

Experiments with a Computational Model of the Cochlea

Richard F. Lyon
Schlumberger Palo Alto Research
3340 Hillview Ave.
Palo Alto, CA 94304

Lounette Dyer
Computer Science Department
California Institute of Technology
Pasadena, CA 91125

Abstract

A computational model of the cochlea has been subjected to a number of experiments to clarify its relation to psychophysical and physiological reality. The experiments are useful in calibrating the model, but more importantly, they are useful in explaining how seemingly contradictory data can be accounted for by simple mechanisms.

For example, measurements of threshold tuning curves and masking functions show clearly that threshold tuning curves are sharper than the cochlear mechanical filter transfer functions, without requiring a second tuned mechanism. Similarly, critical bandwidths or spread of masking are seen to be wider than the filter transfer functions. These effects are shown to be consequences of the coupled automatic gain control model which was originally introduced as an aid to handling a large dynamic range; the AGC's effect on tuning is a side effect.

1 Introduction

For several years we have been experimenting with computational models of hearing, including both peripheral (cochlear) processing and more central sound interpretation processing [1, 2, 3, 4]. The models studied include adaptations of previously proposed models of cochlear filtering by the basilar membrane [5], transduction by the hair cells [6], and central processing via correlations [7, 8]. In addition, several stages of coupled automatic gain control are included, and dominate the computational cost of the cochlear model.

The coupled or multichannel automatic gain control (MAGC) stages have been developed from an engineering perspective to allow the model to cope with a very wide input dynamic range (greater than 100 dB), while maintaining a robust representation in a very limited output dynamic range. While the technique is motivated by the physiological notion of lateral inhibition [9], cochlear mechanisms for such a function have not been demonstrated. It seems a reasonable speculation, however, that the efferent nerves going to the outer hair cells somehow implement an adjustment of gain or sensitivity, and that other more local, perhaps mechanical, mechanisms in the cochlea may also implement very fast adaptation.

This paper addresses an experimental approach to assessing the plausibility of the model as presently implemented, concentrating on the effects of the nonlinear/time-varying MAGC stages. The relationships between phenomena measured on live mammalian cochleae and similar phenomena measured experimentally on the cochlear model are discussed.

2 Model Overview

The computational model of the cochlea is implemented as a cascade of multi-channel signal processing operators, programmed as signal types in the signal representation language SRL [10]. The overall structure of the model is shown in figure 1.

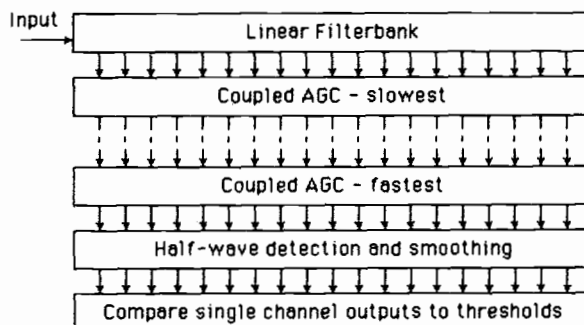


Figure 1: Block diagram of the Cochlear Model.

The first stage is a time-invariant linear filterbank implemented as a cascade of biquadratic digital filters with output taps between each pair of filters. This filterbank is a simplification of the more general cascade/parallel structure for the case of identical poles in the resonators and the following notch filters, and without micro-mechanical anti-resonances (see [1]). The filter structure and transfer functions are motivated by the simplified linear long-wave approximation of the one-dimensional hydrodynamic model of the cochlea with a forward-only traveling wave [5]. The transfer functions of a pair of stages and the overall transfer functions to a pair of adjacent output taps is shown in figure 2.

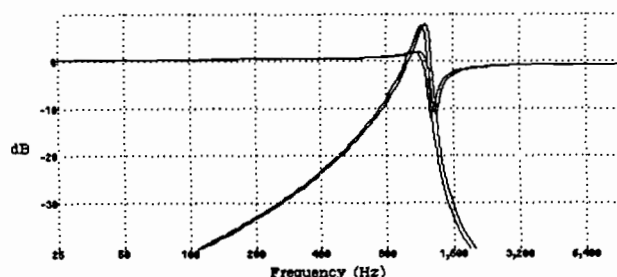


Figure 2: Cochlear filter transfer functions.

Each filter stage contributes a resonance and an antiresonance. The cascade of these stages plus an initial stage with pre-emphasis and a zero at DC yields the overall bandpass functions shown, with very sharp high-side rolloff. The zero at DC approximately models the hair cells' sensitivity to membrane velocity that is caused by viscous drag bending the cilia of the hair cells; for implementation efficiency the zero is put in the first stage of the cascade, rather than after each tap.

The following four stages of the model are MAGC stages of identical structure but with differing parameters. The first has the highest target level, the longest time constant, and the widest coupling across channels; the fourth stage has the lowest target and most local adaptation. Each stage uses feedback of half-wave detected outputs through a time-space lowpass filter to reduce forward gains; gains are reduced enough to keep the average of the output across time and space below the target level of the stage. Thus, the gain of a channel will be reduced by detected activity on other channels near it.

The time constants used are 640, 160, 40 and 10 milliseconds, and the target levels are 0.0032, 0.0016, 0.0008, and 0.0004 (on the same scale, a unit-amplitude sine wave is arbitrarily termed +125 dB, so there is not much compression for signals significantly below +57 dB). The spatial spread functions are exponentially decreasing away from the affected channel with spatial decay constants of 64, 32, 16, and 8 channels respectively, which occur naturally with 20 kHz sample rate using the above time constants and equal weights of 1/3 in the previously published algorithm [11].

Other model stages depend on the experiment. In some cases, a realistic hair cell model might be used, with perhaps a hyperbolic tangent half-wave detection nonlinearity. For the experiments reported here, an ideal half-wave rectifier was used, and only averaged detected output levels were observed. This is motivated by analogy with physiological experiments on neural thresholds, in which the detection criterion is an observed increase in neural firing rate by a specified amount, and detailed patterns of activity are ignored. According to Jeffress [12], this ignorance of phase information is consistent with masking data when the masker is a noise.

3 Experimental Technique

The basic tuning curve experimental paradigm is to pick a single output channel (or neural unit) and to measure the level of sinusoidal input required to cause the output of the channel to reach a criterion level, as a function of stimulus frequency. This is typically done by slowly ramping up the amplitude of the sine wave, so that the level can be measured as the time of threshold crossing. In the computational model, unlike in physiological experiments, we are able to monitor all channels simultaneously, which speeds up the experimentation.

Since we are interested in the relationship of the MAGC targets to detection criteria, we set a variety of detection thresholds and make separate tuning curves for each. These separate detection thresholds can also be monitored concurrently, so little extra time is required. We have used detection thresholds of 1/32, 1/16, 1/8, 1/4, 1/2, and 1/1 of the final (lowest) MAGC target level.

For signals below or near the lowest threshold, the MAGC stages have little effect and the tuning curve shapes are nearly equal to the inverses of the filter transfer functions. At higher thresholds, nonlinear effects come into play. The threshold that

gives the most realistic tuning curve shapes will imply the appropriate scaling of the hair cell model detection nonlinearity and the neuron model parameters.

For masked tuning curves, a masker of fixed frequency and amplitude is added to the increasing amplitude test tone, and the thresholds for each channel are taken as the response to masker alone plus the above increments. This is analogous, but not identical, to a fixed increase in neural firing rate in physiological experiments.

Since threshold experiments are essentially steady-state measurements, the dynamics of the MAGC loops are not important. For most of the experiments we use an alternative implementation in which time constants and space constants are separately controlled, so that we can speed up the response and use much faster amplitude sweeps. Results are consistent across the two implementations except for the case of test tones close in frequency to a masker, because in that case the beating interferes with the fast sweep, leading to increased error.

In all cases the threshold crossing time quantization has been adjusted to correspond to 1 dB amplitude increments, which leads to some visible noise in the resulting curves. For low frequency test tones, a beating between the sine frequency and the time sample rate leads to some aliasing noise, since only a crude anti-aliasing filter is used to smooth the response.

4 Experimental Results

Tuning curves for every fourth channel are shown in figure 3. Since the model was run with nominally four channels per Bark, the curves are spaced about one Bark apart. The overall threshold, or the trace of the minima of the tuning curves (not shown, but apparent from figure 3), can be viewed as an equal-loudness curve. Apparently, the equal-loudness curve is tilted about twice as much as would be expected between 200 Hz and 4 kHz, due to excessive pre-emphasis or inappropriate filter gains. This effect can be easily adjusted.

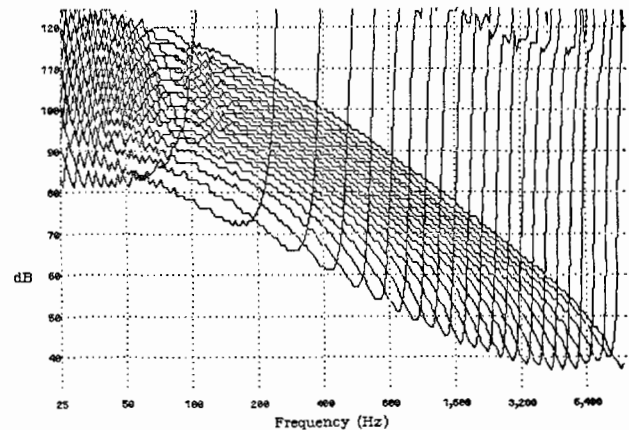


Figure 3: Tuning curves for threshold 1/8.

Figure 4 shows tuning curves for a single channel (channel #44, CF=2340 Hz) with the detection threshold as a parameter. While the tuning curve shapes for the lowest thresholds resemble the inverse of the filter transfer function, the higher thresholds lead to increased amounts of sharpening of the tuning curves, especially in the tails. The tuning curve tip sensitivities are about 6 dB apart, due to the factor of two spacing of detection thresholds, except for the case of the highest thresholds.

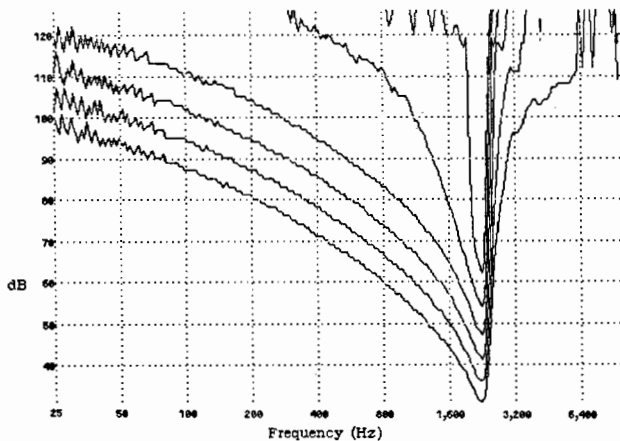


Figure 4: Tuning curves for CF = 2340 Hz, all thresholds.

5 Understanding the Coupled AGC

At low signal levels, the MAGC stages operate with unity gain; as the signal level increases, the gains reduce so that a higher input level is required to reach a fixed detection threshold. Hence, tuning curves are elevated at high input levels, relative to the shape of the filter transfer functions.

For detection thresholds near the MAGC targets, the MAGC gains are already significantly reduced when a signal is first detected at the unit with corresponding CF. For lower detection thresholds, gains are nearly constant until the signal is significantly above the minimum detection level, so that only the tails of the curve are elevated.

Even between the two lowest curves in figure 4, some elevation of the high-side tail is evident, since the high input level causes gain reduction in channels near the channels that are responding strongly. Thus, even though the output is well below the MAGC target on the channel under consideration, a spread of inhibition from other strongly responding channels causes a threshold elevation.

The shapes of the low-frequency tails of the tuning curves, which resemble measured neural data for thresholds around 1/2 to 1/4 of the MAGC target, are thus evidently dependent on the characteristics of the adaptation mechanism, as well as on the detection criterion used in measuring the curve. An interesting physiological experiment, therefore, would be to see if corresponding shape changes can be seen by using criteria more and less sensitive than "perceptible change in neuron firing rate". A synchrony measure would have to be used to make a more sensitive criterion.

If this interpretation is correct, the data seem to imply that a significant gain reduction occurs in the cochlea (in the mechanics, the hair cells, or the primary auditory neurons) at stimulus levels below that needed to cause a significant increase in neural firing rate. That is, the adaptation might be based on feedback from a synchrony detection process.

6 Masking Results

The above tuning curve measurements were repeated with a 1 kHz masker at 25, 45, 65, 85, and 105 dB. As a sample of the data, figure 5 shows curves for every fourth channel with detection threshold 1/8, for a masker at 85 dB. Figure 6 shows tuning curves as a function of detection threshold for a channel with CF = 2340 Hz.

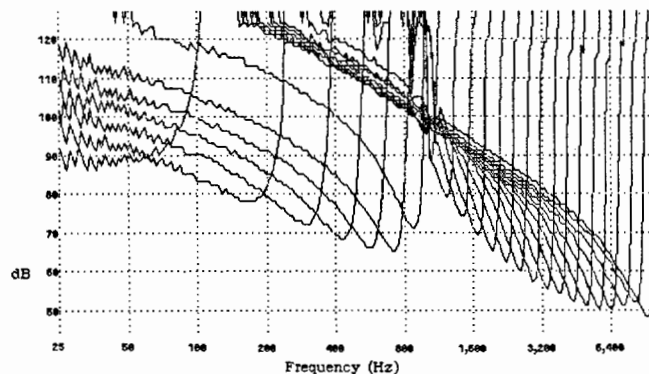


Figure 5: Masked tuning curves for threshold 1/8, 1 kHz masker at 85 dB.

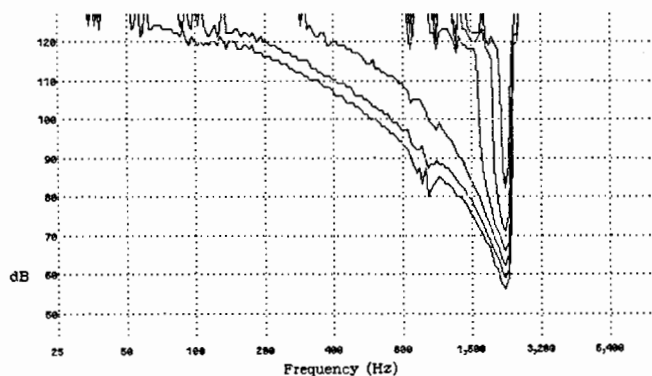


Figure 6: Masked tuning curves for CF = 2340 Hz, all thresholds, 1 kHz masker at 85 dB.

An overall detection curve is again taken as the trace of the minima of the individual channel tuning curves, and is shown in figure 7. The difference between these curves and the corresponding unmasked detection curves is a measure of the effect of masking, shown in figure 8.

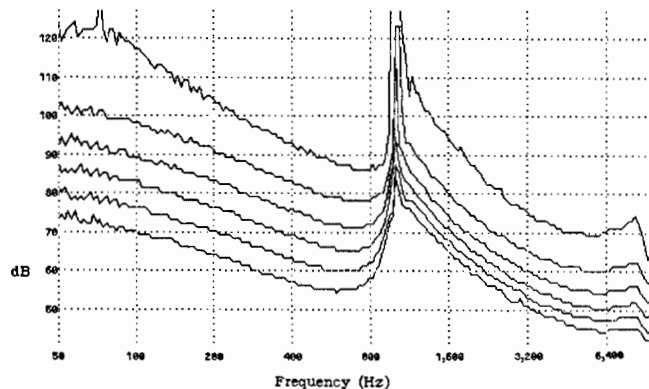


Figure 7: Overall threshold with 1 kHz masker at 85 dB, for 6 detection threshold levels.

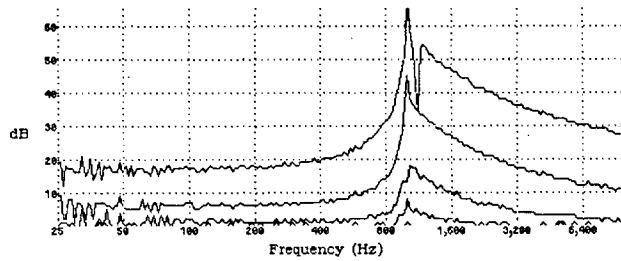


Figure 8: Overall threshold elevation due to masking, for five masker levels separated by 20 dB, 1 kHz masker.

The spread of masking indicated by these curves is more than would be due to the filter transfer functions alone; some of the masking is due to a spread of gain reduction via the MAGC mechanism. The masking due to gain reduction can be viewed as "two-tone suppression", as it will suppress the absolute (rate) response of a single unit to a tone when the suppressor is sub-threshold for the unit under consideration; this effect will occur for maskers both above and below CF with the present model.

Figures 7 and 8 should be compared with figures 1 and 2 in [12], where it is explained that some of the complex shape of Wegel and Lane's original masking curves [13] is due to beating. The fast amplitude sweep and the effect of the MAGC dynamics on beats in the present experiments has resulted in rather chaotic behavior at high signal levels. The general form and level dependence of the curves is encouraging, and more study should reveal model changes that would improve the fit. Perhaps measurement of two-tone suppression regions would provide a more sensitive measure of the model fit.

7 Cochleagrams for Speech Recognition

For use in speech recognition, a model stage that uses phase to emphasize resonances and other spectral peaks such as resolved harmonics is included [11]. Experiments in processing speech have been begun with a corpus of single pronunciations of "one" and "nine" by 112 adult talkers from the TI connected digit database. Experiments with spectral tilt and loudness variations show that the MAGC effectively de-emphasizes these effects, favoring more local contrasts; similarly, formant amplitude and bandwidth are de-emphasized, while formant locations are clearly represented. Current results will be presented at the conference. A typical vector quantized digit cochleagram is shown in figure 9.

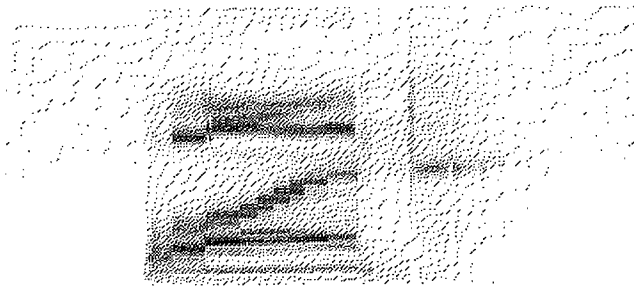


Figure 9: Vector quantized cochleagram of the digit "one".

8 Conclusions

While it is difficult to draw firm conclusions from simple experiments on models, it seems that the general idea of adaptation to a wide dynamic range of auditory stimuli via some form of coupled automatic gain control is reasonably consistent with many measured effects. In particular, the fact that neural tuning curves are sharper than the mechanical transfer function of the basilar membrane can be easily explained as a side effect of a coupled AGC, requiring no tuned mechanism after the cochlear mechanics (i.e., sharpening is realized with no "second filter").

The same kind of model that mimics subtle physiological phenomena appears to be well suited to extracting the information needed for speech recognition, which was the original motivation for the model development.

9 References

- [1] Richard F. Lyon, "A Computational Model of Filtering, Detection, and Compression in the Cochlea," *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Paris, May 1982.
- [2] Richard F. Lyon, "A Computation Model of Binaural Localization and Separation," *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Boston, Apr. 1983.
- [3] Richard F. Lyon, "Computational Models of Neural Auditory Processing," *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, San Diego, Mar. 1984.
- [4] Mitchel Weintraub, "The GRASP Sound Separation System," *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, San Diego, Mar. 1984.
- [5] G. Zweig, R. Lipes, and J. R. Pierce, "The Cochlear Compromise," *JASA* 59, pp. 975-982, 1976.
- [6] Jont B. Allen, "A Hair-Cell Model of Neural Response," in *Mechanics of Hearing*, ed. E. deBoer and M. A. Viergever, Delft University Press, 1983.
- [7] L. A. Jeffress, "A Place Theory of Sound Localization," *J. Comp. Physiol. Psychol.* 41,35-39, 1948.
- [8] J. C. R. Licklider, "A Duplex Theory of Pitch Perception," *Experientia* 7,128-133 (1951), reprinted in *Psychological Acoustics*, E. Schubert ed.; Dowden, Hutchinson, and Ross, Inc., Stroudsburg, PA, 1979.
- [9] G. von Békésy, *Sensory Inhibition*, Princeton University Press, 1967.
- [10] Gary Kopec, "The Signal Representation Language SRL," *IEEE Trans. ASSP* 33:4, Aug. 1985.
- [11] Richard F. Lyon and Niels Lauritzen, "Processing Speech with the Multi-Serial Signal Processor," *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Tampa, Mar. 1985.
- [12] Lloyd A. Jeffress, "Masking," in *Foundations of Modern Auditory Theory 1*, Jerry V. Tobias ed., Academic Press, 1970.
- [13] R. L. Wegel and C. E. Lane, "The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear," *Phys. Rev.* 23,266-285 (1924), reprinted in *Psychological Acoustics*, E. Schubert ed.; Dowden, Hutchinson, and Ross, Inc., Stroudsburg, PA, 1979.