APPLICATION OF A HUMAN AUDITORY MODEL TO LOUDNESS PERCEPTION AND HEARING COMPENSATION

Douglas M. Chabries, David V. Anderson, Thomas G. Stockham, Jr., and Richard W. Christiansen

Brigham Young University, Provo, Utah

ABSTRACT

A model is proposed which mathematically transforms an acoustic stimulus into a form which is believed to be more nearly related to that used by the auditory cortex to interpret the sound. The model is based upon recent research towards understanding the response of the human auditory system to sound stimuli. The motivation and approach in developing this model follow the philosophy pursued in the development of a similar model for the human visual system [1]. Application of this model to the problem of hearing compensation for impaired individuals is shown to yield a bank of bandpass filters each followed by a homomorphic multiplicative AGC. Clinical tests on hearing impaired subjects suggest that this approach is far superior to other current hearing compensation schemes.

1. A Process Model of the Human Auditory System

Several treatments of the auditory system describe in detail the functioning of the ear [2], [3], [4], [5]. One particularly insightful discussion is provided by Lyon and Mead [6].

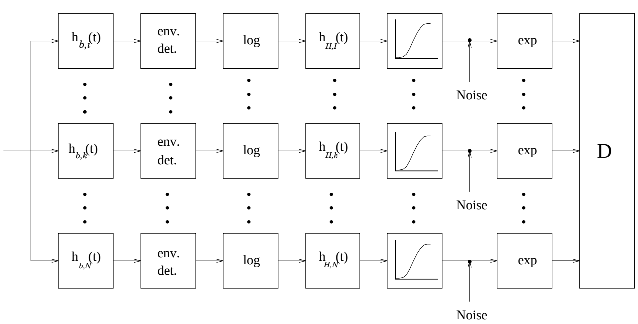

Figure 1: Loudness perception model.

A proposed model describing the human auditory response to sound and which includes the effects of the human cochlea, inner and outer hair cells, as well as loudness perception is shown in Figure 1. The proposed model performs several essential functions which are basic to the human auditory system: (1) a set of bandpass filters which provide the filtering performed by the basilar membrane, $h_{b,k}(t)$ (see figure 2); (2) an envelope detector which estimates the intensity of the signal from each of the bandpass filters as detected by the hair cells; (3) a non–linear intensity gain which logarithmically compresses the signal and is of the form $\frac{\beta_k \log \alpha_k I_k}{I_k}$, where I_k is the intensity of the sound signal from the k^{th} envelope detector, and β_k and α_k are parameters related to the gain of the outer hair cells (OHCs) and response of the inner hair cells for each respective channel of bandpass filters; (4) a high pass filter, $h_H(t)$, which provides loudness adaptation; (5) a hyperbolic tangent which functionally mimics the firing of the inner hair cells (IHCs) and the associated neural network; (6) a multiplicative intrinsic noise source which is additive in the logarithmic domain; and (7) an exponentiator and detector that in a simple way mimic the function loudness detection in the brain.

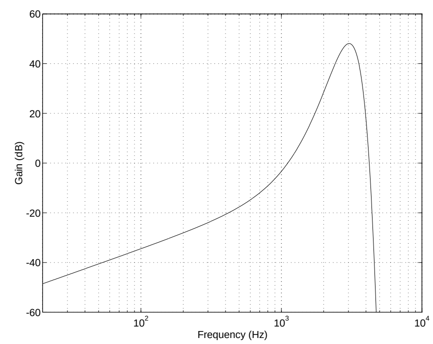

Figure 2: Simulated basilar membrane filter response following the model of [6] for the 1600^{th} stage of a 3500 stage cascaded filter bank with $Q = 0.77$ followed by a differentiator.

Creation of the model which provides these characteristics of the auditory tract enables the modeling of hearing for acoustically impaired subjects. It will be shown that the model (which transforms an input sound signal into a meta space related to normal hearing perception) can be combined with an equivalent inverse model of a damaged auditory system (which transforms from the normal-hearing meta space back into a pre-warped sound pressure space) to enable near normal hearing perception for a hearing impaired ear. Surprisingly, the combination of a forward model of a normal auditory system and an inverse model of a hearing impaired auditory system (in other words, a hearing aid) reduces to a set of bandpass filters, each of which is followed by a homomorphic automatic gain control (AGC) of the type proposed by Stockham [7].

1.1. Band Pass Filtering and Log Sensitivity

The input filter $h_{b,k}(t)$ of Figure 1 is derived from the cascaded second order transmission line model proposed by Lyon and Mead [6] to represent the propagation of acoustic energy along the basilar membrane.[1] The sharp tuning

[1] Use of a fixed filter $h_{b,k}(t)$ followed by a nonlinear operation (the log operation) is an approximation to the auditory filter-

of the filter is accomplished via active mechanical feedback into the basilar membrane by the OHCs. IHCs sense the output from each of the second order stages and the OHCs provide a gain which is dependent upon the intensity of the stimulus signal. An envelope detector mimics part of the processing performed by the hair cells and the olivocochlear complex to determine that intensity. A half wave rectified version of the acoustic stimulus is low pass filtered to provide an intensity estimate.[2] The output of the low pass filter is subsequently scaled by a factor α_k which is the inverse of the acoustic intensity required to raise the neural firing above the spontaneous rate. In order to accommodate individual hearing characteristics (especially for those with auditory losses), it is necessary to measure these threshold values as a function of frequency.

In the cochlea, the basilar membrane displacement is sensed and amplified by positive feedback from the OHCs. This compresses the dynamic range of the input signal approximately logarithmically. The fact that IHCs fire at a rate proportional to the logarithmic intensity combined with a non-linear saturation has been measured for animal cochlea (see [8], [9] for example) and several observations of a logarithmically compressive gain have been reported, and have been variously described [8]. Thus the gain of the OHCs is directly proportional to the logarithm of the intensity detected by the OHCs divided by that same intensity. Since the model output from the envelope detector is a scaled intensity estimate, a logarithm of the output [10], [8] is used to provide the intensity of the basilar membrane displacement.

1.2. Inner Hair Cell Transduction

The inner hair cell transduction of the acoustic stimulus is accomplished in two parts. First, the basilar membrane displacement results in a fluid velocity which acts on the IHCs. Mathematically, the expression for the basilar membrane displacement must first be spatially differentiated to obtain an expression for acceleration and then integrated temporally to obtain a function which represents the fluid velocity [11]. The final part of the IHC transduction occurs when the fluid which surrounds the IHCs moves relative to the IHC sterocilia thereby stimulating primary auditory neuron firings.

The model must reflect both the spatial differentiation and temporal integration inherent in the transformation from a membrane displacement to a fluid velocity as well as the neural firing rate associated with fluid velocity. The output of the logarithmic operation in the model of Figure 1 is the intensity of the envelope of the acoustic pressure field rather than the instantaneous fluid velocity which drives the IHCs; hence, a high pass filter is required.[3] The relationship between the velocity and neural firing rate is a hyperbolic tangent (see [12]) given by

$$V = \frac{V_m}{2}\left[\tanh\left(E \cdot X - F\right) + 1\right] \tag{1}$$

where V_m is the maximum output firing rate (see [13]), X is the input variable, and E and F are constants to be determined.

The block diagram of the proposed auditory model also includes three other subsystems: a noise source which adds noise to the nerve in the logarithmic domain and is therefore multiplicative, an exponentiator and then a detector that provides a simple model of the decision processes carried out later in the human auditory system. Depiction of the corresponding model of the human visual system may be found in Xie and Stockham [1].

2. Application to Auditory Processing

An acoustic signal $s(t)$ may be represented as

$$s(t) = e(t)v(t) \tag{2}$$

where $e(t)$ is a slowly-varying positive-valued envelope and $v(t)$ is a rapidly varying vibration. In the human ear the neural firing rate is determined by the intensity of the acoustic stimulus envelope, $e(t)$, as sensed at differing spatial locations along the basilar membrane. Encoding of the frequency information, derived from $v(t)$, is done both by the tonotopic organization of the auditory nerve and the IHCs as well as by the synchronization of the firings with the vibration of the stimulus for lower frequencies. The philosophy underlying the human auditory model is that an improved metric for processing is found in the output of the model which relates to sensation. The proposed process model of the auditory system maps the envelope, $e(t)$, and preserves information about $v(t)$. These are believed to be the features used by the auditory cortex for sound perception.

The hypothesis which will guide the analysis of hearing compensation is the assumption that the VIIIth nerve is normal, therefore the neural firing patterns and information provided to the VIIIth nerve ought to be the same from the compensated- degraded ear mechanism as from the normal ear mechanism. Hence, an inverse model is developed which will be used to transform the normal hearing meta space signal at the outputs of the hyperbolic tangent subsystem of the model, referred to hereafter as a perceptual space, to the input acoustic stimulus required to produce a similar meta space response for the damaged ear. Mathematically, the human auditory system model, which operates on an acoustic input stimulus and transforms the signal into the perceptual space, is characterized by the operator $H_n()$ for normal hearing and $H_d()$ for damaged hearing. For simplicity in the figures, only a single channel of the multiple

ing. The auditory system, with mechanical feedback from the OHCs, provides varying levels of "tuning" and gain along the basilar membrane. A fixed bandpass filter—fixed both in gain and bandwidth—is used to simplify the formulation,.

[2] The implementation of the half wave rectifier for the processing may alternatively utilize a full wave rectifier or the RMS value of the input with an appropriate scaling adjustment. Values for the cutoff frequency may be selected for each band to be equal to $\frac{1}{8}^{th}$ of the corresponding "critical" bandwidth.

[3] The transformation suggested by Lyon and Mead to convert from displacement to velocity, must be modified to reflect an equivalent result when operating on the envelope. In the passband of the filter, the OHC reponse to an input provides the logarithmic gain previously discussed. This gain has a scaling β_k. The differentiation which occurs is perhaps not an essential part of the signal analysis for most processing applications, but it provides the equivalent of visual Mach bands for hearing. For most applications, a gain of β_k may be used in place of the filter indicated.

parallel channels outlined earlier will be shown and channel variables will be designated by the subscript k.

The normal hearing perceptual space signal is given as

$$x(t) = H_n(s(t)) \tag{3}$$

where $s(t)$ is the acoustic stimulus and $x(t)$ is the perceptual space model subsystem output. Representing the damaged ear equivalent perceptual space quantities with accented variables, $x(t)$ is replaced with $\hat{x}(t)$ given by

$$\hat{x}(t) = H_d(s(t)). \tag{4}$$

One may express the inverse transformation required to transform the normal perceptual hearing space intensity into an input stimulus pre-warped by the inverse operations for the damaged ear as[4]

$$\hat{s}(t) = H_d^{-1}(x(t), v(t)) \tag{5}$$

The damaged ear operating on this pre-warped acoustic stimulus provides the perceptual space output

$$H_d(\hat{s}(t)) = H_d(H_d^{-1}(x(t), v(t))) = x(t) \tag{6}$$

Under the hypothesis that $H_d()$ sufficiently mimics the operation of the human auditory system, the hearing impaired individual listening to the pre-warped signal, $\hat{s}(t)$, provides the final auditory system operation $H_d()$ in the foregoing equation with their own ear. Therefore, the task required here is to provide an $\hat{s}(t)$ such that the damaged ear will perceive $x(t)$. The transformation required is

$$\hat{s}(t) = H_d^{-1}(x(t), v(t)) = H_d^{-1}(H_n(s(t)), v(t)) \tag{7}$$

2.1. Simplification of $H_d^{-1}(H_n(s(t)), v(t))$

The set of transformations suggested by $H_d^{-1}(H_n(s(t)), v(t))$ will be specifically defined as a series of subsystems and operations to allow a reduction in the complexity of the required processing (see also [5]). In order to simplify the discussion to follow,[5] a single orthogonal set of ideal bandpass filters which completely span the audio band of interest will be chosen as the bandpass filters in the initial portion of the human auditory model. Each of the bandpass filters, $h_k(t)$, is selected to be equivalent in bandwidth to the corresponding bandpass filter, $h_{b,k}(t)$, earlier described. The output of each of the bandpass filters is

$$s_k(t) = h_k(t) * s(t) \tag{8}$$

where the symbol $*$ indicates convolution. The sum of the individual component outputs forms the original signal

$$s(t) = \sum_{k=1}^{N} s_k(t). \tag{9}$$

[4]It should be noted that the model output is based primarily upon the slowly varying characteristics of the signal intensity at the output of each of the auditory filter channels. To reconstruct a signal which contains both the mapped slowly varying intensity, $x(t)$, and the vibratory information, the inverse model operation requires a knowledge of the vibration $v(t)$.

[5]This simplification will not limit the power of the result we are seeking, but it significantly simplifies the analysis.

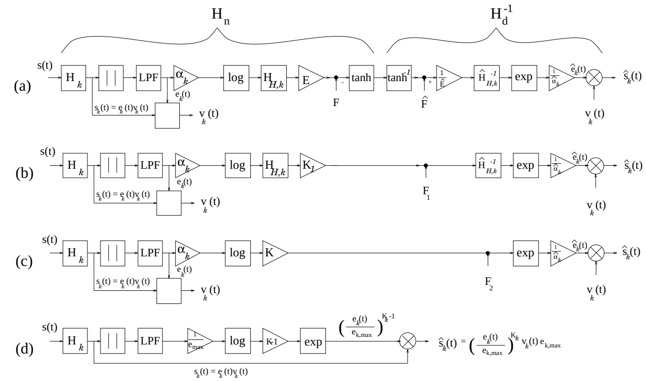

Figure 3: Forward and inverse model simplification.

Figure 3a contains a block diagram of the individual component operations in a single band represented by the notation $H_d^{-1}(H_n(s(t)), v(t))$. As the last stage of the reconstruction process, the mapped intensity function, $\hat{e}_k(t)$, is combined with $v_k(t)$ to provide $\hat{s}_k(t) = \hat{e}_k(t)v_k(t)$. The detail in figure 3a may be reduced to the diagram of figure 3b by noting that the sequence of a hyperbolic tangent and inverse hyperbolic tangent exactly cancel each other and combining the multiplication by E followed by a subtraction of F, a sum with $\hat{F}$ and a multiplication by $\frac{1}{\hat{E}}$. K_1 and F_1 represent the new constants required to maintain equivalency. Since the high-pass filters $H_{H,k}$ and $\hat{H}_{H,k}$ in figure 3b differ only by a gain constant, the high pass filter may be combined with the inverse high pass filter and results in the simplification shown in figure 3c.

At this point, one may recognize that a single channel of the human auditory model for compensation is a homomorphic automatic gain control of the type proposed by Stockham [7]. Motivated by this observation, the system gains can be reallocated and the result is shown in figure 3d. The parameter values in figure 3d are defined for each band as

$$\begin{aligned} e_{k,max} &= \left(\textit{ the UCL acoustic intensity } \right) \\ K_k &= \left[1 - \left(\frac{\textit{hearing loss } (dB)}{UCL\ (dB)\ -\ NHT\ (dB)} \right) \right] \end{aligned}$$

where UCL is the upper comfort level, the hearing loss is measured at threshold, and NHT is the normal hearing threshold. This multiplicative AGC provides no gain for signal intensities above the UCL and a gain equivalent to the hearing loss for signal intensities associated with the normal hearing threshold. The simplicity of the result is somewhat surprising, since the hearing models which lead to this result include recruitment effects and the nonlinearities of the inner ear, processing in the olivocochlear complex and synapse to the VIIIth nerve. The model adaptively and correctly (without explicit feedback or feedforward and with a minimum of delay) compresses the full range of audio inputs to the more limited range of the hearing impaired. The gain for each channel of the compensation system is

$$Gain_k(t) = \left(\frac{e_k(t)}{e_{k,max}} \right)^{K_k - 1}. \tag{10}$$

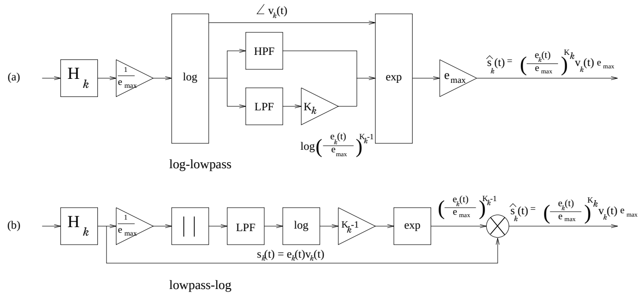

Figure 4: Comparison of the Stockham multiplicative AGC with a log–lowpass filter operation and (b) the multiplicative AGC resulting from the human auditory model with a low–pass log.

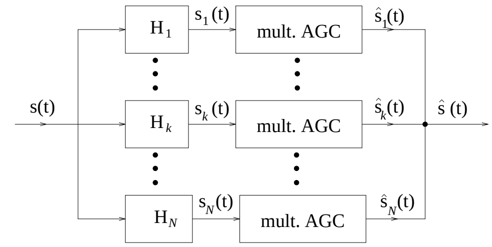

Figure 5: An N channel hearing compensation system. Each channel possessing a "critical bandwidth" bandpass filter, $H_k(f)$; a multiplicative AGC with an exponentiation factor K_k determined for that band; and a sum of the multiplicative AGC outputs.

2.2. Alternative Implementations

Figure 4a depicts the multiplicative automatic gain control proposed by Stockham [7]. In this configuration, a lowpass filter function is not present prior to the logarithmic operation as is the case in the human auditory model and the output of the logarithm is complex. Because the envelope of the signal $e(t)$ is slowly varying and positive, the $\log(e_k(t))$ is also slowly varying. Likewise $v_k(t)$ is a rapidly varying and so $\log|v_k(t)|$ has more high frequency components. If the linear filter proposed by Stockham is implemented with a low-frequency response K_k, and a unit passband response at higher frequencies, the $\log(e(t))$ will be multiplied by K_k while the $\log|v_k(t)|$ will be largely unaffected. The output is then

$$\hat{s}_k(t) = e_k^{K_k}(t) v_k(t). \tag{11}$$

If the input signal is normalized such that $e_k(t) < 1$, then for K_k less than unity, the variations in the envelope away from unity are reduced and a gain results.

The configuration proposed earlier by Stockham (see figure 4a) differs from the human auditory model in that the filtering occurs after the logarithmic operation. Figure 4b shows an equivalent implementation of the multiplicative AGC with low and high pass filters replacing $H(f)$. The $\log(e_k(t))$ is filtered through a 16 Hz lowpass filter with gain K_k in the passband, while a high-pass filter is used to separate the rapidly varying $\log|v_k(t)|$.

The Stockham multiplicative AGC and the multiplicative AGC (compared in figure 4) resulting from application of the human auditory model differ in the fact that the low pass filtering of $\log(e_k(t))$ in figure 4a occurs subsequent to the logarithmic operation while in figure 4b the lowpass filtering occurs prior to the logarithmic operation. Treatment of the lowpass–log and log–lowpass are considered by Stockham [1] in the context of a human visual model. A complete diagram which includes all channels of the hearing compensation system is shown in figure 5.

2.3. Hearing Aid Performance

Preliminary tests of the hearing aid were done in comparison with the Resound® and 3M Multimate® digitally programmable analog hearing aids. The hearing aid proposed herein yielded higher intelligibility scores and was generally prefered in subjective listening tests [14].

3. References

[1] Z. Xie and T. G. Stockham, Jr. Toward the unification of three visual laws and two visual models in brightness perception. *IEEE Trans. on Systems, Man and Cybernetics*, 19(2):379–387, March/April 1989.

[2] W. A. Yost and D. W. Nielsen. *Fundamentals of Hearing*. Holt, Rhinehart, and Winston, 1985.

[3] D. O. Kim. Functional roles of the inner– and outer–hair–cell sub-systems in the cochlea and brainstem. In Charles I. Berlin, editor, *Hearing Science*, pages 241–262. College–Hill Press, San Diego, 1984.

[4] J. Syka and R. B. Masterson (Eds.). *Auditory Pathway; Structure and Function.* Plenum Press, New York, 1982.

[5] D. V. Anderson. Model based development of a hearing aid. Master's thesis, Brigham Young University, Provo, Utah, April 1994.

[6] R. R. Lyon and C. Mead. An analog electronic cochlea. *IEEE ASSP Trans.*, 36(7):1119–1134, July 1988.

[7] T. G. Stockham, Jr. The application of generalized linearity to automatic gain control. *IEEE Trans. on Audio and Electroacoustics*, AU-16(2):267–270, June 1968.

[8] C. D. Geisler. Effects of a compressive nonlinearity in a cochlear model. *JASA*, 78(1):257–260, July 1985.

[9] N. Y. S. Kiange, et al. *Discharge Patterns of Single Fibers in Cat's Auditory Nerve.* MIT, Cambridge, Mass., 1965.

[10] W. S. Hellman and R. P. Hellman. Intensity discrimination as the driving force for louness. application to pure tones in quiet. *JASA*, 87(3):1255–1265, Mar. 1990.

[11] S. T. Neely and D. O. Kim. A model for active elements in cochlear biomechanics. *JASA*, 79(5):1472–1480, May 1986.

[12] T. F. Weiss and R. Leong. A model for signal transmission in an ear having hair cells with free–standing stereocilia. iv. mechanoelectric transduction stage. *Hearing Research*, 20:175–195, 1985.

[13] M. C. Liberman and N. Y. S. Kiang. Single-neuron labeling and chronic cochlear pathology. vi. stereocilia damage and alterations in rate– and phase–level function. *Hearing Research*, 16:75–90, 1984.

[14] D. V. Anderson and R. W. Harris. Evaluation of a hearing compensation algorithm. To be published, 1994.