Aki Härmä (1), Matti Karjalainen (1), Lauri Savioja (1), Vesa Välimäki (1), Unto K.Laine and Jyri Huopaniemi (2),
(1) Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland.
(2) Nokia Research Center, Speech and Audio Systems Laboratory, Helsinki, Finland

Presented at
the 108th Convention
2000 February 19-22
Paris, France

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Frequency-Warped Signal Processing for Audio Applications

Aki Härmä, Matti Karjalainen,
Lauri Savioja, Vesa Välimäki, Unto K. Laine, and Jyri Huopaniemi*
Helsinki University of Technology,
Laboratory of Acoustics and Audio Signal Processing,
P. O. Box 3000, 02015, Espoo, Finland
* Nokia Research Center, Speech and Audio Systems Laboratory,
Helsinki, Finland

Abstract

Modern audio techniques, such as audio coding and sound reproduction, emphasize the importance of auditory perception as one of the cornerstones for system design. In this article we present a methodology, the frequency-warped digital signal processing, as a means to design or implement DSP algorithms directly in a way which is relevant for auditory perception. Several audio applications are considered in which this approach shows advantages when used as a design or implementation tool or as a conceptual framework of design.

0 Introduction

The human auditory system is a very complex analyzer that is nonlinear, time-variant, and adaptive in many ways. Thus, models of auditory perception are necessarily complex, and audio techniques that utilize such principles are intricate. Only few properties of the auditory system are such that they can be easily and systematically exploited in audio signal processing.

The most often utilized auditory feature in this sense is pitch scales, i.e., auditory 'frequency' scales, that are nonlinear and nonuniform in relation to the Hertz scale. Examples thereof are the mel scale [1], the Bark scale (critical band rate scale) [2], and the ERB (equivalent rectangular bandwidth) rate scale [3]. A close relative to them is the logarithmic scale that has a long tradition of use in audio technology.In many audio applications, it would

be desirable to design signal processing systems and algorithms that work directly on some of these auditory scales. For example, audio equalizers typically should have such properties, and the psychoacoustic models of audio codecs approximate this kind of behavior.

The conventional frequency scale used in digital signal processing systems is linear in relation to the Hertz scale, that is, the inherent frequency resolution is uniform for the whole band from DC to the Nyquist limit, i.e., half of the sampling frequency f_s. The origin of this fact is the property of the unit delay $D(z) = z^{-1}$, the basic DSP building block, that it delays signal components of all frequencies the same amount: $delay = 1/sample\_rate$. For example, a sample sequence, when Fourier transformed, results in frequency bins that are equidistant in frequency.

There are of course ways to avoid this property of uniform frequency resolution. Recursive filters (IIR filters) can easily be focused to desired portions of the Nyquist band to have sharper resonances and magnitude response transitions than other parts of the frequency scale. However, it would be useful to be able to design digital filters and algorithms directly on nonuniform frequency scales, such as the Bark scale.

In this paper we discuss a general approach to design and implement digital signal processing techniques on a warped frequency scale that approximates well the Bark scale. The paper begins with an introduction to auditory frequency scales in Section 1. In Section 2, the theory of frequency-warped DSP is introduced, and it is studied how well a warped system can approximate auditory frequency representation. This is followed by an introduction to the design and implementation of warped FIR-type and IIR-type filters which are the basic building blocks in signal processing algorithms.

Several audio applications where frequency-warped techniques have shown clear advantages are introduced in Section 3. The warped FFT and filterbank techniques are reviewed and it is demonstrated how a conventional design of a uniform filterbank directly produces a computationally efficient auditory filterbank if it is implemented using warped filters. Parametric techniques for spectral estimation, e.g., warped LPC modeling and adaptive filtering techniques are also available. It is demonstrated that warped LPC is a potential technique for wideband speech and audio coding. The gain that can be obtained by warping an LPC type coding algorithm is shown in listening test results.

Finally, specific applications of warped filters to digital loudspeaker equalization, physical modeling of a guitar body, and implementation of head-related transfer functions are studied. There is also an example where warped

techniques have been used to solve a slightly different problem. In this application, warped filters are used to reduce dispersion errors in a digital waveguide mesh which is used in physical modeling of musical instruments or acoustic spaces. Many other recent applications are also briefly reviewed.

1 Pitch scales and resolutions

1.1 Cochlear mapping

Based on measurements of the mechanical motion on the cochlea and neural recordings, Greenwood introduced a general analytic expression for the cochlear *frequency-position* function [4]. For man, it is given by

$$f = 165.4\,(10^{0.06x} - 1) \tag{1}$$

$$\Leftrightarrow x = \frac{1}{0.06}\log_{10}\left(\frac{f - 165.4}{165.4}\right), \tag{2}$$

where x is the location on the cochlea (in mm) and f is its best frequency in Hertz corresponding to that position. It is usually assumed that the same frequency representation is preserved also at higher neural stages of the hearing mechanism [5].

The first derivative of (1) is

$$\frac{df}{dx} = 22.9 \times 10^{0.06x} = 22.9\,\frac{f + 165.4}{165.4}. \tag{3}$$

df/dx is a function describing the bandwidth related to a one millimeter range on the cochlea.

1.2 Psychoacoustic scales

Psychoacoustic scales are usually based on a filterbank model of the hearing mechanism, see, e.g., [6] for review. Estimates for the bandwidths of the filters, i.e., auditory filters, may be measured in many different ways. Once a function representing the bandwidths of the filters is found, a psychoacoustic frequency-position function may be obtained by integrating over the frequency scale.

A classical and still widely used psychoacoustic frequency scale is based on an assumption that the ear may be modeled as a filterbank of a large number of overlapping rectangular filters. The bandwidth of those filters is called the *critical bandwidth*. Based on several different experiments, Scharf [2] concluded that listeners react in one way when the stimuli are wider than

the critical band and in another way when the stimuli are narrower. The theory of critical bandwidth postulates that the same bandwidth is used in loudness summation, phase sensitivity, harmonic discrimination, and several other psychoacoustic phenomena.

Recently, several authors have pointed out that this is not necessarily the case [5]. The built-in assumption that the filters of the auditory system have rectangular shapes have also been criticized and it has been shown that the off-band listening which was not counted in original experiments may change observed effects significantly [6]. Furthermore, the technique used in determining the critical bandwidths especially at low frequencies may be inaccurate and based on incorrect assumptions [7].

Based on Scharf's data, a useful analytic expression for the critical bandwidth as a function of frequency is given by [8]

$$\Delta f_{CB} = 25 + 75(1 + 1.4(\frac{f}{kHz})^2)^{0.69}, \tag{4}$$

where the unit is Bark. Scharf remarked that the values may differ as much as $\pm 15\%$ among subjects.

A corresponding frequency-position function defines a scale called the *Bark rate scale*. A suitable approximation is given by [8]

$$v = 13 \arctan(0.76\frac{f}{\text{kHz}}) + 3.5 \arctan(\frac{f}{7.5\text{kHz}})^2. \tag{5}$$

A technique which was designed so that the abovementioned problems can be eliminated was introduced in [6]. In this technique, a notched noise masker was used in a novel way to find the shapes of auditory filters of the filterbank model. The bandwidth of an obtained auditory filter is often characterized by its *Equivalent Rectangular Bandwidth* (ERB), i.e., the bandwidth of a rectangular filter which passes the same amount of signal energy as the filter. The following analytic expression for the ERB as a function of frequency was proposed in [3]:

$$ERB = 24.7 + 0.108 f_c, \tag{6}$$

where f_c is the center frequency of the filter. The equation is actually reliable only in the frequency range from 100 Hz to 10 kHz because ERB's have not been measured at very low and very high frequencies. The corresponding ERB scale can be calculated by

$$\int dx = \int \frac{df}{24.7 + 0.108\ f} \tag{7}$$

$$\Leftrightarrow\ x = \frac{1}{0.108} \ln(24.7 + 0.108\ f) + C$$

$$= 21.3 \lg(24.7 + 0.108 * f) + C$$
$$= 21.3(\lg(1 + \frac{0.108}{24.7} f) - \lg(24.7)) + C.$$

Since $f(x = 0) = 0$, the equation reduces to

$$x = 21.3 \lg(1 + \frac{0.108}{24.7} f), \tag{8}$$

or

$$f \approx 229 * (10^{x/21.3} - 1). \tag{9}$$

1.3 Comparison of different auditory scales

The ERB and Bark bandwidth estimates are compared in Figs. 1 and 2. In Fig. 1, the bandwidths are plotted on a log-log scale, and in Fig. 2 the ratios between the Bark, ERB, and third octave bands, and the bandwidth expression of (3) are illustrated.

As shown in the figures, the ERB's are very close to the bandwidth measure derived from Greenwood's formula in (3): The Bark bands are wider at low and high frequencies but the correspondence is excellent at the central range of hearing, i.e., from 700 Hz to 4 kHz. In Fig. 2, it can be seen that the Bark bands are two to four times wider than ERB's below 200 Hz and 1.5 to two times wider above 5 kHz. The corresponding frequency-to-position mappings are compared in Fig. 3. All auditory frequency scales are in between linear and logarithmic scales.

Since the ERB rate scale found using psychoacoustic listening tests and the frequency to position scale derived from physiological and neural measurements coincide, the ERB rate scale may be considered to be a more accurate model than the Bark scale.

2 Frequency warping

The technique of computing non-uniform resolution Fourier transforms using first-order allpass filters was introduced by Oppenheim, Johnson, and Steiglitz in [9]. They used the warping technique with the FFT to compute a non-uniform spectral representations of a signal. The idea of frequency-warped transfer functions dates back to a paper by Schüssler [10]. A technique was proposed where the unit delays of a digital filter are replaced with first-order allpass filters to obtain a variable digital filter that can be controlled by adjusting the coefficient of the allpass element. The idea was also

presented by Constantinides [11]. In 1976, Johnson published a paper where frequency warping was applied to recursive variable-cutoff digital filters. He also discussed about the realizability problem of the warped recursive filters with delayless loops [12].

Strube [13] pointed out that the warping effect may be adjusted to approximate the spectral representation occurring in the human auditory system. He also designed a speech coding scheme where the frequency-warped autocorrelation method of linear prediction is used to estimate coefficients of warped analysis and synthesis filters of the codec. A related technique of cepstral analysis and synthesis on a mel frequency scale was proposed by Imai in [14]. Laine [15, 16] presented a more general theory of frequency warping in terms of his classes of FAM and FAMlet functions. The current authors have used frequency warped techniques in audio coding, e.g., in [17, 18, 19], and in several other audio signal processing tasks [20, 21].

2.1 Allpass filter chain

The transfer function of a first-order allpass (AP) filter is given by

$$D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}. \tag{10}$$

By definition, the magnitude response of the filter is a constant. The normalized phase response of $A(z)$ for various real values of λ is shown in Fig. 4. If the parameter $\lambda = 0$ the transfer function (10) reduces to a single unit delay with linear phase and constant group delay.

The group delays of the same filters are shown in Fig. 5. For example, the group delay of a filter with $\lambda = 0.723$ is approximately 6 samples at low frequencies but less than 0.2 samples at very high frequencies.

Cascading a set of AP filters an AP filter chain of Fig. 6 is built. In this paper, the block diagram for an allpass filter element is unconventional. This was chosen because it is compact and it can be easily used in all structures.

Now, if a signal is fed into an AP chain, the nonlinear group delay of the elements makes low-frequency components proceed slower and high-frequency components faster than in a chain of unit delays. For instance, if the three sinusoidal signals of Fig. 7 are put into an AP-chain of 1000 elements with $\lambda = 0.723$ and the outputs a_i are read after 1000 samples, a *warped signal* consisting of sinusoids of Fig. 8 is obtained. Studying the spectra of the original and warped signal, we immediately see that the frequencies of the sinusoids have been changed. In fact, the mapping from the natural frequency

domain to a warped frequency domain is determined by the phase function of the AP filter which is given by

$$f' = \arctan \frac{(1-\lambda^2)\sin(f)}{(1+\lambda^2)\cos(f) - 2\lambda}. \tag{11}$$

As shown in Fig. 8, warping also changes the temporal structure of the original signal so that low-frequency components are shorter and high-frequency components longer than the original signal. The change in length of a sinusoidal signal is controlled by the group delay function of the AP-filter. The *turning point frequency* f_{tp} where a warped sinusoid is as long as the original sinusoid, or frequency warping does not change the frequency, is where the group delay is equal to one sample period. A convenient expression for f_{tp} as a function of λ and sampling frequency f_s is given by [22]

$$f_{tp} = \pm \frac{f_s}{2\pi} \arctan \sqrt{\lambda^{-2} - 1}. \tag{12}$$

For a certain value of λ, the frequency transformation closely resembles the frequency mapping occurring in the human auditory system. Smith and Abel [23] derived an analytic expression for λ so that the mapping, for a given sampling frequency f_s, matches the *Bark* rate scale mapping. The value is given by

$$\lambda_{f_s} \approx 1.0674(\frac{2}{\pi}\arctan(0.06583 f_s))^{1/2} - 0.1916. \tag{13}$$

At the 44.1 kHz sampling rate, $\lambda = 0.756$. The mapping occurring in the allpass chain is very close to the Bark rate scale mapping and therefore it has been referred to as the *Bark bilinear mapping* or *Bark warping.* If the low-frequency part of the mapping is emphasized, a slightly higher value, e.g., $\lambda = 0.78$ gives the best fitting to both the Bark rate scale and Greenwood's frequency-to-position function. However, in practice the significance of the second decimal of λ is small.

In Fig. 9, the AP mappings are compared to Greenwood's frequency-to-position curve and to the Bark rate scale mapping. Due to the shape of the first-order allpass mapping, it is not possible to find a parameter value λ for which the mapping would fit well with Greenwood's mapping. Globally, the closest match is found at $\lambda \approx 0.74$ and if the optimization is focused to low frequencies, $\lambda \approx 0.8$.

2.2 Warping as a conformal bilinear mapping

Denoting

$$\tilde{z} = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}, \tag{14}$$

a warped signal is given by

$$S(z) = \sum_{k=0}^{N-1} a_k \tilde{z}^k. \tag{15}$$

In the time domain, this is an a_i-weighted superposition of the impulse responses of the outputs of the AP-chain and therefore a method to *synthesize* a time-domain signal $s(n)$ from its warped counterpart $a(k)$ [24]. Since the warping effect is generally a *shift-variant* operation, it is possible to use this approach only for a finite or truncated impulse response or coefficient sequence. It is important to keep in mind that there are actually two different views to the warping techniques. Firstly, it is possible to warp a signal segment as was done in Fig. 8. Warped signals have rather strange characteristics and therefore this should be done very carefully. Secondly, one may warp a transfer function, coefficient sequence, or an impulse response. This approach is more straightforward and is actually used in all the application examples of this paper.

The relation between the $\mathcal{Z}$-transform of a warped signal and an unwarped signal given by

$$S(z) = \sum_{i=0}^{M} s_i z^{-1}, \tag{16}$$

is a bilinear transformation determined by the following mapping:

$$z^{-1} \rightarrow \tilde{z} = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \tag{17}$$

This is a conformal mapping from the unit disk onto another unit disk.

A useful characteristic of the mapping is that the inverse transformation is produced simply by replacing λ with $-\lambda$. This is easy to show by manipulating (17).

2.3 Warped FIR filters

A warped FIR filter, denoted here WFIR, is obtained by replacing the unit delays of a conventional filter structure with first-order allpass filters (see Fig. 10). In fact, the name WFIR is misleading because the filter has an infinite impulse response, but the structure of the filter is like in a typical FIR filter. Correspondingly, a warped FIR type lattice filter is shown in Fig. 11. Warped filters are closely related to so called digital Laguerre's filters [24, 25, 26, 27]. The only difference between the Laguerre filter and the WFIR filter [10] in this paper is that in a Laguerre filter there is an additional prefilter placed before the AP filter chain. The role of the prefilter

is to orthogonalize the basis functions of the system but its influence in practical applications is marginal. Therefore, all the techniques that work with Laguerre filters are also suitable for frequency warped filters.

WFIR filters may be designed using any conventional FIR design method if the frequency response of the filter is specified in a warped, e.g., Bark domain. Coefficients of a WFIR filter may also be derived from a non-warped FIR filter in the following way. The desired impulse response $h(n)$ and its z-transform $H(z)$ must be equal to the impulse response $\tilde{h}(k)$ and its z-transform $\tilde{H}(\tilde{z})$ in the warped domain, i.e.,

$$H(z) = \sum_{k=0}^{\infty} \tilde{h}(k)\tilde{z}^{-k} \text{ and } \tilde{H}(\tilde{z}) = \sum_{n=0}^{\infty} h(n)z^{-n} \tag{18}$$

Mappings between sequences $h(n)$ and $\tilde{h}(k)$ are linear but not shift-invariant. The first form specifies the WFIR realization (= synthesis) structure yielding

$$H_{\text{WFIR}}(z) = \sum_{n=0}^{M} \tilde{h}(n)\tilde{z}^{-n} = \sum_{n=0}^{M} \beta_n \{D(z)\}^n \tag{19}$$

and the second form of (18) yields a method to compute the WFIR coefficients (= analysis). It is easy to show from (17) that both forms of (18) may be computed with the same warping structure but using coefficient λ for synthesis and $-\lambda$ for analysis. That is, reversing the mapping (17) the second part of (18) may be written as

$$\tilde{H}(\tilde{z}) = \sum_{n=0}^{\infty} h(n)z^{-n} = \sum_{n=0}^{\infty} h(n) \left(\frac{z^{-1} + \lambda}{1 + \lambda z^{-1}} \right)^{-n} \tag{20}$$

Notice that both forms of (18) yield responses of infinite length even if the sequence to be mapped is finite. Since the coefficient sequence β_i must in practice be finite, we have to approximate $\tilde{h}(i)$ by truncation to M samples as in (19).

The reflection coefficients of the warped lattice filter may be obtained from coefficients of the WFIR using the same recursive computation as in the case of non-warped filters, e.g., in [28].

2.4 Warped IIR filters

A general form for the transfer function of a warped IIR (WIIR) filter is given by

$$H_{\text{WIIR}}(z) = \frac{\sum_{i=0}^{M} \beta_i [D(z)]^i}{1 + \sum_{i=1}^{R} \alpha_i [D(z)]^i} \tag{21}$$

Since the implementation of the numerator of the transfer function, i.e., a WFIR filter, is straightforward, we concentrate on the denominator, i.e., a warped all-pole filter.

The main problem in implementation of a warped direct form all-pole filter of Fig. 12a is that the filter contains delay-free recursive loops. The filter cannot be implemented directly. However, it is possible to implement the filter using a two-step technique proposed in [29, 30]: The output of the filter is first computed using a modified difference equation; After that the inner states of the filter are updated using the computed output value. From the same formulation it is also possible to derive a new modified structure where the delay-free loops are eliminated. The modified filter is shown in Fig. 12b. The coefficients σ_i of the modified structure may be computed efficiently from α_i's using a simple algorithm introduced in [31], and independently in [20]. The two-step implementation of the original filter in Fig. 12a is computationally less expensive than the modified structure of Fig. 12a only if the filter coefficients α_i are updated at each sample period.

A *predictor* type formulation of a warped IIR lattice is shown in Fig. 13a. The filter contains even more recursive delay free loops. However, the two-step procedure can be used and, as in the case of the direct form IIR filter, a new modified structure shown in Fig. 13b can be found. There is also a recursive algorithm for computing the coefficients c_i from reflection coefficients k_i of the system [29]. The modified structure is computationally more efficient than the two-step implementation only if the coefficients are held constant over several hundred sample periods.

These four filters may be combined to form warped pole-zero filters. It is possible to use practically any conventional method of designing IIR filters to design WIIR filters if the filter specification is given in the warped domain. In addition, it is possible to convert any ordinary IIR filter to a warped filter by applying the technique presented in Section 2.3 separately to the numerator and denominator of its transfer function. However, this usually requires higher filter order than the original filter, and therefore it may be an impractical way of designing warped filters. A more efficient method is to solve the poles p_i and zeros m_i of the transfer function and map them explicitly to the warped domain. This may be done using the following formulas:

$$\tilde{p_k} = \frac{p_k + \lambda}{1 + p_k\lambda}, \quad \tilde{m_k} = \frac{m_k + \lambda}{1 + m_k\lambda}, \quad k = 1, 2, 3, \cdots N, \tag{22}$$

where $\tilde{p_k}$ and $\tilde{m_k}$ are the corresponding poles and zeroes in the warped $\tilde{z}$-domain. Yet another alternative is to first compute the impulse response of

the original filter, warp it using the synthesis technique represented in Section 2.3 and then use, e.g., Prony's method [32] to find a pole-zero model to approximate the warped impulse response. This gives directly the coefficients of a warped pole-zero filter. However, it is often advantageous to modify the target impulse response in advance, for example, so that it has a minimum phase.

3 Audio applications of frequency warping

Since frequency warping is just a bilinear mapping from an unit disk onto another unit disk, most of the techniques for parametric spectral estimation, adaptive filtering, and predictive coding are immediately available for warped systems, too. Frequency-warped techniques may be combined with most of the conventional DSP methods including digital filtering, filterbanks, AR, MA, and ARMA modeling. In audio applications the main advantage of using frequency-warped techniques is the automatic utilization nonuniform frequency representation as was discussed above.

The authors have recently published a free Matlab [33] toolbox for frequency-warped signal processing. The toolbox is available at

`http://www.acoustics.hut.fi/software/warp`

and it contains many of the techniques presented in the previous section and some examples related to the applications in this section.

3.1 Warped FFT and filterbanks

A frequency-warped spectrum may be computed directly by applying the Fast Fourier Transform (FFT) to the outputs of an AP filter chain [9] as illustrated in Fig. 14. This technique may also be interpreted as a nonuniform resolution filterbank. Bark-warped filterbank with 16 channels is shown in Fig. 15. The filterbank is computationally very efficient and easy to design and implement. However, the sidebands have too high a level for many practical audio applications. It is possible to enhance the filterbank by using a longer an AP-chain and by combining the neighboring channels [34]. Also a more suitable window function can be used to improve the sidelobe attenuation.

Warped filterbanks may also be designed directly. In [35], Karjalainen introduced an auditory filterbank where the center frequencies of the filters follow Bark-warping and the shapes of the filters are given by the classical approximation for the psychoacoustic *spreading function* or selectivity curves

of hearing [36]. The design of the filters is straightforward because the shape of an auditory filter is suggested to be uniform on the Bark-scale, i.e, the domain where the filters are designed. It is also possible to design IIR-type warped filterbanks. A set of filters from a filterbank which consists of 24 fifth-order warped Butterworth filters is shown in Fig. 16. The bandwidth of each rectangular shaped filter is one Bark.

The main problem with warped filterbanks is that they are based on IIR filters and therefore the critical sub-sampling with perfect reconstruction (PR) is impossible in most cases. Therefore, warped filterbanks are probably not as suitable for coding applications as conventional filterbanks. Nevertheless, there are techniques where perfect reconstruction after sub-sampling can be obtained. Laine [37] has introduced a warped block-recursive algorithm which çan be interpreted as an approximately PR nonuniform resolution filterbank. The approximation error is small in typical applications of the method. Recently, Evangelista [38] introduced a frequency-warped wavelet transform having the PR characteristics. However, the method is based on time reversal of the entire signal, thus real-time applications cannot be considered.

3.2 Warped linear prediction and adaptive filtering

LPC is a powerful technique to model a spectrum, e.g., in coding applications. It is well motivated in terms of human hearing because it gives an allpole spectral representation which concentrates on modeling spectral peaks. The ear is known to be relatively insensitive to spectral zeroes. The representation of spectral information as a small set of parameters which can be quantized very efficiently is a beneficial feature especially in coding applications. LPC is a standard technique in speech coding [39] and it is finding its way to wideband audio coding applications [40], too.

Warped linear predictive coding, WLPC, was first proposed by Strube in 1980 [13] but the idea of performing linear predictive, LP, analysis on a modified frequency scale was invented earlier, see e.g. [41].

In classical forward linear prediction [42] an estimate for the next sample value $x(n)$ is obtained as a linear combination of N previous values given by

$$\hat{x}(n) = \sum_{k=1}^{N} a_k x(n-k), \text{ or } \hat{X}(z) = [\sum_{k=1}^{N} a_k z^{-k}]X(z), \tag{23}$$

where a_k are fixed filter coefficients. Here z^{-1} is a *unit delay filter* or a *shift operator*, which may be replaced by a first-order allpass filter, which is here

denoted by $D(z)$, to obtain

$$\hat{X}(z) = [\sum_{k=1}^{N} a_k D(z)^{-k}]X(z). \tag{24}$$

In the time domain, $D(z)^{-k}$ can be interpreted as a *Generalized Shift Operator* defined as

$$d_k[x(n)] \equiv \underbrace{h(n) * h(n) * \cdots * h(n)}_{k\text{–fold convolution}} * x(n), \tag{25}$$

where the asterisk is a convolution and $h(n)$ is the impulse response of $D(z)$. Furthermore, we denote $d_0[x(n)] \equiv x(n)$. The minimum mean square error of the estimate may now be written as

$$e = E\left[|x(n) - \sum_{k=1}^{N} a_k d_k[x(n)]|^2\right], \tag{26}$$

where $E[\cdot]$ is expectation. Minimization of this with $\partial e/\partial a_k = 0$ and $k = 1, 2, \cdots, N$ leads to a system of *normal equations*

$$E[d_m[x(n)]d_0[x(n)]] - \sum_{k=1}^{N} a_k E[d_k[x(n)]d_m[x(n)] = 0, \tag{27}$$

with $m = 0, \cdots, N-1$. Since $D(z)$ is a linear filter, it is straightforward to show that

$$E[d_m[x(n)]d_k[x(n)]] = E[d_{m+p}[x(n)]d_{k+p}[x(n)]], \; \forall j, k, p, \tag{28}$$

which means that the same correlation values appear in the both parts of (27). Therefore (27) can be seen as a generalized form of the *Wiener-Hopf* equations. The correlation terms can be easily computed using the autocorrelation network of Fig. 17. The *optimal* coefficients a_k can be solved efficiently using, e.g, the *Levinson-Durbin* algorithm equally as in the conventional autocorrelation method of linear prediction. Correspondingly, we now have a *prediction error filter* given by

$$A(z) = 1 - \sum_{k=1}^{N} a_k D(z)^{-k}, \tag{29}$$

which can be implemented directly by replacing all the unit delays of a conventional FIR structure with $D(z)$ blocks. It is also possible to implement a *synthesis filter* given by

$$A^{-1}(z) = \frac{1}{1 - \sum_{k=1}^{N} a_k D(z)^{-k}}, \tag{30}$$

using, e.g., techniques presented in [29].

In the spectral domain the solution of (27) is equivalent to matching the power spectrum $P(f)$ of the signal with an estimate given by the warped allpole model

$$P(f') \sim \frac{G^2}{\left|1 + \sum_{k=1}^{M} a_k e^{-i*2\pi f'/N}\right|^2}, \tag{31}$$

where G is a gain term and f' are warped frequency bins given by (11). It is now easy to see that the matching of the model is done on the warped frequency scale, see, e.g., [13] for details.

Figs. 18a and b illustrates this and show how the approach differs from conventional LP modeling. The three curves in Figs. 18a show a power spectrum of an excerpt of a sound of the *clarinet* and the spectral estimates that have been obtained by a conventional and warped LP analysis. The order of the model is 40 in both cases. In Fig. 18b the same data is shown on a warped frequency scale which approximates the Bark scale. The warped model has been optimized on this scale and therefore the frequency resolution is significantly better at low frequencies while the conventional model pays too much attention to insignificant spectral details at very high frequencies at the cost of a poorer resolution at low frequencies.

A characteristic of classical D*PCM [43], or *residual driven* [44], LP coding is that the spectrum of the quantization error signal in the decoded signal has the same spectral shape as the estimated allpole model. Hence, the *frequency masking* effect of hearing is automatically utilized at least to some extent. In warped LP coding this feature is even more pronounced. The spectral estimates in Fig. 18b can be assumed to be close to the *masked threshold* that is related to the original signal. In Fig. 19 the quantization noise spectra in a simple WLP based D*PCM codec is compared with an MPEG I layer 3 codec at the output of a computational model of human hearing. Even when the WLP codec has no auditory model controlling the quantization process, the noise spectra in the two cases are very similar.

Practically all conventional LP coding techniques can be warped. One may take, e.g., any coder from the vast speech coding literature and warp it using the techniques presented above. It can be assumed that in many cases a warped coder can be made to work better than a conventional coder with a wider frequency bandwidth. However, there are some specific techniques, e.g., Barnwell's adaptive autocorrelation method [45], where the application of warping techniques produces unexpected stability problems, and the adaptive least-square lattice method [46], for which the corresponding warped adaptation algorithm is difficult to formulate, see, e.g., discussion in

[47].

3.2.1 Evaluation of warped LPC

To show the difference between LPC and WLPC, the first author of this paper performed a set of preliminary listening tests with a simulated generalized LP based audio codec. A more detailed paper about this topic is going to be published during the year 2000. Both warped and conventional LPC were used in the listening tests.

The structure of the encoder is shown in Fig. 20a. Here, an LPC model is estimated using the autocorrelation method of LPC (or WLPC) with a 20 ms Hamming window. The windows overlap by 10 ms and the *reflection coefficients* are linearly interpolated between frame centers. The residual signal $r(n)$ is produced by filtering the original signal with a time-varying inverse filter given by (29). Next, the residual is quantized using a simple 5-bit backward adaptive scalar quantizer [48] to produce a quantized residual $\hat{r}(n)$. In the decoder, see Fig. 20b, the quantized residual is first subtracted from the original residual to produce a quantization error signal $q(n) = r(n) - \hat{r}(n)$, which is approximately white noise but follows roughly the energy envelope of the original signal. The excitation for the time-varying synthesis filter (30) is a weighted sum of the original residual and the quantization error signal given by

$$\tilde{r}(n) = r(n) + \sqrt{10^{-SNR/10}} G_r q(n), \tag{32}$$

where G_r is a gain coefficient which is used to scale $q(n)$ so that $E[|G_r q(n)|^2] = E[|r(n)|^2]$. In listening tests, a subject may adjust the parameter SNR in real time to find the threshold of audibility for the quantization noise in presence of the signal. The parameter SNR is the *Signal-to-Noise Ratio* for the residual signal and therefore it has, roughly, the following relation to the bit-rate of the quantizer:

$$SNR/\mathrm{dB} = 6b + \delta, \tag{33}$$

where b is the number of bits and δ is some constant, see e.g., [49].

The preliminary listening test reported in this paper was done in a standard listening room with one loudspeaker. The two listeners used a computer mouse to adjust a slider, corresponding to the SNR parameter, in a graphical user interface. The listening test system works in real-time so that the change in SNR can be immediately heard. The test material consisted of 11 steady-state musical and speech sounds. The duration of each sample was one second but the sounds were played in a continuous loop.

The warped and conventional LPC simulations were tested at four different sampling rates (8, 16, 32, and 48 kHz) and with three different orders of the LPC or WLPC model, i.e, 20, 40, and 50.

The average listening test results over all test samples and subjects are shown in Fig. 21. At sampling rates of 48 kHz and 32 kHz, the SNR for residual in the warped LPC is approximately 6 dB below that of a conventional LPC. According to (33) this means that a sufficient bitrate for residual in WLPC is one bit per sample less, i.e., 48kb/s or 32 kb/s less, than in LPC. At the 16 and 8 kHz sampling rates the difference between WLPC and LPC is a decreasing function of model order. In the case of 50th order model at the 16 kHz sampling rate, or 35th order model at the 8 kHz sampling rate the use of warped LPC brings no gain compared to the conventional case. However, below that the difference is clear.

It is not shown in the figures, but the curves can be almost linearly extrapolated towards lower orders of the model. For example, at 8 kHz sampling rate the difference between a 10th order conventional and a warped LPC model is around 5 dB. Basically, this means that the order of the model can be significantly lower in the warped LPC compared to the conventional LPC. Similar results with narrow band speech coding were also obtained by Krüger and Strube [50] and, e.g., Koishida et al. [51].

3.2.2 Warped adaptive filtering

Warped or Laguerre adaptive filtering have been studied by several authors. Den Brinker [52] used a well known LMS-type algorithm with a Laguerre filter, and Tokuda et al. [53] introduced a speech coding method where the coefficients of a warped filter are updated using their adaptive mel-cepstral analysis technique. Fejzo and Lev-Ari [47] studied the properties of an adaptive warped lattice filter where the coefficients are updated using the Gradient Adaptive Lattice (GAL) algorithm. The adaptive warped filters share the same characteristics as warped linear predictive methods, i.e., the frequency resolution follows from the characteristics of the warping function.

A perceptual audio codec based on backward adaptive Bark-warped lattice was introduced in [19]. Since the codec is warped it has similar advantages to those of the WLP codec presented in the previous section, e.g., the noise masking characteristics of the ear are automatically utilized. In addition, the use of backward adaptation makes it possible to minimize the coding delay of the codec. In fact, the coding delay in its first prototype is equal to one sample period. In conventional audio codecs, where the auditory modeling is realized as a separate block where the noise masking characteristics are

determined from a long term FFT spectrum, it is probably not possible to achieve that low coding delay. This is one obvious advantage of using a warped DSP system where the auditory model is incorporated into the coding process.

3.3 Audio equalization

A traditional use of the frequency warping technique is the design of digital filters for audio equalization. Moorer [54] discussed the modification of the resonance or cutoff frequencies of parametric equalizers, shelving filters, and notch filters using the first-order allpass mapping. It can be used to derive closed-form formulas for the coefficients of digital equalization filters [54, 55].

Recently, it was shown that the use of frequency warping is advantageous in fixed-point implementations of digital audio filters. An audio equalizer that is designed and implemented as a warped filter is less sensitive to coefficient quantization and round-off noise than conventional digital filters [56]. The resulting quantization noise level is low and the noise has a lowpass characteristic.

3.3.1 Loudspeaker equalization

Loudspeaker response equalization by digital inverse filtering is becoming a well known technique although relatively few commercial implementations exist. The most common method is FIR equalization but IIR filters have also been used. The equalization is applied either to the magnitude response only or to both magnitude and phase. The applicability of different equalizer filter structures, including warped structures has been studied in [57].

It turns out that FIR filters are very efficient at high frequencies. This is due to the fact that FIRs inherently yield a uniform frequency resolution while in audio the response specifications as well as response measurements are given on a logarithmic scale. Thus FIRs are particularly problematic to equalize at low frequencies. IIR filters avoid some of the problems with FIRs but they are more difficult to design and share the frequency resolution problem.

It was shown [35] that WFIR and WIIR filter structures are good competitors to traditional filters. Figure 22 shows a set of magnitude responses for a less-than-medium quality speaker, including the original response and three equalized ones. The WIIR equalizer (inverse filter) design was based on the warped Prony's method. Very low filter orders (less than 10) already show good overall equalization. Figure 22 depicts also a comparison to tra-

ditional FIR filter equalizer (order 105) which yields about the same degree of equalization as the WIIR filter of order 24. Notice also that while the FIR filter does best job at high frequencies, the WIIR filters work best at middle to low frequencies (depending on the amount of warping).

Loudspeaker equalization techniques based on warped filters have been recently studied by other authors, e.g., in [56, 58]. There is also available a free software package for loudspeaker equalization which uses WFIR filters [59].

3.4 Physical modeling of the guitar body

The next application example concerns model-based synthesis of the acoustic guitar. The modeling and real-time synthesis of string vibration is relatively well understood and the body can be simulated efficiently by commuted synthesis (body response is used as excitation) but simulation of the body as a digital filter is computationally very expensive [60].

A typical magnitude response of the acoustic guitar body is shown in Fig. 23a. For a sample rate of 22 kHz, simple FIR implementation of it requires a filter order of about 2000–5000 for good result since the lowest resonances are sharp and they decay slowly. On the other hand, the high-frequency modes are much broader in bandwidth and thus they decay faster. FIR modeling is not well suited and an IIR model fits better. Using linear prediction an all-pole model of order 500–1000 works relatively well.

Warped FIR filters of order 500 are comparable to those mentioned above. WIIR filters yield the lowest order so that a denominator order of 100–200 and numerator order of 50–100, designed using Prony's method in the warped domain, are comparable. Since the warped structures are inherently more complex, a small efficiency advantage over traditional filters remains when implementing the warped structures using typical DSP processors. In this body modeling case the frequency warping has a double match to the problem. Firstly, physically, the warping means balancing of the resonance Q values so that in the warped domain the sharp low frequency peaks will be broadened to be more similar to the high frequency resonance peaks (Fig. 23b). Secondly, the warping has a natural match to the auditory resolution and Bark scale so that the filter order which is needed is minimized.

3.5 HRTF filtering

Real-time digital modeling of human spatial hearing cues is often referred to as 3-D sound spatialization or auralization. The static cues of spatial

hearing are contained in head-related transfer functions (HRTF). Traditionally, HRTF filters have been created using minimum-phase reconstruction and different FIR and IIR design methods (see [61] for detailed summary). The use of warped filters in binaural and transaural filter design was investigated in [62]. The use of a psychoacoustically based frequency scale is well motivated, and considerable reduction of filter order can be achieved using warped designs. The transfer function expressions of warped filters may be expanded (dewarped) to yield equivalent IIR filters of traditional form, such as direct form II filters. Such implementations have been reported in the literature [63]. An alternative strategy is presented in [61, 64], where implementation is carried out directly in the warped domain using warped FIR and IIR structures.

Theoretical and empirical investigations [61] have shown that dewarped WIIR structures outperform traditional FIR and IIR design methods. In Fig. 24, a comparison of design methods is illustrated. A dummy head HRTF was used, and two filter orders (48 and 16 taps for FIR, orders 24 and 8 for IIR and WIIR) were tested. It can clearly be seen from the results that the fit at lower frequency is enhanced in WIIR designs with a trade-off of reduced high frequency matching. According to the psychoacoustic theory, this can be tolerated. in summary, the use of auditorily motivated filter design in 3-D sound applications has a clear computational advangate without sacrificing perceptual accuracy.

3.6 Improved Digital Waveguide Mesh Simulations

The digital waveguide mesh was introduced in 1993 [65]. It is a finite-difference time-domain simulation, where the vibrating surface has been discretized. The model consists of a rectangular grid where the signal value at every node is updated at each sampling interval using the state of the four neighboring nodes. It was shown that the method is suitable for sound synthesis of percussion instruments although it suffers from direction-dependent dispersion [65]. In 1994, Savioja et al. extended the use of the digital waveguide mesh to three dimensions, and presented simulation results of wave propagation in acoustic spaces [66].

The main weakness of the digital waveguide mesh technique is the dispersion error that increases with frequency. For this reason, the method can be used for accurate numerical simulations at low frequencies only, that is, the sampling frequency (in both time and space) must be very high in acoustic simulations. The dispersion error appears as a frequency error in the simu-

lation results: the modes at high frequencies occur at incorrect frequencies. The frequency-dependence implies that the error caused by dispersion varies as a function of propagation direction: the frequency of standing waves that are formed in the diagonal direction - with respect to the sampling grid - are exact, while the frequencies of standing waves in any other direction are too small, and the error increases with frequency, i.e., higher modes are displaced more than the lower ones.

A triangular waveguide mesh [67, 68, 69] has been developed to overcome the direction-dependent wave propagation characteristics and dispersion error. It is based on the idea that the sampling points are at corners of equilateral triangles instead of squares. The interpolated mesh was also devised to reduce the error while still using the convenient rectangular sampling grid [70, 21]. The key idea was that sample updates should account for more propagation directions than just four, as in the original mesh. The interpolation effectively inserts new nodes in the mesh - the contribution of the hypothetical nodes is then spread on the existing neighboring nodes to obtain a realizable structure. This is a multi-dimensional application of fractional delay filters [71].

Using the interpolated digital waveguide mesh, it was only possible to reduce the direction-dependence while the dispersion was not much affected - it was merely rendered almost independent of direction. The triangular mesh also improved the direction dependence although the dispersion itself was also much reduced. Luckily, the remaining dispersion error can be made considerably smaller in both cases using frequency warping. It is implemented by post-processing the output signal of the mesh using a warped-FIR filter. The best results so far have been obtained using the triangular mesh together with a WFIR filter [72, 21]. Also, the interpolated rectangular mesh is improved using frequency warping [73, 21].

In this application, the frequency warping method can be used since the error to be corrected is almost identical in all directions and the error function is relatively smooth [73, 72, 21]. The frequency warping is implemented with a warped-FIR filter in which the tap coefficients are set equal to the output samples of the digital waveguide mesh algorithm. When a unit impulse is fed into the filter, the warped signal is obtained at the output. The extent of warping is controlled by the value of the allpass filter coefficient λ, which can be optimized using one of many optimization methods, such as those based on the least squares or the minimax criterion.

We present results from a simulation of a square membrane with rigid boundaries using different digital waveguide mesh algorithms. For details of

the simulation, see [73, 72, 21]. Figure 25 shows the magnitude spectra of the simulated membrane in three cases: (a) the original, (b) the warped optimized interpolated, and (c) the warped triangular digital waveguide mesh. Also, the ideal magnitude spectrum is given for comparison in each case.

The error in the frequency of each mode of the membrane is given in Fig. 26. The error behavior of the original mesh reveals that some eigenmodes occur at nearly the correct frequency while others are too low by several percent. The errors in the warped interpolated mesh are smaller - they are within $\pm$ 1.5% in the frequency band from 0 to 25% of the sampling rate. However, the warped triangular mesh is still better: the maximum error on the same frequency band is only 0.60%.

Our example demonstrates the fact that the frequency warping turns the digital waveguide mesh simulations into an accurate method for obtaining impulse responses of acoustic systems. However, a suitable waveguide mesh algorithm needs to be used as well, i.e., either the interpolated rectangular mesh or the triangular mesh.

3.7 Other applications

In speech signal processing, the warped techniques have been used in various applications. Strube's pioneering work [13] with WLP was soon followed by Imai's [14] paper on warped cepstral analysis, where the spectral representation is warped and the magnitude scale is logarithmic. His *mel-cepstral* analysis technique was further generalized in [74]. They have published several articles where the mel-cepstral or *mel-generalized cepstral* analysis techniques have been used, e.g., in speech analysis [75], and coding [51] applications. They have also published a free software package, which has implementations for most of these techniques [76]. Frequency warping has also been applied to speaker verification [77] or to measure the objective quality of coded speech [78]. Recently, warped filterbanks were used in a speech enhancement application [79]. Frequency warped LPC has shown its power also in speech synthesis [80].

4 Conclusions

This article introduced a methodology, frequency-warped signal processing, which can be used to modify frequency representation in digital signal processing systems. The idea is not new but it was almost forgotton after the pioneering works on various distinct fields in 1970's and 1980's. During the

last five years there have been clear signs of a renaissance for this methodology.

It was shown in the article that practically any digital signal processing algorithm can be warped. After presenting the basic methodology, it was shown how this can be utilized in filter design and implementation, non-parametric and parametric spectral estimation, coding, and in a specific application to reduce errors that are induced by a multidimensional discrete-time structure.

The main focus in this paper was in audio and speech signal processing techniques where it is advantageous to design the system so that it takes into account the non-uniform frequency resolution of human hearing. It is shown in the paper that a warped DSP system can be designed to approximate accurately the frequency resolution of hearing. In several presented applications this brings about obvious gains in terms of quality, computational complexity, and digital storage.

Acknowledgments

The work of A. Härmä and V. Välimäki has been supported by the Academy of Finland.

References

[1] S. S. Stevens, J. Volkmann, and E. B. Newman, "A scale for the measurement of the psychological magnitude of pitch," *J. Acoust. Soc. Am.*, vol. 8, pp. 185–190, 1937.

[2] B. Scharf, "Critical bands," in *Foundations of Modern Auditory Theory*, J V Tobias, Ed., pp. 159–202. Academic Press, 1970.

[3] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.*, vol. 88, pp. 132–140, 1990.

[4] D. D. Greenwood, "A cochlear frequency-position function for several species - 29 years later," *J. Acoust. Soc. Am.*, vol. 87, no. 6, pp. 2592–2605, 1990.

[5] B. C. J. Moore, "Frequency analysis and masking," in *Hearing*, B. C. J. Moore, Ed., Handbook of Perception and Cognition. Academic Press, 2 edition, 1995.

[6] R. D. Patterson, "Auditory filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.*, vol. 56, pp. 640–654, 1976.

[7] A. Sek and B. C. J. Moore, "The critical modulation frequency and its relationship to auditory filtering at low frequencies," *J. Acoust. Soc. Am.*, vol. 95, no. 5, pp. 2606–2615, 1994.

[8] E. Zwicker and H. Fastl, *Psychoacoustics: facts and models*, Springer-Verlag, 1990.

[9] A. V. Oppenheim, D. H. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the Fast Fourier Transform," *Proc. of IEEE*, vol. 59, pp. 299–301, 1971.

[10] W. Schüssler, "Variable digital filters," *Arch. Elek. Übertragung*, vol. 24, pp. 524–525, 1970.

[11] A. G. Constantinides, "Spectral transformations for digital filters," *Proc. IEEE*, vol. 117, no. 8, pp. 1585–1590, August 1970.

[12] D. H. Johnson, "Application of digital-frequency warping to recursive variable-cutoff digital filters," in *Journal of the Illuminating Engineering Society IEEE Electron and Aerosp Syst Conv (EASCON '76)*, September 1976.

[13] H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. Am.*, vol. 68, no. 4, pp. 1071–1076, 1980.

[14] S. Imai, "Cepstral analysis synthesis on the mel frequency scale," in *Proc. ICASSP'83*, Boston, 1983, pp. 93–96.

[15] U. K. Laine, "Famlet, to be or not to be a wavelet," in *Proc. Int. Symp. Time-Frequency and Time-Scale Analysis*, Victoria, Canada, October 1992, IEEE-SP, pp. 335–338.

[16] U. K. Laine, M. Karjalainen, and T. Altosaar, "WLP in speech and audio processing," in *IEEE Int. Conf. on Acouctics, Speech, and Signal Processing*, Adelaide, 1994, vol. III, pp. 349–352.

[17] A. Härmä, U. K. Laine, and M. Karjalainen, "WLPAC – a perceptual audio codec in a nutshell," in *AES 102nd Conv. preprint 4420*, Munich, 1997.

[18] A. Härmä, U. K. Laine, and M. Karjalainen, "An experimental audio codec based on warped linear prediction of complex valued signals," in *Proc of ICASSP*, Munich, 1997, vol. 1, pp. 323–327.

[19] A. Härmä, U. K. Laine, and M. Karjalainen, "Backward adaptive warped lattice for wideband stereo coding," in *Proc. of EUSIPCO'98*, Greece, 1998, pp. 729–732.

[20] M. Karjalainen, A. Härmä, J. Huopaniemi, and U. K. Laine, "Warped filters and their audio applications," in *IEEE Workshop on ASPAA 97*, Mohonk, New York, 1997.

[21] L. Savioja and V. Välimäki, "Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques," *accepted for publication in IEEE Trans. Speech and Audio Processing*, 1999.

[22] A. Härmä, "Perceptual aspects and warped techniques in audio coding," M.S. thesis, Helsinki Univ. of Tech./Laboratory of Acoustics and Audio Signal Processing, 1997.

[23] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transform," *Trans. Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, November 1999.

[24] R. E. King and P. N. Paraskevopoulos, "Digital Laguerre filters," *Circuit Theory Applicat.*, vol. 5, pp. 81–91, 1977.

[25] B. Maione and B. Turchiano, "Laguerre z-transfer function representation of linear discrete-time systems," *Int. J. of Control*, vol. 41, no. 1, pp. 245–257, 1985.

[26] B. Wahlberg, "System identification using Laguerre models," *IEEE Trans. on Autom. Control*, vol. 36, no. 5, pp. 551–562, 1991.

[27] T. Oliveira e Silva, "Optimality conditions for truncated Laguerre networks," *IEEE Trans. Signal Processing*, vol. 42, no. 9, pp. 2528–2530, 1995.

[28] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, Macmillan, 2 edition, 1992.

[29] A. Härmä, "Implementation of recursive filters having delay free loops," in *Proc of ICASSP'98*, Seattle, Washington, 1998, IEEE, vol. III, pp. 1261–1264.

[30] A. Härmä, "Implementation of frequency-warped recursive filters," *Signal Processing (to appear)*, vol. 80, no. 3, 2000.

[31] T. Kobayashi, S. Imai, and Y. Fukuda, "Mel-generalized log spectral approximation filter," *Trans. IECE*, vol. 68, pp. 610–611, 1985.

[32] T. W. Parks and C. S. Burrus, *Digital Filter Design*, Wiley, New York, 1987.

[33] The Mathworks, Natick, USA, *Matlab – the language for Technical Computing*, January 1997.

[34] U. K. Laine and A. Härmä, "On the design of Bark-FAMlet filterbanks," in *Proc. of Nordic Acoustical Meeting (NAM)*, Helsinki, 1996, pp. 277–284.

[35] M. Karjalainen, A. Härmä, and U. K. Laine, "Realizable warped IIR filter structures," in *Proc. of NORSIG'96*, Helsinki, 1996.

[36] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech codders by exploiting masking properties of the human ear," *J. Acoust. Soc. Am.*, vol. 99, pp. 1647–1652, 1979.

[37] U. K. Laine, "Critically sampled pr filterbanks of nonuniform resolution based on block recursive FAMlet transform," in *Proc. EUROSPEECH'97*, Rhodos, 1997.

[38] G. Evangelista and S. Cavaliere, "Discrete frequency warped wavelets: Theory and applications," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 874–885, 1998.

[39] N. S. Jayant and P. Noll, *Digital coding of waveforms*, Prentige-Hall, New Jersey, 1984.

[40] N. Iwakami and T. Moriya, "Transform-domain weighted interleave vector quantization (TwinVQ)," in *AES 101st Convention preprint 4377*. AES, November 1996.

[41] J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding in predictive coding of speech," *Trans. on ASSP*, vol. 27, no. 1, pp. 63–73, 1979.

[42] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, Number 12 in Communication and Cybernetics. Springer-Verlag, 1976.

[43] P. Noll, "On predictive quantization schemes," *The Bell Syst. Tech. J*, vol. 57, no. 5, pp. 1499–1532, 1978.

[44] J. D. Gibson, "Adaptive prediction in speech differential encoding systems," *Proc. of the IEEE*, vol. 68, no. 4, pp. 488–524, 1980.

[45] T. Barnwell, "Recursive autocorrelation computation for LPC analysis," in *Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing*, Hartford, May 1977, pp. 1–4.

[46] J. D. Gibson, S. K. Jones, and J. L. Melsa, "Sequantially adaptive prediction and coding of speech signals," *Trans. on Comm.*, vol. 22, no. 11, pp. 1789–1797, 1974.

[47] Z. Fejzo and H. Lev-Ari, "Adaptive Laguerre-lattice filters," *IEEE Trans. Signal Processing*, vol. 45, no. 12, pp. 3006–3016, 1997.

[48] N. S. Jayant, "Adaptive quantization with one word memory," *The Bell Syst. Tech. J*, pp. 1119–1144, 1973.

[49] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall Inc., New Jersey, 1978.

[50] E. Krüger and H. W. Strube, "Linear prediction on a warped frequency scale," *IEEE Trans. Acoust. Speech, and Signal Proc.*, vol. 36, no. 9, pp. 1529–1531, September 1988.

[51] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "CELP coding system based on mel-generalized cepstral analysis," in *Proc. of ICSLP'96*, 1996, vol. 1.

[52] A. C. Den Brinker, "Adaptive modified Laguerre filters," *Signal Processing*, vol. 31, no. 1, pp. 69–79, 1993.

[53] K. Tokuda, T. Kobayashi, S. Imai, and T. Fukada, "Speech coding based on adaptive mel-cepstral analysis and its evaluation," *Electr. and Comm. in Japan, Part 3*, vol. 78, no. 6, pp. 50–61, 1995.

[54] J. A. Moorer, "The manifold joys of conformal mapping: applications to digital filtering in the studio," *J. Audio Eng. Soc.*, vol. 31, no. 11, pp. 826–841, November 1983.

[55] D. K. Wise, "A survey of biquad filter sutrctures for application to digital parametric equalization," in *AES 105th Convention*, San Francisco, USA, September 1998.

[56] C. Asavathiratham, P. E. Beckmann, and A. V. Oppenheim, "Frequency warping in the design and implementtion of fixed-point audio equalizers," in *Proc. WASPAA'99*, New Paltz, New York, October 1999, IEEE, pp. 55–59.

[57] M. Karjalainen, E. Piirila, A. Järvinen, and J. Huopaniemi, "Comparison of loudspeaker equalization methods based on DSP techniques," *J. Audio Eng. Soc.*, vol. 47, no. 1/2, pp. 15–31, January/February 1999.

[58] J. A. Pedersen, P. Rubak, and M. Tyril, "Digital filters for low frequency equalization," in *AES 106th Convention*, Munich, Germany, May 1999.

[59] A. Torger, "nwfiir – a software equalizer for high resolution digital audio," http://www.ludd.luth.se/~torger/filter.html, November 1999.

[60] M. Karjalainen, "Warped filter design for the body modeling and sound synthesis of string instruments," in *Proc. Nordic Acoustical Meeting (NAM'96)*, helsinki, Finland, June 1996, pp. 445–453.

[61] Jyri Huopaniemi, Nick Zacharov, and Matti Karjalainen, "Objective and subjective evaluation of head-related transfer function filter design," vol. 47, no. 4, pp. 218–239, April 1999.

[62] J. Huopaniemi and M. Karjalainen, "Review of digital design and implementation methods for 3-D sound," in *102nd AES Convention preprint 4461*, Munich, Germany, March 1997.

[63] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," in *98th AES Convention preprint 3980*, Paris, France, February 1995.

[64] Jyri Huopaniemi, *Virtual acoustics and 3-D sound in multimedia signal processing*, Ph.D. thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 1999.

[65] S. Van Duyne and J. O. Smith, "Physical modeling with the 2-D digital waveguide mesh," in *Proc. Int. Computer Music Conf. (ICMC'93)*, Tokyo, Japan, Sept. 1993, pp. 40–47.

[66] L. Savioja, T. Rinne, and T. Takala, "Simulation of room acoustics with a 3-D finite difference mesh," in *Proc. Int. Computer Music Conf. (ICMC'94)*, Aarhus, Denmark, 12-17 Sept. 1994, pp. 463–466.

[67] F. Fontana and D. Rocchesso, "A new formulation of the 2D-waveguide mesh for percussion instruments," in *Proc. XI Colloquium on Musical Informatics*, Bologna, Italy, 8-11 Nov. 1995, pp. 27–30.

[68] S. Van Duyne and J. O. Smith, "The 3D tetrahedral digital waveguide mesh with musical applications," in *Proc. Int. Computer Music Conf. (ICMC'96)*, Hong Kong, 19-24 Aug. 1996, pp. 9–16.

[69] F. Fontana and D. Rocchesso, "Physical modeling of membranes for percussion instruments," *Acustica united with Acta Acustica*, vol. 84, no. 3, pp. 529–542, May/June 1998.

[70] L. Savioja and V. Välimäki, "Improved discrete-time modeling of multi-dimensional wave propagation using the interpolated digital waveguide mesh," in *Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP'97)*, Munich, Germany, 19-24 April 1997, vol. 1, pp. 459–462.

[71] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay – tools for fractional delay filter design," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30–60, Jan. 1996.

[72] L. Savioja and V. Välimäki, "Reduction of the dispersion error in the triangular digital waveguide mesh using frequency warping," *IEEE Signal Processing Letters*, vol. 6, no. 3, pp. 58–60, March 1999.

[73] L. Savioja and V. Välimäki, "Reduction of the dispersion error in the interpolated digital waveguide mesh using frequency warping," in *Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP'99)*, Phoenix, Arizona, 15-19 March 1999, vol. 2, pp. 973–976.

[74] K Tokuda, T Kobayashi, S Imai, and T Chiba, "Spectral estimation of speech by mel-generalized cepstral analysis," *Electr. and Comm. in Japan, Part 3*, vol. 76, no. 2, pp. 30–43, 1993.

[75] T Kobayashi and S Imai, "Spectral analysis using generalized cepstrum," *IEEE Trans. Acoust. Speech, and Signal Proc.*, vol. ASSP-32, no. 5, pp. 1087–1089, October 1984.

[76] K Tokuda, T Masuko, and K Koishida, "Speech signal processing toolkit ver 1.0," http://kt-lab.ics.nitech.ac.jp/~tokuda/SPTK/, 1998.

[77] H. Noda, "Frequency-warped spectral distance measures for speaker verification in noise," in *IEEE Int. Conf. on Acouctics, Speech, and Signal Processing*, New York, 1988, IEEE, pp. 576–579.

[78] S. Wang, A. Sekey, and A. Gersho, "Auditory distortion measure for speech coding," in *IEEE Int. Conf. on Acouctics, Speech, and Signal Processing*. IEEE, 1991, vol. 1, pp. 493–496.

[79] B. Krzysztof and A. P. Alexander, "Speech enhancement system for hands-free telephone based on the psychoacoustically motivated filter bank with allpass frequency transformation," in *Proc. Eurospeech'99*, Budapest, Hungary, September 1999, vol. 6, pp. 2555–2558.

[80] M. Karjalainen, T. Altosaar, and M. Vainio, "Speech synthesis using warped linear prediction and neural networks," in *IEEE Int. Conf. on Acouctics, Speech, and Signal Processing*, Seattle, USA, May 1998, IEEE, vol. 2, pp. 877–880.

5 Figures

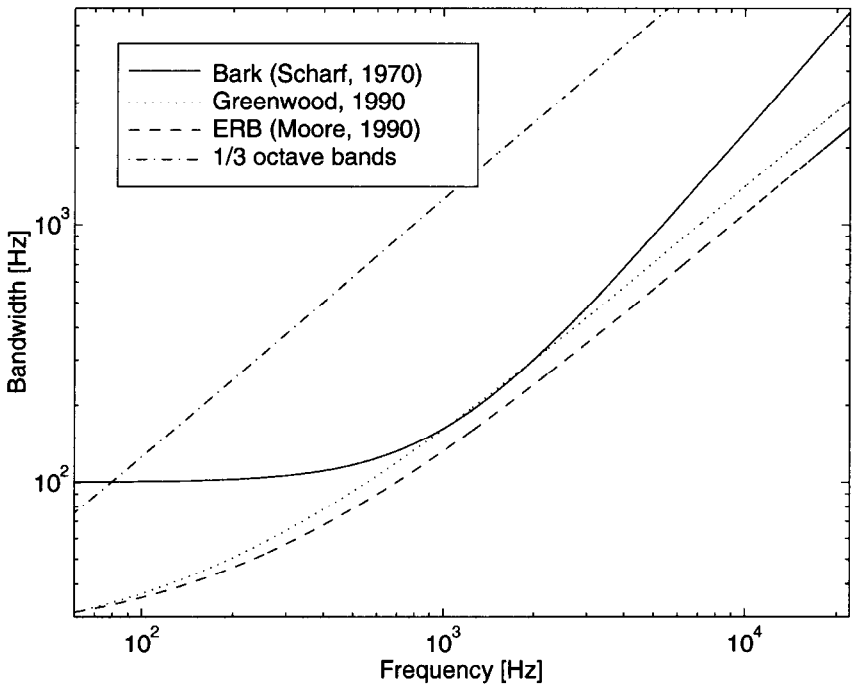

Figure 1: A bandwidth of 1 mm on the cochlea derived from Greenwood's formula, ERB's, Bark's, and third-octave bands as a function of center frequency.

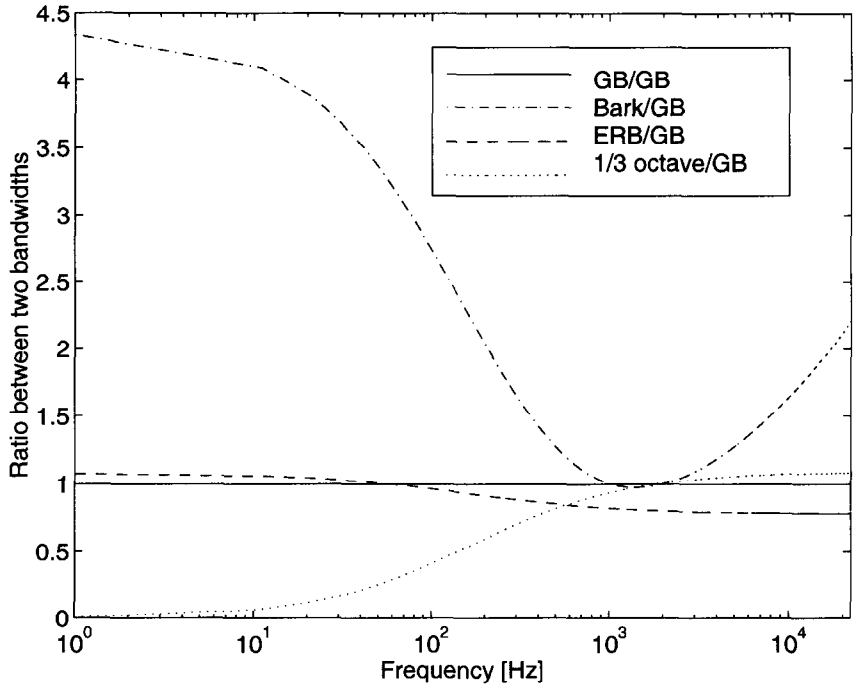

Figure 2: The differences between estimates on the bandwidth of an auditory filter illustrated in terms of the ratio between a bandwidth function and the bandwidth derived from Greenwood's mapping. The latter is given by (3) and denoted by GB in the sub-figure.

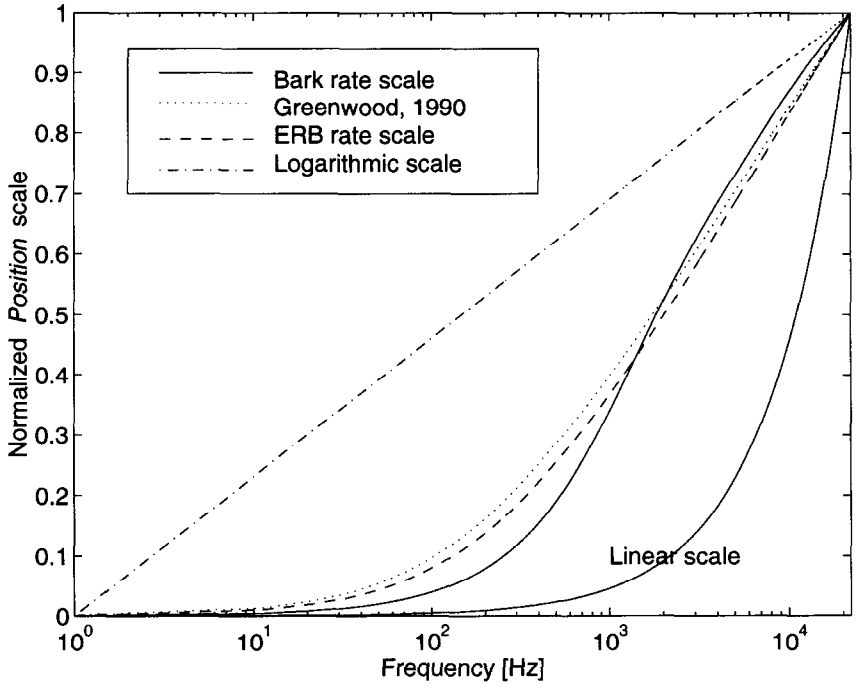

Figure 3: Greenwood's expression, ERB rate scale, Rark rate scale, logarithmic, and linear frequency mappings.

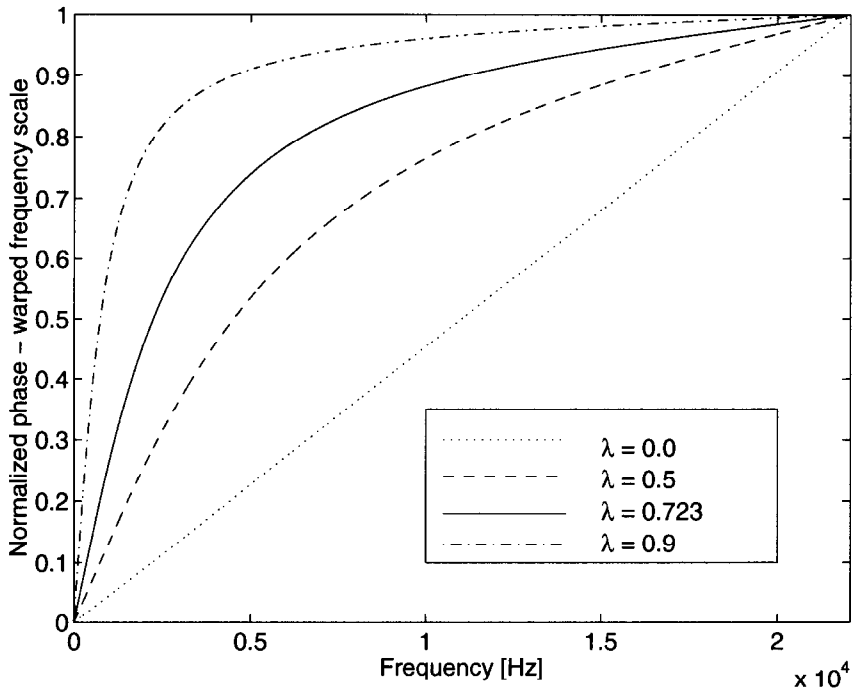

Figure 4: Phase response of a first-order AP-filter for several values of λ.

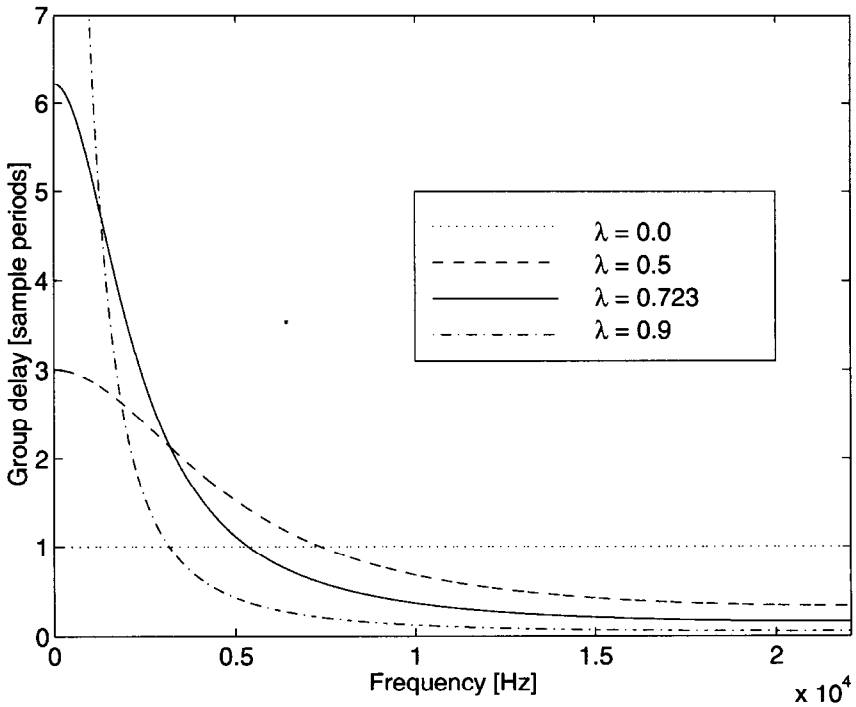

Figure 5: Group delay for positive values of λ. For negative values the curves are mirror images of those curves.

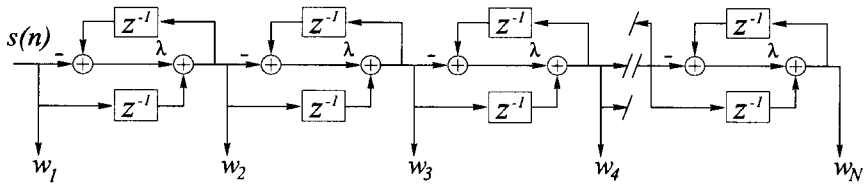

Figure 6: A direct form II implementation of an AP-chain.

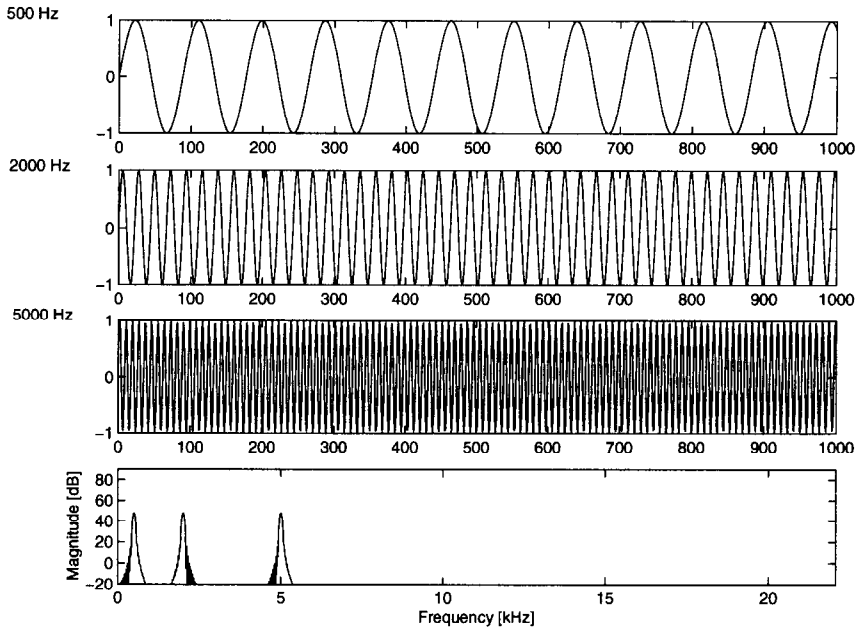

Figure 7: A set of sinusoidal signals and their spectra.

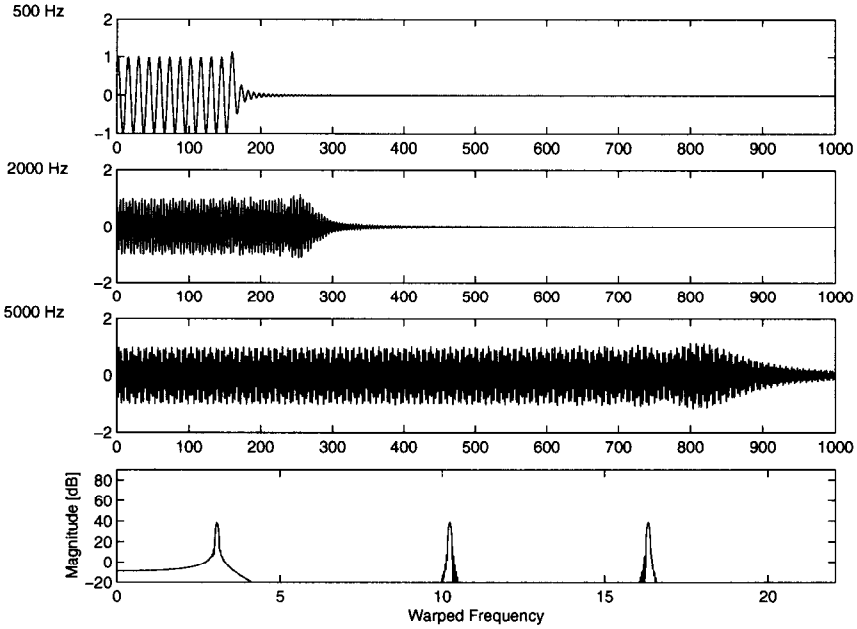

Figure 8: Warped signals and their spectra.

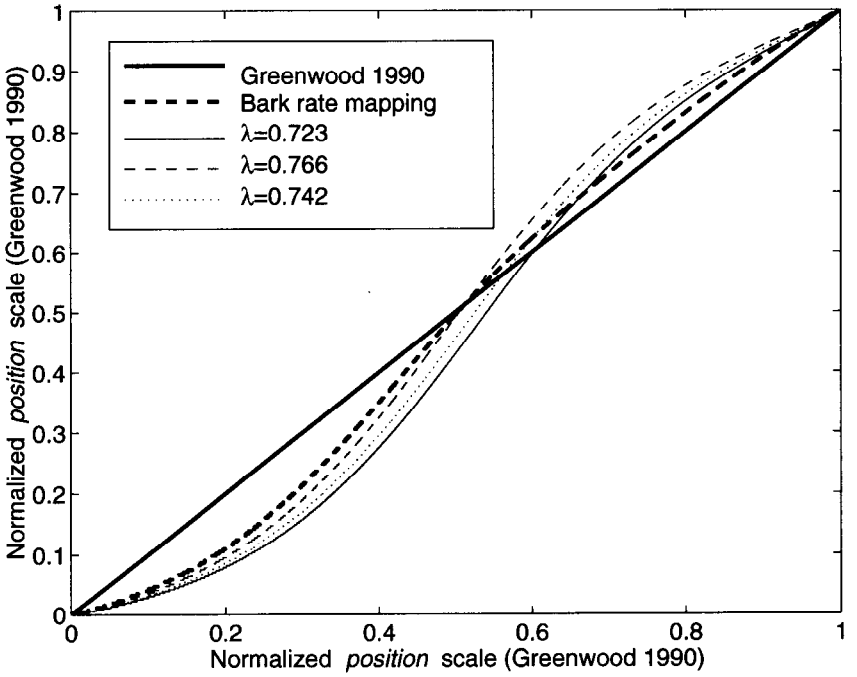

Figure 9: Bark rate scale mapping and frequency transformation occuring in an AP chain for some values of λ. In the figure, both the x and y-axis represent normalized auditory frequency scale from Greenwood's expression.

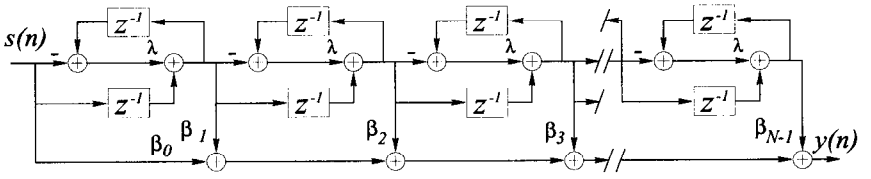

Figure 10: A warped FIR type filter where the unit delays of a conventional filter are replaced with first order allpass filters $A(z)$. The name FIR (Finite Impulse Response) is used to illustrate the structural similarity with a non-warped FIR filter. Naturally, the filter has an infinite impulse response because $A(z)$ is an IIR filter.

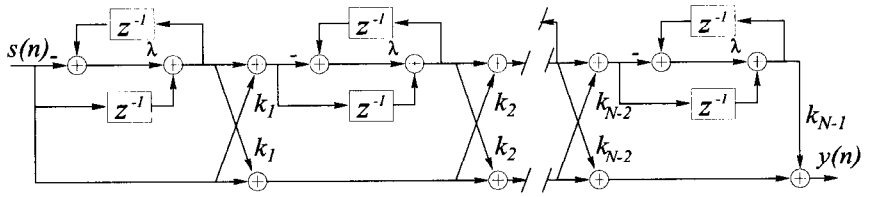

Figure 11: A warped FIR lattice filter.

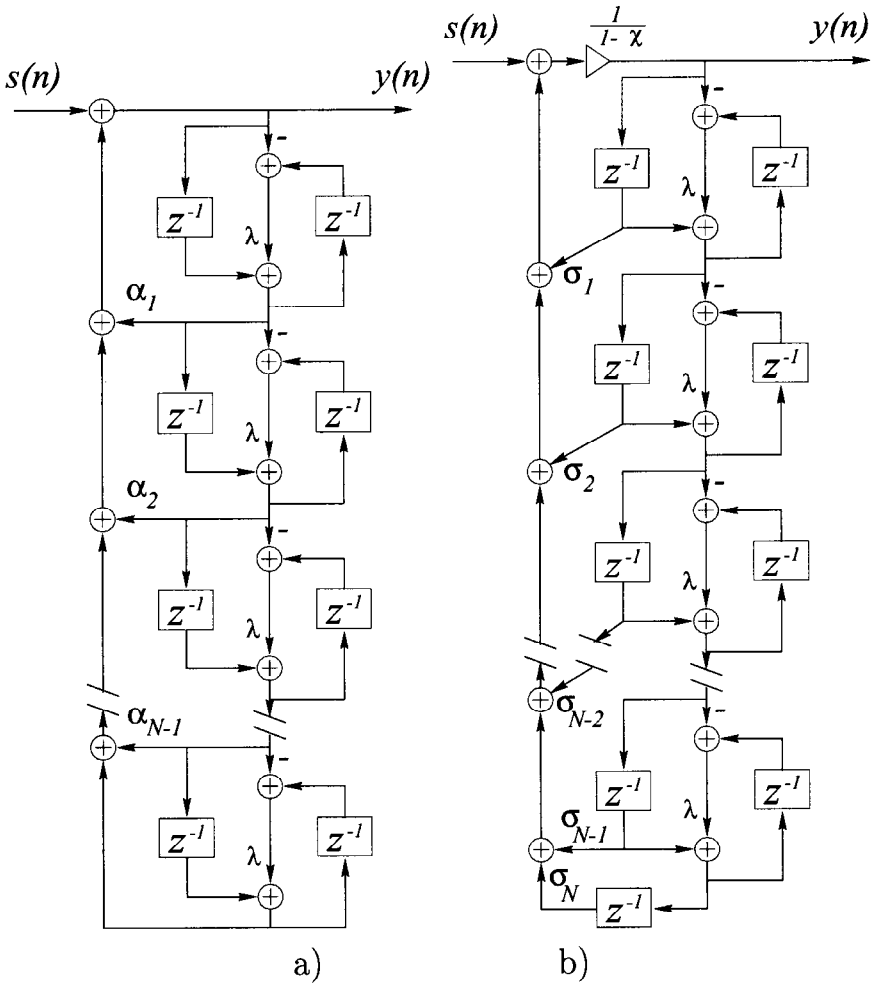

Figure 12: a) A warped IIR filter. This filter cannot be implemented directly because it has delay-free recursive loops. b) A directly realizable warped IIR filter.

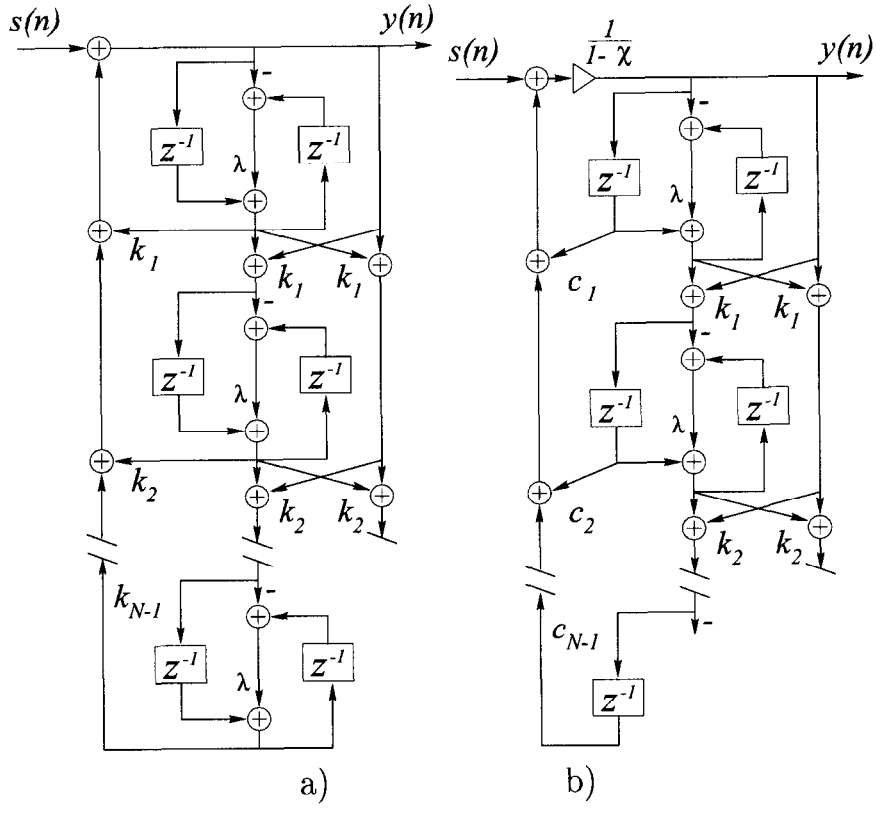

Figure 13: a) A warped IIR lattice filter. This structure cannot be implemented directly because it contains delay-free recursive loops. b) A directly realizable warped lattice IIR filter.

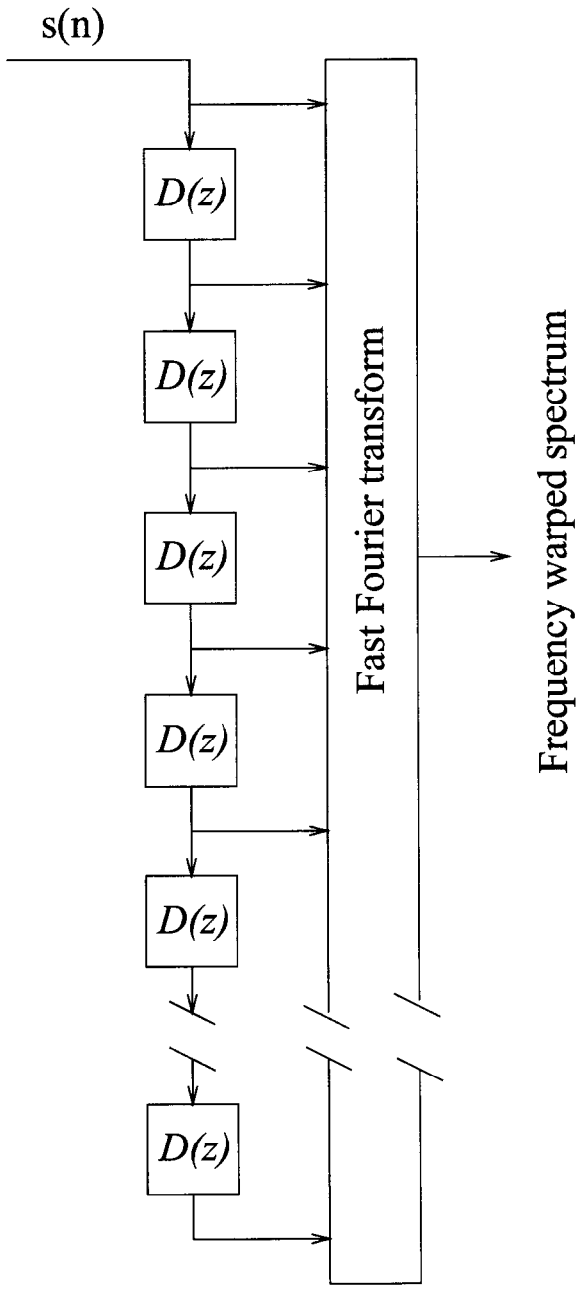

Figure 14: A network for computing warped FFT spectra.

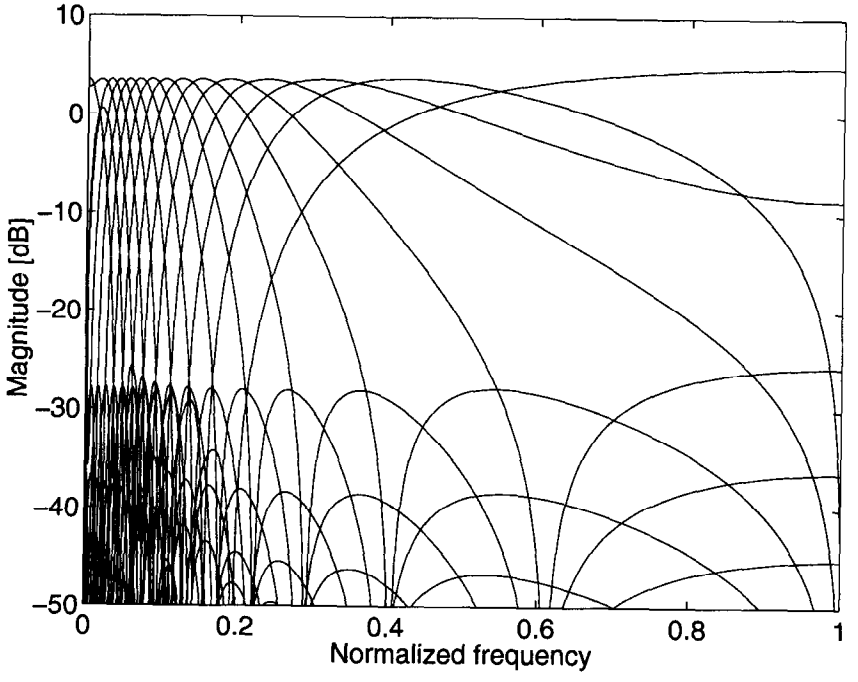

Figure 15: An example of frequency responses of a 16 channel warped filterbank. The filterbank was designed using an AP filter chain of 16 elements and hamming window was used before the FFT.

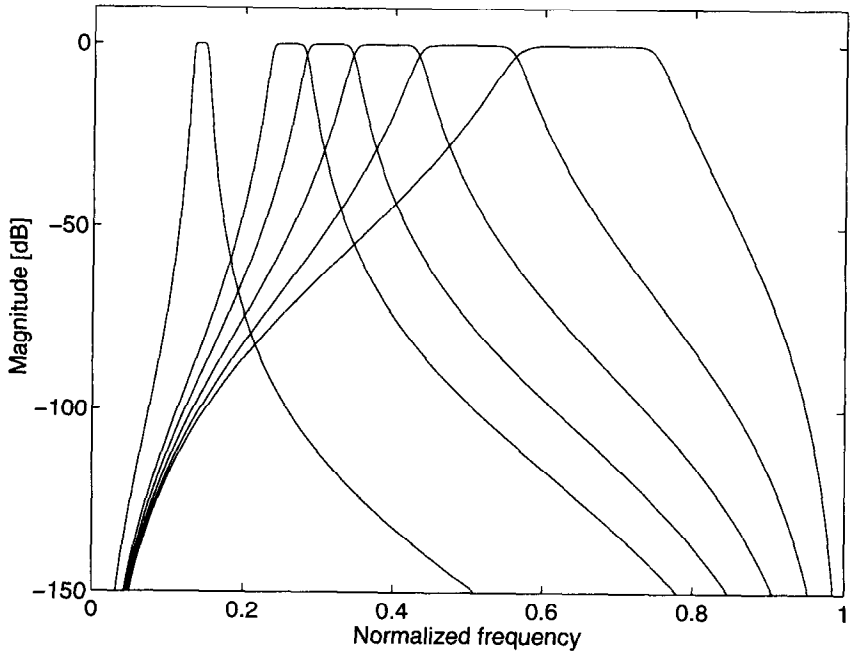

Figure 16: A design example. A filterbank consisting of rectangular fifth-order warped Butterworth's bandpass filters each having the bandwidth of one Bark. Only the channels 10, and 19 – 23 are plotted.

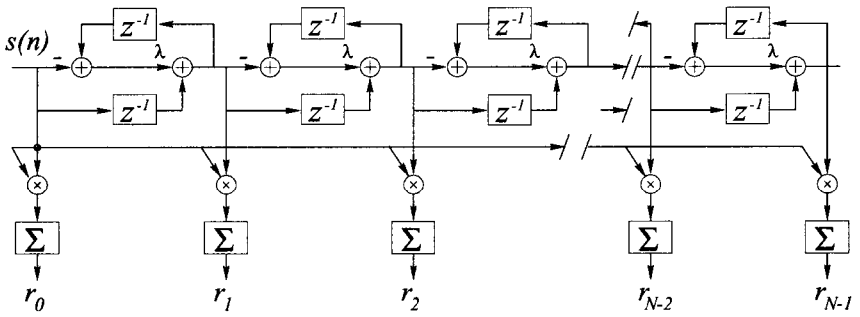

Figure 17: A warped autocorrelation network. This computes an N -tap warped autocorrelation values continuously from the input sequence $s(n)$.

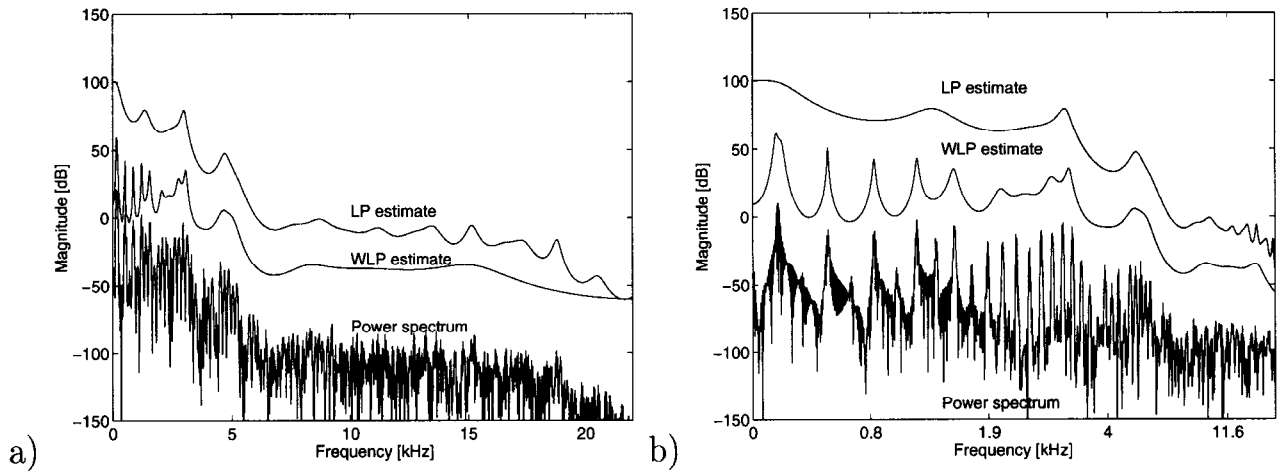

Figure 18: Power spectrum of a musical signal (clarinet) and power spectral estimates given by a conventional and a warped 40th order LPC model on a) linear and b) warped frequency scale.

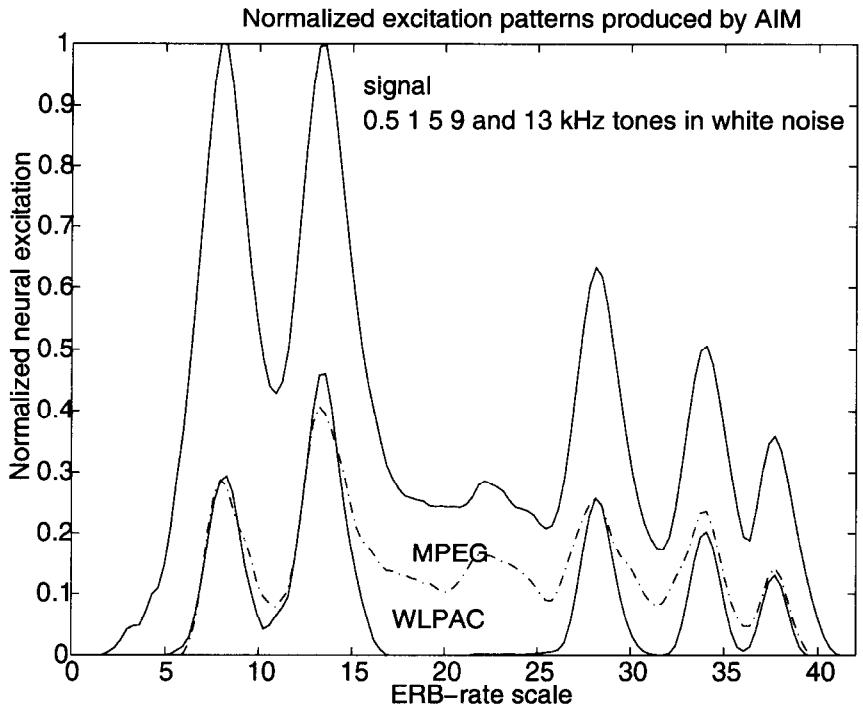

Figure 19: Noise processes in WLP coding. The original signal consist of a set of sinusoids and white background noise. The spectral shape of the error has the same shape as the original signal due to open loop coding scheme. The corresponding coding error signal in an MPEG I *layer* 3 codec is plotted on the same figure for comparison (dashed-dotted curve).

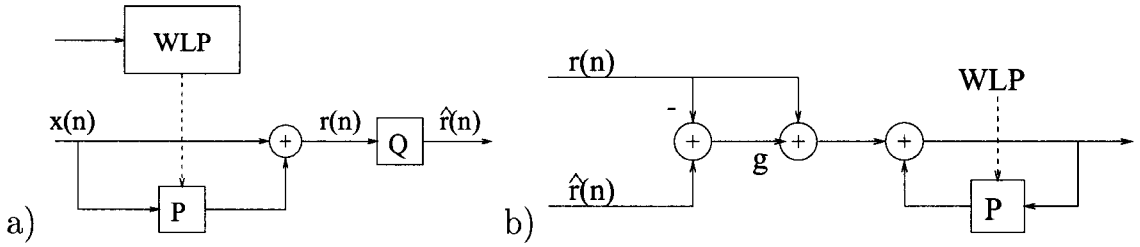

Figure 20: A simulated generalized WLPC based a) encoder and b) decoder used in listening tests.

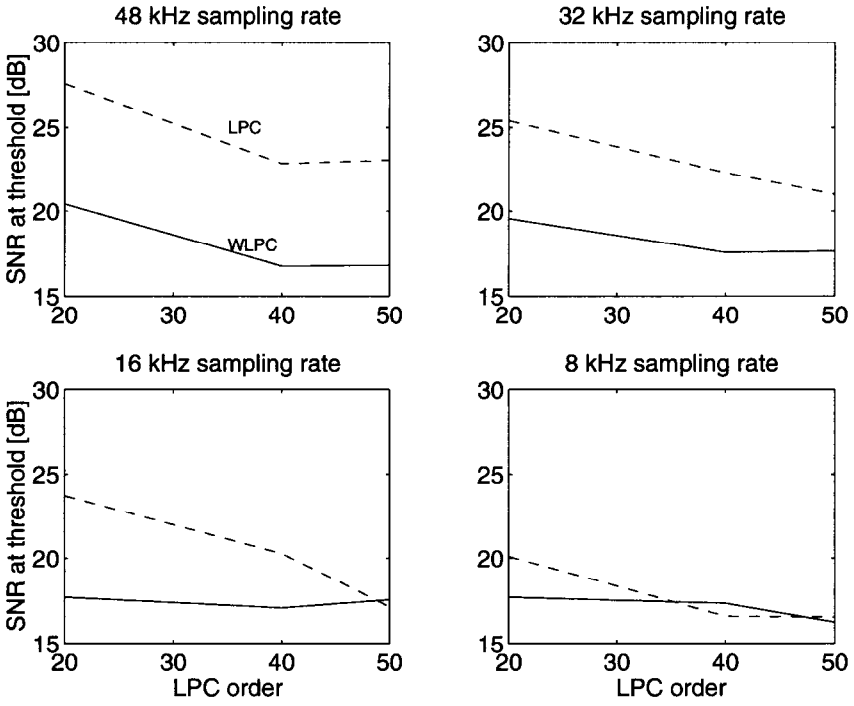

Figure 21: Preliminary listening test results at four different sampling rates as a function of the order of a WLP of LPC filter. Only average data over 11 test signals and two subjects are shown in the figures.

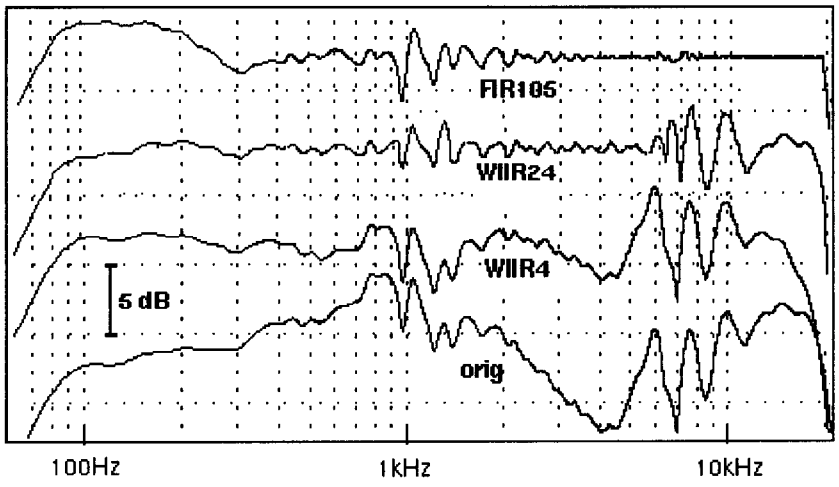

Figure 22: Loudspeaker equalization curves; *orig*: original magnitude response, *WIIR4*: warped IIR equalization, filter order 4, *WIIR24*: warped IIR order 24, *FIR105*: FIR filter equalization, filter order 105.

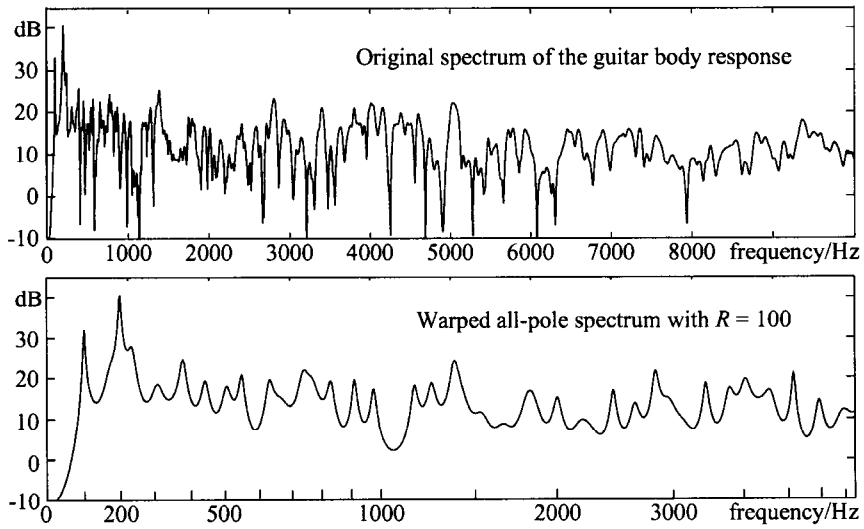

Figure 23: Modeling of the guitar body response: a) original magnitude response, b) magnitude response using WIIR modeling (notice the warped frequency scale).

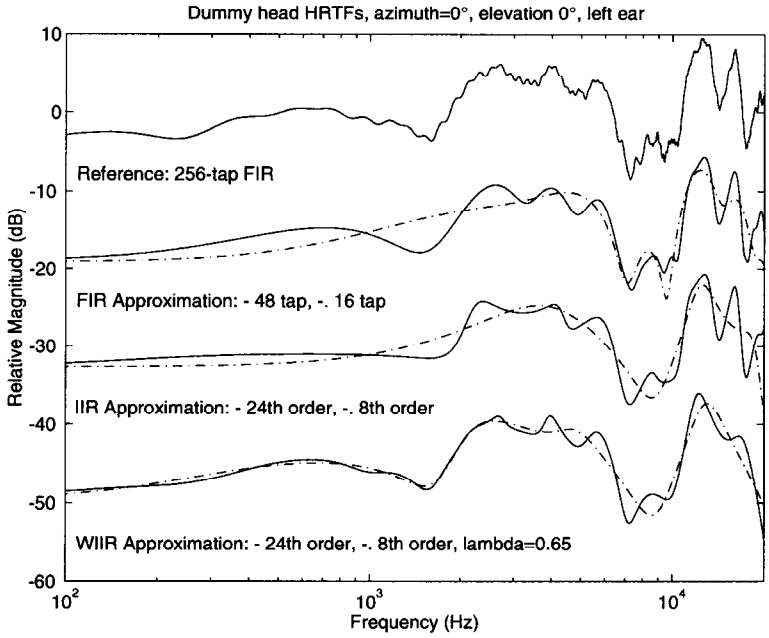

Figure 24: HRTF filter design for dummy head measurement. Frequency-sampling FIR design, Prony's method IIR and WIIR designs.

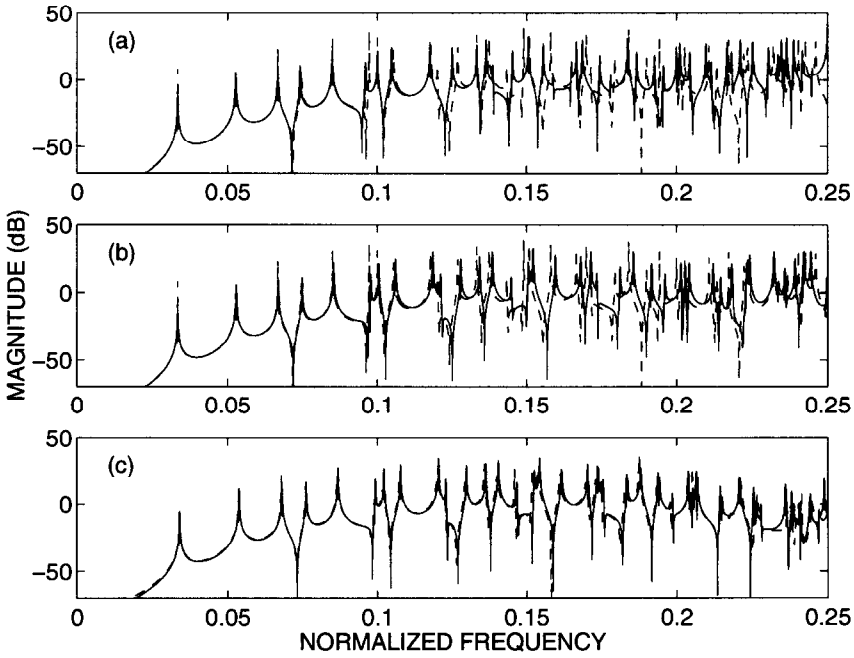

Figure 25: Analytically solved magnitude spectrum of an ideal rectangular membrane (dashed lines) together with the simulation results obtained with the (a) original, (b) the warped optimally interpolated, and (c) the warped triangular mesh.

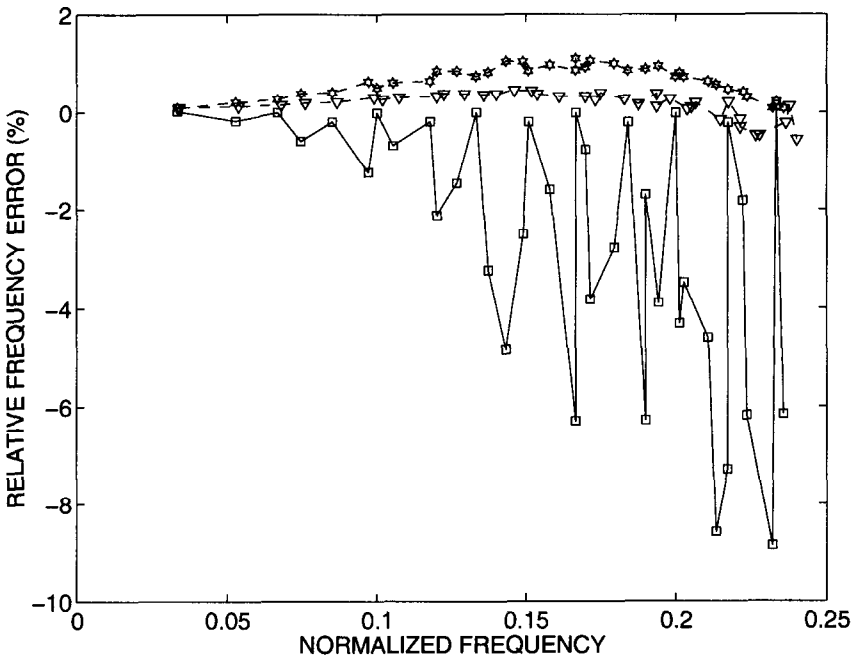

Figure 26: Error in eigenmode frequencies of the simulated rectangular membrane ($\square$ = original mesh, $\star$ = warped interpolated mesh, $\triangledown$ =warped triangular mesh).