are constants. This is the ‘‘Gabor function,’’ and the example shows how it was derived using the constraint that the required function satisfy minimal uncertainty in a joint time-frequency representation.

c. The Mellin operator

Cohen (1991, 1993) introduced the concept of a scale operator into signal processing in the form:

$$\mathcal{E} = \frac{1}{2}(\mathcal{T}\mathcal{W} + \mathcal{W}\mathcal{T}) = \mathcal{T}\mathcal{W} - \frac{1}{2}j. \quad (\text{A11})$$

Previously it had been known as the operator representing an affine variable in quantum mechanics (Klauder, 1980). The corresponding transform, that is, ‘‘the scale transform’’ (Cohen, 1993), is

$$D(c) = \frac{1}{\sqrt{2\pi}} \int_0^\infty f(t) t^{-jc-1/2} dt. \quad (\text{A12})$$

The correspondence between the scale transform and the Mellin transform is revealed by setting $p = -jc + \frac{1}{2}$ in Eq. (A1). Thus, Cohen’s scale transform is the Mellin transform with a specific argument. In Eq. (A12), the argument is restricted in range; we can, however, extend it to cover the entire complex plane by the introduction of two real constants c_0 and μ as follows:

$$p = -j(c - c_0) + (\mu + \frac{1}{2}). \quad (\text{A13})$$

The corresponding Mellin operator is

$$\mathcal{E}_m = \mathcal{T}\mathcal{W} + \{c_0 + j(\mu - \frac{1}{2})\}. \quad (\text{A14})$$

Since we are concerned with signal processing by an auditory filterbank, we introduce a ‘‘frequency-shift’’ term ω_0 into the operator to specify the individual filters. The form of the operator becomes

$$\mathcal{E}_a = \mathcal{T}(\mathcal{W} - \omega_0) + \{c_0 + j(\mu - \frac{1}{2})\}. \quad (\text{A15})$$

The frequency-shift term can be removed later following consideration of the fluctuation of components at the output of the auditory filter (Irino, 1996). The commutator between time and this operator is

$$[\mathcal{T}, \mathcal{E}_a] = [\mathcal{T}, \mathcal{E}_m] = [\mathcal{T}, \mathcal{E}] = j\mathcal{T}. \quad (\text{A16})$$

The operators in Eqs. (A14) and (A15) are not Hermitian except when $\mu=0$; nevertheless, $(\mathcal{E}_a - \langle\mathcal{E}_a\rangle)$ is Hermitian and, thus, the eigenvalue is real. The function that satisfies minimal uncertainty between time and the quantity represented by the operator in Eq. (A14) is the solution to the equation

$$(\mathcal{E}_a - \langle\mathcal{E}_a\rangle)s(t) = \lambda(\mathcal{T} - \langle t \rangle)s(t), \quad (\text{A17})$$

where

$$\lambda = \langle [\mathcal{T}, \mathcal{E}_a] / 2(\Delta\mathcal{T})^2 \rangle = j\langle t \rangle / 2(\Delta t)^2. \quad (\text{A18})$$

Equation (A15) expands to

$$t \left(-j \frac{d}{dt} \right) s(t) - (\omega_0 + j\alpha_1) t s(t) + (-c_1 + j\alpha_2) s(t) = 0, \quad (\text{A19})$$

where $\alpha_1 = \langle t \rangle / 2(\Delta t)^2$, $\alpha_2 = \mu - \frac{1}{2} - \text{Im}\langle c_a \rangle + \langle t \rangle^2 / 2(\Delta t)^2$, and $c_1 = \text{Re}\langle c_a \rangle - c_0$, Re and Im indicate the real and imaginary parts. The solution is

$$\begin{aligned} s(t) &= a t^{\alpha_2 + jc_1} \exp(-\alpha_1 t + j\omega_0 t) \\ &= a t^{\alpha_2} \exp(-\alpha_1 t) \exp(j\omega_0 t + jc_1 \ln t), \end{aligned} \quad (\text{A20})$$

where a is a constant. The envelope $t^{\alpha_2} \exp(-\alpha_1 t)$ is a gamma distribution function $\gamma(t)$. The instantaneous frequency is $\omega_0 + c_1/t$; that is, a fractional function of time. When played as a sound, the carrier would be a chirp, and hence the name ‘‘gammachirp’’ function. When $c_1=0$, Eq. (A20) becomes a gammatone function. Thus, the gammatone function is a first order approximation to the gammachirp function.

APPENDIX B: THE AMPLITUDE SPECTRUM OF THE GAMMACHIRP FUNCTION

The Fourier spectrum of the gammachirp function can be derived analytically. For convenience, we consider a simplified version of the complex form of the gammachirp filter in Eq. (2).

$$\begin{aligned} g_c(t) &= a t^{n-1} \exp(-b't) \exp(j\omega_r t + jc \ln t) \quad (t > 0) \\ &= a t^{n-1+jc} \exp(-b't + j\omega_r t) \quad (t > 0), \end{aligned} \quad (\text{B1})$$

where $b' = 2\pi b \text{ERB}(f_r)$, $\omega_r = 2\pi f_r$, and the phase term ϕ is ignored. The Laplace transform of Eq. (B1) is

$$\begin{aligned} G_C(s) &= a \frac{\Gamma(n+jc)}{\{s - (-b' + j\omega_r)\}^{n+jc}} \\ &= a \frac{\Gamma(n+jc)}{|s - (-b' + j\omega_r)|^{n+jc} e^{j\theta \cdot (n+jc)}} \\ &= a \frac{\Gamma(n+jc)}{|s - (-b' + j\omega_r)|^n e^{-c\theta} \cdot |s - (-b' + j\omega_r)|^{jc} e^{jn\theta}}, \end{aligned} \quad (\text{B2})$$

where $\theta = \arg\{s - (-b' + j\omega_r)\}$. Thus, the absolute value is

$$|G_C(s)| = \frac{|a\Gamma(n+jc)|}{|s - (-b' + j\omega_r)|^n e^{-c\theta}}. \quad (\text{B3})$$

Substituting $s = j\omega = j2\pi f$ into Eq. (B3) to derive the amplitude of the Fourier spectrum of the gammachirp function,

$$|G_C(f)| = \frac{|a\Gamma(n+jc)|}{|b' + j2\pi(f - f_r)|^n} \cdot e^{c\theta}, \quad (\text{B4})$$

where $\theta = \arg\{b' + j2\pi(f - f_r)\}$.

¹Several auditory filters with gamma distribution envelopes and monotonically frequency-modulated (FM) carriers have appeared recently. First, Lyon (1996) has reported an ‘‘all-pole gammatone filter (APGF)’’ based on reduction of zeros from the Laplace transform of the gammatone filter in the s plane. In this case, the intent was to simulate basilar partition motion

‘as a function of stimulus level. Although the impulse response of the APGF is not mathematically equivalent to the gammachirp function, it is similar in having a monotonic FM carrier and a gamma distribution envelope. Second, in an attempt to produce an asymmetric gammatone filter, Baker (1995) replaced the pure-tone carrier with a monotonic FM carrier. Again, the result is not strictly a gammachirp function, but the impulse response has an FM carrier and a gamma distribution envelope. Finally, Laine and Härmä (1996) have suggested similar filters for an auditory filterbank on the Bark scale. The gammachirp function can be viewed as providing the theoretical background to the larger family of auditory filters with gamma distribution envelopes and chirp carriers.

²In their paper Rosen and Baker (1994) presented total-squared-error values of 103.6, 105.0, 110.9, 110.9, 111.0, 111.9, and 158.1 dB², respectively, for these conditions.

³To construct such a filterbank, we would need to develop a mechanism to measure the output level of each filter on a moment to moment basis to specify the appropriate value of c , and thus the filter’s asymmetry, at any given moment. A mechanism of this sort has been developed by Lyon (1982) for a nonlinear filterbank simulating cochlear mechanisms. Thus, it would not appear to be an insurmountable problem to develop one for a gammachirp filterbank. It is, however, beyond the scope of the present paper.

Baker, R. (1995). Personal communication.

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. Neurophysiol.* **60**, 1653–1677.

Cohen, L. (1991). “A general approach for obtaining joint representations in signal analysis and an application to scale,” in *Advanced Signal-Processing Algorithms, Architectures, and Implementation II*, edited by T. L. Franklin, Proc. SPIE, **1566**, 109–133.

Cohen, L. (1993). “The scale representation,” *IEEE Trans. Signal Process.* **41**, 3275–3292.

de Boer, E., and de Jongh, H. R. (1978). “On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,” *J. Acoust. Soc. Am.* **63**, 115–135.

Fant, G. (1970). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fletcher, H. (1940). “Auditory patterns,” *Rev. Mod. Phys.* **12**, 47–65.

Fletcher, N. H., and Rossing, T. D. (1991). *The Physics of Musical Instruments* (Springer-Verlag, New York).

Gabor, D. (1946). “Theory of communication,” *J. IEE (London)* **93**, 429–457.

Giguère, 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.

Greenwood, D. D. (1990). “A cochlear frequency-position function for several species—29 years later,” *J. Acoust. Soc. Am.* **87**, 2592–2605.

Irino, T. (1995). “An optimal auditory filter,” in *Proc. 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 *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-96)*, **2**, 981–984, Atlanta, GA.

Johannesma, P. I. M. (1972). “The pre-response stimulus ensemble of neurons in the cochlear nucleus,” in *Symposium on Hearing Theory (IPO, Eindhoven, Holland)*, pp. 58–69.

Klauder, J. R. (1980). “Path integrals for affine variables,” in *Functional Integration: Theory and Applications*, edited by J. P. Antoine and E. Tirapgui (Plenum, New York).

Laine, U. K., and Härmä, A. (1996). “On the design of Bark-FAMlet Filterbanks,” in *Proc. Nordic Acoustical Meeting*, Helsinki, Finland.

Lutfi, R. A., and Patterson, R. D. (1984). “On the growth of masking asymmetry with stimulus intensity,” *J. Acoust. Soc. Am.* **76**, 739–745.

Lyon, R. F. (1982). “A computational model of filtering, detection, and compression in the cochlea,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-82)*, 1282–1285, Paris, France.

Lyon, R. F. (1996). “The all-pole gammatone filter and auditory models,” in *Proc. Forum Acusticum ’96*, Antwerp, Belgium.

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.

Patterson, R. D. (1976). “Auditory filter shapes derived with noise stimuli,” *J. Acoust. Soc. Am.* **59**, 640–654.

Patterson, R. D. (1994). “The sound of a sinusoid: Spectral models,” *J. Acoust. Soc. Am.* **96**, 1409–1418.

Patterson, R. D., Allerhand, M., and Giguère, 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, 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.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988). *Numerical Recipes in C* (Cambridge U.P., Cambridge).

Rosen, S., and Baker, R. J. (1994). “Characterising auditory filter nonlinearity,” *Hear. Res.* **73**, 231–243.

Titchmarsh, E. C. (1948). *Introduction to the Theory of Fourier Integrals* (Oxford U.P., London), 2nd ed.