Arbitrary Audio FIR Filter Design by Bode Plot Smoothing using Tuneable Approximate Piecewise Linear Regression

Anthony Zaknich (1), Gareth E Lee (2)

(1) School of Electrical, Energy and Process Engineering, Murdoch University, Murdoch 6150, Australia
(2) School of Electrical, Energy and Process Engineering, Murdoch University, Murdoch 6150, Australia

ABSTRACT

A method for the design of arbitrary minimum or linear phase FIR filters is developed for audio applications by using the Tuneable Approximate Piecewise Linear Regression (TAPLR) method to smooth the required FIR magnitude frequency response from a prototype Bode plot model. The TAPLR method incorporates a set of contiguous piecewise linear (affine) sections, which are coupled and smoothed by a single tuning parameter associated with a set of amplitude weighted Radial Basis Functions (RBFs) assigned to each linear section. The Bode plot also consists of a set of contiguous linear asymptotes plotted on a log-log scale, which makes it a perfect candidate for modelling and smoothing by the TAPLR method. The TAPLR smoothing turns the artificial asymptotic magnitude plot into a realisable magnitude response curve, which can be made to be band-limited with a finite impulse response by the appropriate degree of smoothing. A typical FIR filter design example for audio system equalisation is provided to show the value and versatility of the method. Also, two Bode plot filter prototypes are presented to show how well the new modelling approach can capture them and adapt them to suitable band-limited FIR filter designs.

INTRODUCTION

One approach to arbitrary Finite Impulse Response (FIR) filter design is to specify the required magnitude response characteristic and then to use it to derive suitable FIR filter coefficients that provide a magnitude response sufficiently close to the original specification. This can be done by either finite length windowing of the corresponding discrete impulse response or by making the magnitude response band-limited and sufficiently smooth in the first place, in which case the discrete impulse response can be taken as the FIR filter coefficients directly. Finite length windowing of the impulse response achieves the required finite impulse response, which also inevitably smooths the required magnitude response characteristic. Both of these approaches allow for the design of either linear phase or minimum phase FIR filters.

Minimum phase discrete-time filters have the property that all their poles and zeros are within the unit circle of the complex z-domain. The phase characteristic of a minimum phase filter is fully determined by the magnitude characteristic as these two characteristics are Hilbert transform pairs. A linear phase filter is one whose phase characteristic is a linear function of frequency and is determined by a symmetrical impulse response. It is easy enough to ensure that the required impulse response is made to be symmetrical during the design process. Therefore, it is possible to produce a linear phase or a minimum phase filter having the same magnitude response characteristic.

It must be accepted that it may not be possible to design a FIR filter from any arbitrarily chosen response characteristic because a characteristic with sharp transition sections may strictly require an infinite impulse response to represent accurately. The required or desired magnitude response characteristic therefore needs to not only be band-limited to at least half the Nyquist sampling frequency but it must also be smooth enough to ensure that the filter is realisable as a FIR filter. A sufficiently smooth band-limited response curve can then be used to determine the required discrete finite impulse response (FIR filter coefficients) directly, by taking a sampled inverse Fourier transform of it.

It is proposed that suitable FIR filter designs may be achieved starting from a traditional Bode plot response characteristic, which is a set of contiguous straight lines on a log-log scale. The sharp transitions due to these connecting straight lines would make the actual Bode response plot unrealisable as a FIR filter, so it needs to be smoothed out sufficiently at these sharp joins. This may be very conveniently achieved by using the Radial Basis Function (RBF) based smoothing and interpolation mechanism associated with the Tuneable Approximate Piecewise Linear Regression (TAPLR) model (Zaknich and Attikiouzel 2000). The TAPLR model is actually made up of a set of contiguous straight lines that are coupled and smoothed by a single tuning parameter, making it a perfect model for Bode plot smoothing.

The TAPLR model is an interpolation model and the precursor to the Integrated Sensory Intelligent System (ISIS) (Zaknich 2003a) and the Sub-Space Adaptive Filter (SSAF) (Zaknich 2003b) models. All these models were originally derived from the Modified Probabilistic Neural Network (MPNN) (Zaknich 1998). The MPNN uses RBFs to interpolate over the data space for the solution of multivariable nonlinear regression problems. It is based on Specht's Probabilistic Neural Network (Specht 1990) and it can be described as a generalization of Specht's General Regression Neural Network (GRNN) (Specht 1991). Both the MPNN and GRNN are similar to the Nadaraya-Watson regression estimator (Naydaraya 1964, Watson 1964).

A review of the generic TAPLR model is presented in the second (next) section of this paper. Section three introduces the details of the frequency domain TAPLR model used for Bode plot modelling, whilst Section four shows how this model is used to achieve FIR filter designs. Some typical design examples are provided in Section five and Section six provides an analysis of results and draws some conclusions about the new design method.

REVIEW OF THE GENERAL TAPLR MODEL

The general MPNN and TAPLR models were both designed to provide a scalar regression output given a set of input(vector)/output(scalar) data pairs. The TAPLR model equation is defined by (1).

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{M} Z_i l_i(\mathbf{x}) f_i(\|\mathbf{x}-\mathbf{c}_i\|,\sigma)}{\sum_{i=1}^{M} Z_i f_i(\|\mathbf{x}-\mathbf{c}_i\|,\sigma)} \quad (1)$$

where:

$f_i(\|\mathbf{x}-\mathbf{c}_i\|,\sigma)$; a suitable RBF, typically Gaussian.
$\mathbf{x}$; an arbitrary p-dimensional input vector.
$\mathbf{c}_i$; the RBF centre-vector i in the input space.
σ ; the single smoothing parameter for model tuning.
M ; the number of unique centre-vectors $\mathbf{c}_i$.
$l_i(\mathbf{x})$; local linear (affine) function output associated with RBF i and its corresponding centre-vector $\mathbf{c}_i$.
Z_i ; a scalar weight associated with RBF i.

If all the local linear models $l_i(\mathbf{x})$ are equally likely, or the *a priori* likelihood is unknown, and the centres of the RBFs are uniformly distributed in the data space (as will be the case for the work in this paper), then all the Z_i may be ignored (all Z_i = 1). Otherwise, the relative values of Z_i might represent the *a priori* likelihood or any required relative weighting of each $l_i(\mathbf{x})$ as appropriate. The functional values provided by each RBF weighted by Z_i, i.e. $Z_i f_i(\|\mathbf{x}-\mathbf{c}_i\|,\sigma)$, are used as a measure of closeness of $\mathbf{x}$ to each of the $l_i(\mathbf{x})$ models, and thus provide the required relative local linear model weightings for suitable interpolation between them. The TAPLR model (1) is essentially a mixture model and, as such, further piecewise linear models can be added to the structure to accommodate required design specifications. The degree of local model coupling or decoupling can be controlled by the adjustment of the single RBF bandwidth or smoothing parameter σ.

A Gaussian radial (spherical) basis function, defined by (2), is often used for the RBF $f_i(\|\mathbf{x}-\mathbf{c}_i\|,\sigma)$. Adjustment of σ controls the degree of weighting of each linear local model associated with each centre-vector. Input vectors $\mathbf{x}$ closest to a centre-vector activate the associated local linear model more than for those further away. For very small σ the local linear model associated with the centre-vector closest to the current input vector dominates, resulting in a linear response in the local space of that centre-vector. For very large σ the model output approaches a fixed weighted average of all the local linear models. Somewhere in between a best overall model results, which provides linear operation close to each centre-vector while deviating from linearity close to centre-vector region boundaries. At the boundaries between centre-vectors a smooth and continuous merging of neighbouring local linear models occurs.

$$f_i(\|\mathbf{x}-\mathbf{c}_i\|,\sigma) = \exp\frac{-(\mathbf{x}-\mathbf{c}_i)^T(\mathbf{x}-\mathbf{c}_i)}{2\sigma^2} \quad (2)$$

Overall model tuning simply involves finding the single optimal tuning parameter σ giving the minimum Mean Squared Error (MSE) of the output $\hat{y}(\mathbf{x}_k)$ minus the corresponding desired output y_k for a representative testing set of known sample vector pairs $\{(\mathbf{x}_k, y_k)| \, k=1,...,NUM\}$. In typical applications there is a unique σ that produces the minimum MSE between the model (1) output and the desired output for the testing set, and it can be found quite easily by trial and error. The relation between σ and MSE is usually smooth with a fairly broad minimal MSE section. Consequently, tuning is not critical to achieve an adequate regression model $\hat{y}(\mathbf{x})$.

The TAPLR model (1) can be usefully seen as a method of MPNN network size reduction. The local linear models $l_i(\mathbf{x})$ in (1) simply replace fixed training values y_i in the MPNN. Another benefit of having the $l_i(\mathbf{x})$ is that it allows for a more sensible extrapolation of the model into the data space for which there is no training data. Although it may be incorrect in specific cases, a linear extrapolation may be the best guess when nothing else is known, especially for short excursions. Finally, the model (1) structure controlled by the weighted RBFs provides a way of decoupling the local linear models $l_i(\mathbf{x})$ from each other while smoothly merging them at their adjoining boundaries.

THE FREQUENCY DOMAIN TAPLR MODEL

The TAPLR model (1) is a general multivariate regression equation that can be used for arbitrary applications in either the time or frequency domains. In the context of FIR filter or system design from Bode plots (1) can be easily modified for log-log continuous-frequency domain scalar function modelling, as defined by (3), by adopting the following parameter set equivalents:

$\mathbf{x} = \log\omega$; the log of arbitrary frequency variable ω.

$\mathbf{c}_i = \log\omega_i$; ith log of local frequency centre variable ω_i.

$\omega_i = \omega^0 \left[(2)^{\frac{1}{m}} \right]^{i-1}$, $i = 1,..,M_\omega$, $m = 12$, which makes the frequency variables $\log\omega_i$ uniformly spaced, therefore Z_i can be ignored since it is a constant, i.e.,

$$Z_i = \log\omega_i - \log\omega_{i-1} = \log\omega_{k+i} - \log\omega_{k+i-1} = 0.025$$

$$20\log\left|\hat{H}(\omega)\right| = \frac{\sum_{i=0}^{M_\omega} l_i(\log\omega) f_i(\|\log\omega - \log\omega_i\|,\sigma_\omega)}{\sum_{i=0}^{M_\omega} f_i(\|\log\omega - \log\omega_i\|,\sigma_\omega)} \quad (3)$$

where:

$|\hat{H}(\omega)|$; the magnitude filter response model estimate.

ω_i ; the i^{th} $1/m^{th}$ octave centre frequency, where $i = 0$ is predefined for 0 Hz and ω^0 is the first non-zero (starting) frequency ω_1.

$|H_i(\omega_i)|$; ith linear function response magnitude at ω_i.

$l_i(\log\omega) = 20\left[n_i \log\frac{\omega}{\omega_i} + \log|H_i(\omega_i)| \right]$; ith linear fn.

n_i ; order of ith linear fn (Bode plot section), is (negative) for positive(negative) slope.

$f_i(\| \log\omega - \log\omega_i \|, \sigma_\omega)$; a suitable RBF centred at $\log\omega_i$, typically Gaussian, i.e., (2).

σ_ω ; the single smoothing parameter for the variable $\log\omega$.

M_ω; the total number of non-zero reference centre frequencies ω_i & local linear functions $l_i(\log\omega)$.

The dB magnitude frequency response model estimate (3) is based on an exponential nonuniform frequency spacing for the ω_i frequencies. In audio applications this is typically a fractional octave spacing. For example, starting with a standard frequency of ω^0 =15.84893192 Hz (Davis and Davis 1977) and a $1/12^{th}$ octave spacing the 128th non-zero frequency would be ω_{128} =23714 Hz. These are the frequencies adopted for this paper and when expressed as logarithms ($\log\omega_i$) as is done for (3), these frequency points are uniformly spaced. This is seen on typical audio frequency response and Bode plots, where the magnitude is also represented in a logarithmic dB scale.

The frequency response model estimate when defined according to (3) using a logarithmic frequency variable will be seen as a set of contiguous line segments, representing a Bode plot, for $\sigma_\omega \to 0$. However, as σ_ω is increased the estimate becomes more heavily smoothed and thus approaches a real response curve, where the Bode plot line segments will appear like asymptotic sections with respect to it.

The exponential fractional octave frequency spacing used here is justified by the fact that many real systems, such as loudspeakers, are typically composed of multiple resonators all of which have approximately the same *Q*, or energy decay characteristics per cycle of decay (Keele 2004, Keele 1994). Consequently, relatively more frequency resolution is required at lower frequencies than at higher frequencies to fully capture the response characteristic without any loss of information. This means that (3) is an efficient response model structure provided that the appropriate fractional octave spacing is chosen, for a given system under test, according to its highest *Q* section.

All design data and test data in this paper have been time collected at a sample rate of 48,000 Hz, on the assumption that this new system model will be used for problems related to typical audio systems. A $1/12^{th}$ octave frequency spacing with 128 non-zero frequency references ω_i as described above, have been chosen to provide useful practical examples. During the design process the reference samples $|H_i(\omega_i)|$ are either directly sampled from a real system or specified as required. The local linear functions $l_i(\log\omega)$ each pass through the values $|H_i(\omega_i)|$ and are assigned a suitable slope by choosing the sign and magnitude of the local linear function order variables n_i (order ≡ poles), where n_i need not be restricted to integer values.

THE NEW TAPLR FIR FILTER DESIGN METHOD

The new FIR filter design method is based on the idea that the desired magnitude response can be made sufficiently smooth by an appropriate choice of σ_ω in (3) such that the resulting impulse response decays fast enough allowing it to be truncated without explicit windowing. This is effectively a band-limited interpolation of the frequency response function.

Given a TAPLR (3) model with an appropriate choice of σ_ω it can be used to provide, say, 2048 uniformly frequency sampled magnitude values. These are placed in order in a real buffer 2048 points long with zero values placed in the corresponding 2048 point imaginary buffer. The buffer is then extended by another 2048 points and the complex values from the first 2048 points transferred to these in reverse order such that the two halves of the full 4096 buffer are symmetrical about the centre. Next, a 4096 point discrete inverse Fast Fourier Transform (FFT) is applied to the full buffer to produce the corresponding 4096 point impulse response of the required magnitude response function. There are two problems with this impulse response. Firstly, it is acausal, being symmetrical about the first buffer point (zero time) because the FFT is cyclic or periodic. Secondly, it is too long to provide an efficient FIR filter design (the impulse response values are the required FIR filter coefficients). To design an odd length *N* (*N*<<4096) causal and efficient linear phase FIR filter all that needs to be done is to circularly shift (*N*-1)/2 points forward in the time buffer.

The TAPLR (3) FIR design method only requires taking the *N* unwindowed points uniformly sampled in a normal frequency scale ω from a sufficiently smoothed model (3), i.e., using a σ_ω that provides a band-limited smooth frequency response curve. If the frequency response samples have already been collected from a band-limited system then σ_ω can be optimised by monitoring for the minimum MSE of some other measured samples between training or reference samples $|H_i(\omega_i)|$ used to construct the TAPLR (3) model. For classical filter design σ_ω can often be chosen by trial and error according to the required or some acceptable smoothing. Otherwise, for arbitrary filter design a way to do it is to oversample during the design phase by, say, a factor of two times the required sampling rate frequency and progressively increasing σ_ω until the out of band energy (above half the require system sampling rate frequency) drops to zero or to a sufficiently small level.

SOME TYPICAL DESIGN EXAMPLES

Three typical design examples are sufficient to show the main features and benefits of the new design approach; a loudspeaker frequency response equalisation, a classical low-pass and an arbitrary Bode construction FIR filter design.

Loudspeaker Frequency Response Equaliser Example

Figure 1 shows a typical loudspeaker frequency response curve and the appropriate equaliser design curve required to flatten the speaker's response from 1000 Hz to about 20,000 Hz. The desired equaliser response was determined by taking a 0 dB flat response from 0 Hz to 1000 Hz, then from 1000 Hz to 20,000 Hz taking the negated speaker's dB response exactly, and from 20,000 Hz up taking the speaker's dB response exactly. When the equaliser is cascaded with the speaker it leaves the speaker's response as is up to 1000 Hz, gives a flat 0 dB response between 1000 Hz and 20,000 Hz, and then rolls off after that. In this example, the speaker's anechoic response below 1000 Hz is already as required because when it is set in a listing room about 0.8 m from a back wall the wall reflection of frequencies below 1000 Hz will add to the direct response to provide an approximately flat combined response from about 100 Hz to 1000 Hz.

Figure 2 shows a $N = 81$ point FIR equaliser design (TAPLR (3) with $\sigma_\omega = 0.01$) compared against the original desired equaliser response (offset by -5dB for clarity). The TAPLR design curve is almost indistinguishable (less than a fraction of a dB difference) from the original design specification and FIR filter design for $N = 81$. The TAPLR design approach is able to provide the correct degree of relative smoothing across the whole frequency range because it uses the same bandwidth RBFs, when implemented on a log scale, that are actually proportional to the exponential frequency spacings in a constant Q sense. Also, in this design much of the desired equaliser response was taken directly from loudspeaker response measurements $|H_i(\omega_i)|$ that were already band-limited, therefore the choice of $\sigma_\omega = 0.01$ was easily found to be a suitable value to preserve the required band-limiting property. Figure 3 shows the impulse response of the FIR filter having the equaliser's magnitude response and linear phase characteristics.

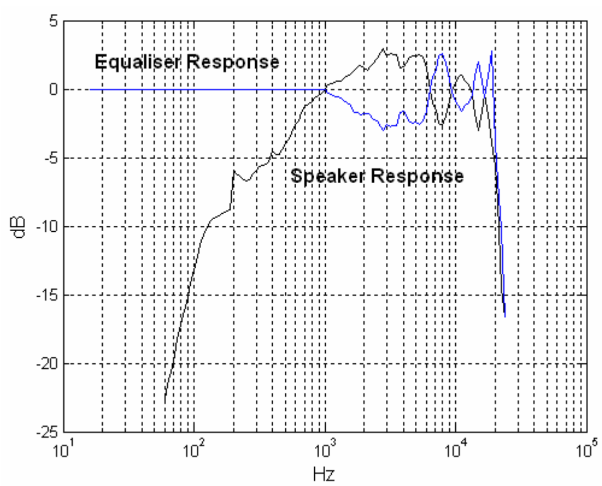

Figure 1. Loudspeaker and Corresponding Required Equaliser Responses

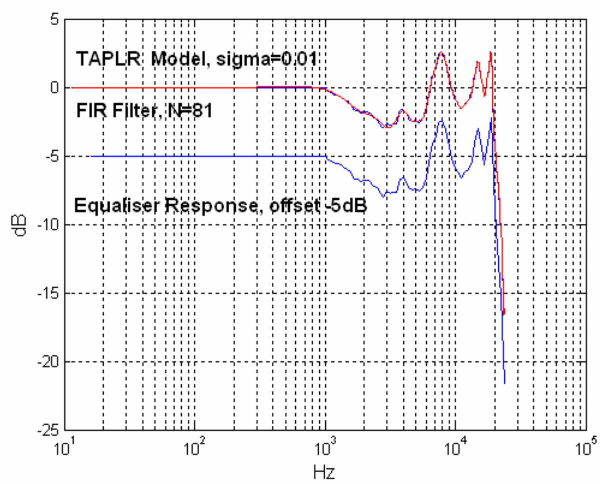

Figure 2. Loudspeaker Equaliser FIR Filter Design, $N = 81$

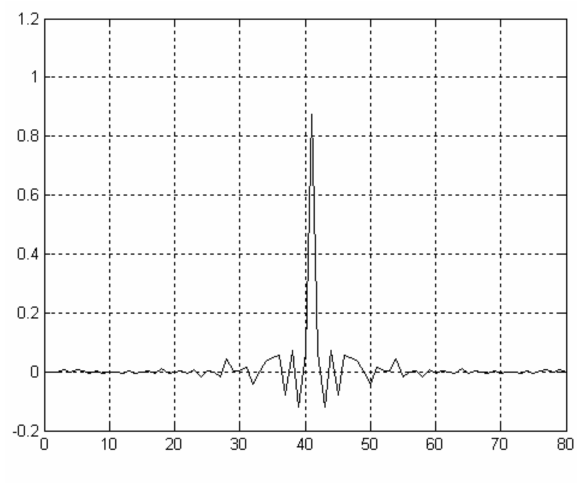

Figure 3. Loudspeaker Equaliser FIR Filter Impulse Response, $N = 81$

Classical Lowpass FIR Filter Design Example

A desired lowpass Bode response is shown in Figure 4, with a cut-off frequency at 2000 Hz and a four pole roll-off. This response has been constructed using the 128 standard $1/12^{th}$ octave frequency spacings starting at ω^0 =15.84893192 Hz and modelled using the TAPLR (3) equation. The $N = 251$ FIR filter frequency response resulting from the TAPLR (3) design prototype for $\sigma_\omega = 0.05$ is shown under the Bode plot in Figure 4. This value of σ_ω was selected to produce a -3dB cut-off at 2000 Hz, approximating a Butterworth magnitude response. The designed FIR filter and TAPLR (3) model responses were nearly indistinguishable showing that the method produces a very good design that is faithful to the design prototype provided that a sufficiently large FIR filter size N is chosen. In this example a four pole lowpass filter has been developed but any degree of pole order, including fractional values such as 4.3, can be applied with equal effectiveness.

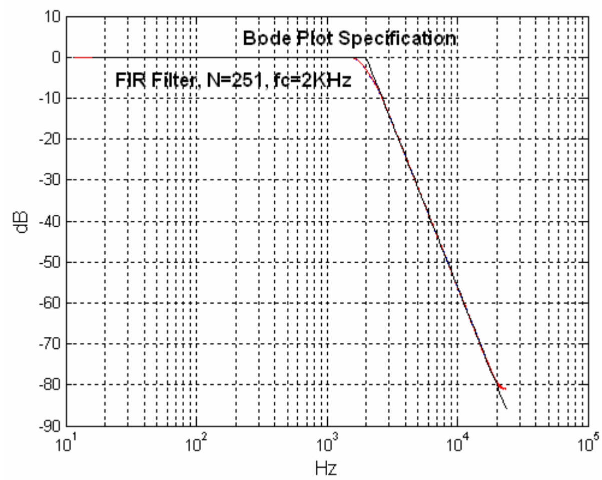

Figure 4. Lowpass FIR Filter Design, $N = 251$

Arbitrary FIR Filter Design Example

The desired Bode response of an arbitrary filter shape is shown in Figure 5, having cut-off frequencies at 1000 Hz and 2000Hz with four pole roll-offs from these. This response has also been constructed using the 128 standard $1/12^{th}$ octave frequency spacings as before and modelled using the TAPLR

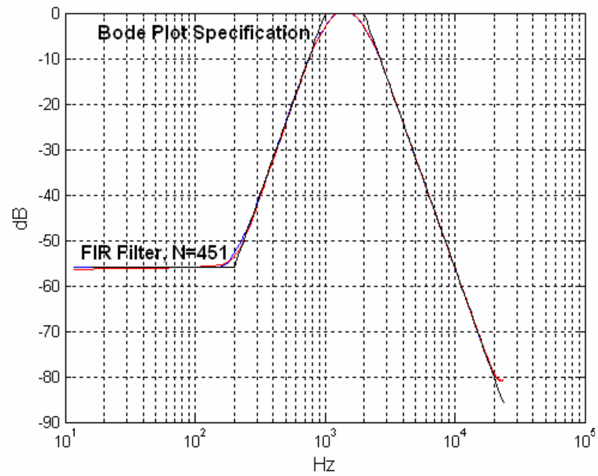

Figure 5. FIR Filter Design, $N = 451$

(3) equation. The $N = 251$ FIR filter frequency response resulting from the TAPLR (3) design prototype for $\sigma_{\omega} = 0.05$ is shown under the Bode plot in Figure 5. Once again the designed FIR filter and TAPLR (3) model responses were nearly indistinguishable.

ANALYSIS OF RESULTS AND CONCLUSIONS

The TAPLR (3) design approach allows for the control of the precise magnitude response from a desired Bode plot construction by the adjustment of a single smoothing parameter. TAPLR model (3) has been adapted from (1) to provide an excellent basis for Bode plot representation and subsequent smoothing by using an exponential frequency scale. The chosen exponential fractional octave frequency spacing is justified for audio applications because it is not only consistent with human perceptual characteristics but it also allows the frequency response to be specified in terms of constant Q bandwidth sections, which are then smoothed automatically in proportion to each other whilst still using the same single smoothing parameter.

This paper has not only shown how the TAPLR (3) model can be used to design very good arbitrary linear phase FIR filters for audio applications from prototype Bode plots but it has also introduced a straight forward method of designing an agreeable loudspeaker equaliser directly from the loudspeaker's own response measurements.

In this paper the Bode plot line segments have been allocated from either direct noiseless experimental measurements or by required design specification at standard $1/12^{th}$ octave frequency spacings. However, it is also possible to automatically find an optimal set of line and sigma parameters for the TAPLR model, given noisy data, to produce an efficient overall frequency response model. This is the subject of ongoing research not only in the context of frequency response modelling but also for more general multi-dimensional system modelling, approximation, regression, and data compression application. So far the work indicates that it is possible to produce both very accurate and efficient models based on the TAPLR structure. Furthermore, it may be possible to construct very accurate linear system transfer functions by simply noting the pole and zero positions after an automated optimal TAPLR regression is performed on the experimentally sampled data set.

REFERENCES

Zaknich, A. and Attikiouzel, Y., 2000, *A tuneable approximate piecewise linear model derived from the modified probabilistic neural network*, IEEE Signal Processing Workshop on Neural Networks for Signal Processing (NNSP), Sydney, Australia, Vol. 1, pp. 45-53.

Zaknich, A., 2003, An integrated sensory-intelligent system for underwater acoustic signal-processing applications, IEEE Journal of Oceanic Engineering, Vol. 28, No. 4, pp. 750-759.

Zaknich, A., 2003, *A practical sub-space adaptive filter*, Neural Networks, Vol. 16 , Nos 5/6, pp. 833-839.

Zaknich, A., 1998, Introduction to the modified probabilistic neural network for general signal processing applications, IEEE Transactions on Signal Processing, Vol. 46, No. 7, pp. 1980-1990.

Specht, D.F., 1990, *Probabilistic neural networks*, International Neural Network Society, Neural Networks, Vol 3, pp. 109-118.

Specht, D.F., 1991, *A general regression neural network*, IEEE Transactions on Neural Networks, Vol. 2, No. 6, pp. 568-576.

Nadaraya, E.A.,1964, *On estimating regression*, Theory of Probability and its Applications, 9, pp. 141-142.

Watson, G.S., 1964, *Smooth regression analysis*, Sankhya Series A, 26, pp. 359-372.

Davis, D. and Davis, C., 1977, *Sound system engineering*, Third Edition, Howard W. Sams and Co., Inc.

Keele, D.B., 2004, *Interpolating linear- and log-sampled convolution*, 117th Convention of the Acoustical Engineering Society, San Francisco, Preprint 6284.

Keele, D.B.,1994, *Log sampling in time and frequency: preliminary theory and application*, 97th Convention of the Acoustical Engineering Society, San Francisco, Preprint 3935.