LOG-MAGNITUDE MODELLING OF AUDITORY TUNING CURVES

L. Lin, E. Ambikairajah and W.H. Holmes
School of Electrical Engineering and Telecommunications
The University of New South Wales, UNSW SYDNEY NSW 2052, Australia

ll.lin@ee.unsw.edu.au, Ambi@ee.unsw.edu.au, H.Holmes@unsw.edu.au

ABSTRACT

In this paper, we propose the novel application of a technique for filter design that can accurately fit measured tuning curves for the auditory fibres in the log-magnitude domain. This method provides pole-zero filters with guaranteed stability, and its log-magnitude domain criterion allows tuning curves with very steep slopes to be accurately modelled with an 8th to 10th order pole-zero filter. Thus, this technique can also be used to design a new set of critical band filters with superior frequency domain characteristics compared with the well-known gammatone filter bank. The filter bank designed using this technique has applications in auditory-based speech and audio analysis.

1. INTRODUCTION

There has been considerable research devoted to modelling the functional roles of peripheral auditory systems. Although computational auditory models have been shown in some cases to outperform conventional signal processing techniques, especially in noisy environments, adequate modelling of the principal behaviour of the peripheral auditory systems is still a difficult problem.

Many auditory models have been proposed previously and earlier models have used transmission line representations to simulate basilar membrane motion [1]-[8]. Gammatone filters were first used by Flanagan [2] to model basilar membrane motion, and were subsequently used by Patterson *et al.* [10] as a reasonably accurate alternative for auditory filtering, and have since become very popular. Irino *et al.* [3] developed a time-varying analysis-synthesis auditory filter bank called the gammachirp filter bank, which they applied to noise suppression. Lyon [8] introduced an all-pole gammatone filter, which was derived by discarding zeros from the gammatone filter. These all-pole models allowed more controlled behaviour of the tuning-curve tail, bandwidth, asymmetry and centre-frequency shift. The use of all-pole gammatone filters has subsequently become popular, for example Kubin and Kleijn [5] applied them to speech coding and Robert and Eriksson [12] applied them to produce a non-linear, active model of the auditory periphery.

Obtaining an all-pole gammatone digital filter from a gammatone impulse response does not provide an accurate frequency domain description of the tuning curves, particularly at the tails. In this paper, we present a technique for pole-zero digital filter design [4][9] that overcomes this problem. The criterion for filter design is based on the minimization of the difference between the log-magnitude of the measured tuning curve and pole-zero filter. Critical band filters designed using this method can achieve high frequency domain accuracy and computational efficiency.

Section 2 reviews the log-magnitude filter design technique. Its application to the modelling of auditory filters is presented in section 3.

2. FILTER DESIGN BASED ON LOG-AMPLITUDE MEASURE

A technique for designing IIR digital filters having an arbitrary log-magnitude frequency response was proposed by Kobayashi [4]. This technique was subsequently modified and applied to encoding sine wave amplitudes in harmonic speech coding by Malik and Holmes [9]. In the technique, a modelling error criterion J is used which is the sum of squared differences, on a logarithmic scale, between a given set of spectral amplitudes $D_r(\omega_k)$ and the magnitude response of a rational function $H(e^{j\omega_k})$ sampled at the same frequencies [4][9]:

$$J = \sum_{k=0}^{N-1} \left[\log D_r(\omega_k) - \log \left| H(e^{j\omega_k}) \right| \right]^2 , \qquad (2.1)$$

where $\omega_k = \frac{2\pi k}{N}$, k = 0, 1, …, N-1, is a set of uniformly spaced frequencies and $D_r(\omega_k)$, in the application described in section 3, is the auditory tuning curve (positive magnitude values) obtained from direct measurement. Also, $H(e^{j\omega_k})$ is the frequency response of the auditory filter approximation, which has the form

$$H(z) \equiv \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{Q} b_i \, z^{-i}}{1 + \sum_{i=1}^{P} a_i \, z^{-i}} , \qquad (2.2)$$

where a_i and b_i are the filter parameters, P is the number of poles, and Q is the number of zeroes.

The minimization of J with respect to the parameters a_i and b_i is a nonlinear problem whose solution would normally require optimization techniques based on gradients. To avoid the use of optimization methods, we consider an iterative procedure originally proposed in [4]. At the mth step, this iterative linearization procedure selects a_i^m and b_i^m to minimize the error measure

$$J^m = \sum_{k=0}^{N-1} W^m(\omega_k) \left| D(\omega_k) A^m(e^{j\omega_k}) - B^m(e^{j\omega_k}) \right|^2 , \qquad (2.3)$$

where $D(\omega_k)$ is a complex-valued function whose magnitudes are identical to the real valued data $D_r(\omega_k)$. The phase of this function can, in general, be specified without any constraint. For simplicity, we will assume $D(\omega_k)$ to be minimum phase, being uniquely obtained from the log-magnitude data sequence $D_r(\omega_k)$ using real and complex cepstral coefficients [4]. The weights

$$W^m(\omega_k) = \frac{\left[\log D_r(\omega_k) - \log\left|B^{m-1}(e^{j\omega_k}) / A^{m-1}(e^{j\omega_k})\right|\right]^2}{\left|D(\omega_k)A^{m-1}(e^{j\omega_k}) - B^{m-1}(e^{j\omega_k})\right|^2} \quad (2.4)$$

are real, non-negative and even functions of the frequency ω_k, and depend only on the parameters a_i^{m-1} and b_i^{m-1} from the previous step.

The minimization of J^m, given in the form of equation (2.3), with respect to the parameters a_i^m and b_i^m, can now be recognized as a linear weighted least squares problem. Equation (2.3) can be expanded to form [4][9]

$$J^m = \begin{bmatrix}\mathbf{a}^{mT} & \mathbf{b}^{mT}\end{bmatrix}\begin{bmatrix}\mathbf{A} & -\mathbf{C} \\ -\mathbf{C}^T & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{a}^m \\ \mathbf{b}^m\end{bmatrix}, \quad (2.5)$$

where

$$\mathbf{a}^m = \begin{bmatrix}1 & a_1^m & a_2^m & \cdots & a_P^m\end{bmatrix}^T \quad (2.6)$$

and

$$\mathbf{b}^m = \begin{bmatrix}b_0^m & b_1^m & b_2^m & \cdots & b_Q^m\end{bmatrix}^T. \quad (2.7)$$

The *uv*th entries of the matrices **A, B** and **C** are given by

$$A_{uv} = \sum_{k=0}^{N-1} W^m\left(\frac{2\pi k}{N}\right)\left|D\left(\frac{2\pi k}{N}\right)\right|^2 e^{j\frac{2\pi|u-v|k}{N}}, \quad (2.8)$$

$$B_{uv} = \sum_{k=0}^{N-1} W^m\left(\frac{2\pi k}{N}\right) e^{j\frac{2\pi|u-v|k}{N}}, \quad (2.9)$$

$$C_{uv} = \sum_{k=0}^{N-1} W^m\left(\frac{2\pi k}{N}\right) D\left(\frac{2\pi k}{N}\right) e^{j\frac{2\pi(v-u)k}{N}}, \quad (2.10)$$

with $u, v = 1, 2, \ldots, P+1$.

Minimization of J^m with respect to the vectors of parameters $\mathbf{a}_i^m$ and $\mathbf{b}_i^m$ gives rise to a system of linear equations in these unknowns and the minimum value $J_{\min}^m$ of J^m:

$$\begin{bmatrix}\mathbf{A} & -\mathbf{C} \\ -\mathbf{C}^T & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{a}^m \\ \mathbf{b}^m\end{bmatrix} = \begin{bmatrix}J_{\min}^m \\ \mathbf{0}\end{bmatrix}. \quad (2.11)$$

The solution of equation (2.11) is used to update the weight function in (2.4) and the process is then repeated. The complete algorithm converges to a sufficiently small error J^m within 2 to 3 iterations in all instances of its use to date. The initial value of the weight function $W^m(\omega_k)$ can simply be set to 1.

3. FITTING AUDITORY NERVE TUNING CURVES

3.1 Fitting Auditory Gammatone Filters

The logarithmic scale of loudness perception of the human auditory system motivates the application of the filter design technique of section 2 to the approximation of gammatone filters. In this section we apply the filter design technique to fit the well known gammatone filter [2][3][8][10][13]. A gammatone is the product of a rising polynomial, a decaying exponential function, and a cosine wave. Its continuous-time impulse response is given by

$$g(t) = at^{N-1}e^{-2\pi bERB(f_c)t}\cos(2\pi f_c t + \phi), \quad (3.1)$$

where N is the order of the filter, f_c is the centre frequency, $ERB(f_c)$ is the equivalent rectangular bandwidth of the auditory filter, and $a, b \in \Re$ are constants. At moderate power levels [13], $ERB(f_c) = 24.7 + 0.108 f_c$. As an example, the filter parameters $N = 4$, $b = 1.019$, $f_c = 1778.5$ Hz and $ERB(f_c) = 217.7$ Hz yield an 8th order gammatone filter with a centre frequency of 1778.7 Hz.

This 8th order gammatone filter has a closed-form Laplace transform with an 8th order denominator and a 4th order numerator [13]. To apply gammatone filters to digital audio and speech processing, bilinear or impulse invariant transforms have traditionally been applied to obtain IIR filters. However, spectral distortion will be introduced due to the nonlinear frequency warping and possibly 'aliasing', and this will be particularly severe near the Nyquist frequency, which is often well within the audible range. The numerator of the IIR filter will be 4th order, regardless of the order of the prototype analogue filter.

To obtain a digital filter that provides an improved approximation to the analogue frequency domain characteristics of the gammatone filters, the filter design technique introduced in section 2 is utilized. The frequency response of this filter is obtained by taking the Fourier transform of the time sequence of the gammatone filter. The magnitude of the frequency response is then used to fit the data to a digital filter based on the log-amplitude measure.

Different orders for the filter have been tested. If a 'perfect' match of the amplitudes is required, the order of the denominator should not be less than 8 while the order of the numerator can be as low as 3. With a slight sacrifice in accuracy, the filter numerator and denominator orders can be chosen as 1 and 8 respectively. The magnitude frequency response of the resulting filter for the above example is shown in Fig. 1, together with the magnitude response of the original gammatone filter. From Fig. 1, it is evident that the filter design technique provides an excellent approximation to the original gammatone filter log-magnitude frequency response, especially in the important range around the centre frequency. Small discrepancies arise at extremely low and high frequencies, however these are not significant since the magnitudes in these regions are 70 dB or more below those at the centre frequency. By contrast, digital filters derived from the bilinear or impulse invariant transforms incur severe distortion due to the zeros introduced at these points.

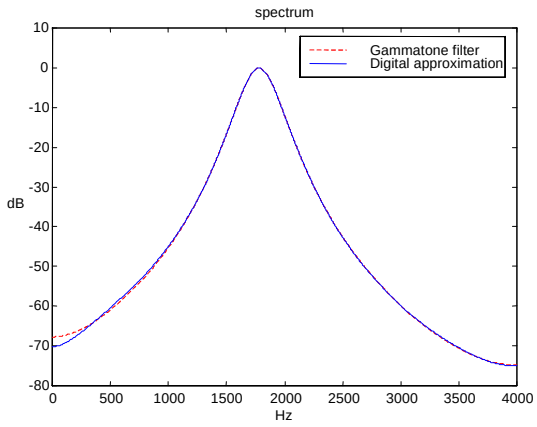

Figure 1. Magnitude responses of a gammatone filter and its digital pole-zero approximation.

A comparison between the impulse responses obtained from the original gammatone filter and its digital pole-zero approximation, illustrated in figure 2, reveals a close similarity. Thus, the digital filter gives a good approximation to the gammatone impulse response, despite the error minimization being in the log-magnitude domain during the filter design.

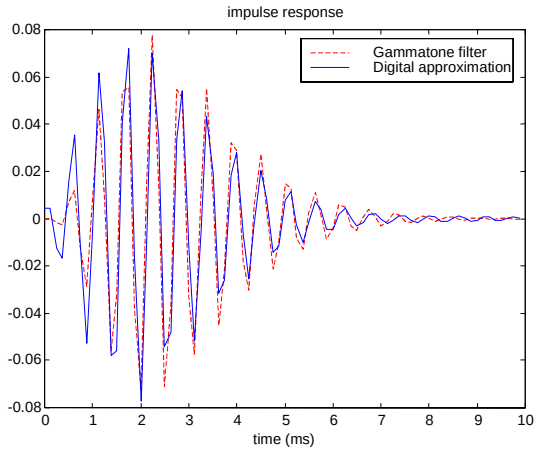

Figure 2. Impulse responses of a gammatone filter and its digital pole-zero approximation

Using the log-magnitude filter design technique, a 10th order all-pole gammatone filter was also designed. The all-pole filter obtained was also found to provide an excellent approximation to both the magnitude response and impulse response of the original gammatone filter. All-pole filters of this kind allow for low complexity speech and audio processing and have the advantage of being easily inverted.

3.2 Fitting Basilar Membrane Magnitude Response

Rhode [11] obtained empirically the amplitude response at a point along the basilar membrane corresponding to a characteristic frequency of 7 kHz (see Figure 3). In comparison with gammatone filter magnitude responses, this amplitude response has a very steep cut-off beyond the characteristic frequency. In the past, transmission line mathematical models [1] have been developed to approximate the amplitude response measured by Rhode. We fit a pole-zero approximation to this amplitude response using the method of section 2. Samples of the measured basilar membrane amplitude response were used as the $D_r(\omega_k)$ in equations (2.1) and (2.3). The best fit obtained was provided by numerator and denominator orders of 6 and 8, respectively. The filter parameters in the transfer function

$$H(z) = \frac{\sum_{i=0}^{6} b_i z^{-i}}{1 + \sum_{i=1}^{8} a_i z^{-i}} \qquad (3.2)$$

were found to be:

$b_0 = 0.6020, b_1 = 0.9036, b_2 = 0.9516, b_3 = 0.0168, b_4 = -0.5394,$
$b_5 = -0.5937, b_6 = -0.2372;$
$a_1 = 0.4347, a_2 = 2.5548, a_3 = 0.4912, a_4 = 2.5183, a_5 = 0.0985,$
$a_6 = 1.2471, a_7 = -0.0218, a_8 = 0.2843.$

The frequency response of the fitted filter (3.2) is illustrated in Fig. 3, and is compared with the original measured response from Rhode [10].

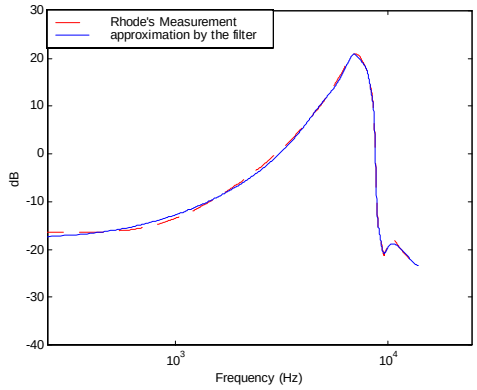

Figure 3. Magnitude response of a basilar membrane magnitude response [10] and its digital pole-zero filter approximation.

This method can thus be applied to fit the measured magnitude response at any point along the basilar membrane without resorting to more complex models such as transmission line techniques. The basilar membrane magnitude responses are the true auditory filters, and provide far superior frequency domain characteristics than the gammatone filters.

3.3 Fitting Auditory-Nerve Tuning Curves

Another application where the technique of section 2 can be used is fitting the auditory tuning curves. Psychoacoustical tuning curves can be obtained by increasing the masker level to the point where a test tone at a given frequency becomes inaudible. A set of tuning curves can normally be obtained using masking thresholds [14], and the shapes of these masking curves are very similar at different centre frequencies a Bark scale. It is also

known that the auditory nerve tuning curves are sharper than the basilar membrane magnitude response.

We choose to apply our technique to three auditory nerve response curves obtained by Liberman [6] from cats raised in a low-noise chamber. The central frequencies of these curves are inside the range 500 to 5000 Hz. It appears that a filter of numerator order 9 and denominator order 9 can fit each measured curve using the technique introduced in section 2. The frequency responses of the filters are shown in Figure 4, together with the original measured tuning curve from [6]. The assumed sampling frequency is 16 kHz.

Testing of this algorithm on other measured tuning curves with arbitrary sharpness also yields approximations with very low spectral distortion. Digital filters obtained directly from auditory tuning curves have a wide variety of applications in speech and audio processing.

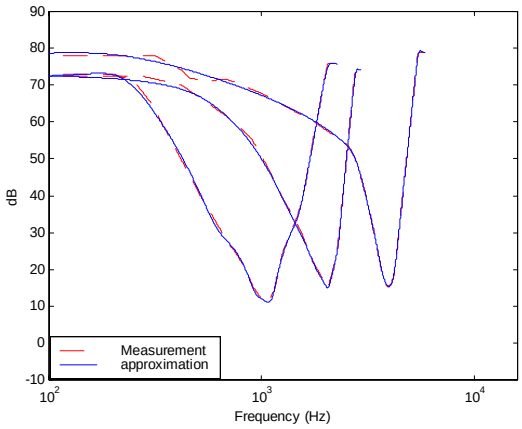

Figure 4. Nerve tuning curves (dashed) and their fitting filters (solid).

4. CONCLUSION

A novel auditory filter design technique was proposed in this paper. Auditory filters were obtained by minimizing the squared difference, on a logarithmic scale, between the measured amplitude of the nerve tuning curve and the magnitude response of the digital filter. The new filter design produces filters whose magnitude and impulse responses are very close to the gammatone magnitude and impulse responses, measured basilar membrane magnitude responses and measured auditory nerve tuning curves. The filters obtained are IIR and are hence computationally efficient. Thus, this technique is highly suitable for the design of auditory filters for speech and audio processing, and provides a new paradigm for critical band auditory filter design with superior frequency domain characteristics.

5. REFERENCES

[1] Ambikairajah, E., Black, N. D. and Linggard, R., "Digital filter simulation of the basilar membrane", *Computer Speech and Language*, 1989, vol. 3, pp. 105-118.

[2] Flanagan, J.L., "Models for approximating basilar membrane displacement", *Bell Sys. Tech. J*, 1960, vol. 39, pp. 1163-1191.

[3] Irino, T and Unoki, M, "A time-varying, analysis/synthesis auditory filter bank using the Gammachirp", *Proc. ICASSP* (Seattle, USA), 1998, pp. 3653-3656.

[4] Kobayashi, T. and Imai, A., "Design of IIR digital filter with arbitrary log magnitude function by WLS techniques", *IEEE Trans. ASSP*, vol. ASSP-38, 1990, pp. 247-252.

[5] Kubin, G. and Kleijn, W.B., "On speech coding in a perceptual domain", *Proc. ICASSP* (Phoenix, USA), 1999, pp. 205-208.

[6] Liberman, M.C. "Auditory-nerve response from cats raised in a low-noise chamber", *J. Acoust. Soc. Am.*, vol. 63, 1978, pp. 442-455.

[7] Lyon, R.F., "A computational model of filtering detection and compression in the cochlea", *Proc. ICASSP*, 1982, pp. 1282-1285.

[8] Lyon, R.F., "The all-pole models of auditory filtering", in *Diversity in Auditory Mechanics*, World Scientific Publishing, Singapore, 1997, pp. 205-211.

[9] Malik, N. and Holmes, W.H., "Encoding sinusoidal amplitudes with a minimum phase rational model", *Proc. ICASSP* (Istanbul), 2000, pp. 1463-1466.

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

[11] Rhode, W.S., "Observation of the vibration of the basilar membrane of the squirrel monkey using the Mossbauer technique", *J. Acoust. Soc. Am.*, vol. 49, 1971, pp. 1218-1231.

[12] Robert, A. and Eriksson, J., "A composite model of the auditory periphery for simulating responses to complex sounds", *J. Acoust. Soc. Am.*, vol. 106, 1999, pp. 1852-1864.

[13] Slaney, M., "An efficient Implementation of the Patterson-Holdsworth auditory filter bank", *Apple Computer Technical Report* #35, Apple Computer, Inc., 1993.

[14] Zwicker, E. and Zwicker, U. T., "Audio engineering and psychoacoustics: matching signals to the final receiver, the human auditory system", *J. Audio Eng. Soc*., vol. 39, No. 3, 1991, pp. 115-125.