A COMPARISON USING SIGNAL DETECTION THEORY OF THE ABILITY OF TWO COMPUTATIONAL AUDITORY MODELS TO PREDICT EXPERIMENTAL DATA

Lisa C. Gresham and Leslie M. Collins

Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708-0291
lcg@ee.duke.edu, lcollins@ee.duke.edu

ABSTRACT

In order to develop improved remediation techniques for hearing impairment, auditory researchers must gain a greater understanding of the relation between the psychophysics of hearing and the underlying physiology. One approach to studying the auditory system has been to design computational auditory models that predict neurophysiological data such as neural firing rates (Patterson *et al.*, 1995; Carney, 1993). To link these physiologically-based models to psychophysics, theoretical bounds on detection performance have been derived using signal detection theory to analyze the simulated data for various psychophysical tasks (Siebert, 1968).

Previous efforts, including our own recent work using the Auditory Image Model, have demonstrated the validity of this type of analysis; however, theoretical predictions often exceed experimentally-measured performance (Gresham and Collins, 1998; Siebert, 1970). In this paper, we compare predictions of detection performance across several computational auditory models. We reconcile some of the previously observed discrepancies by incorporating phase uncertainty into the optimal detector.

1. INTRODUCTION

One goal of auditory research is to design improved remediation devices that are capable of restoring impaired hearing to "normal." By establishing a link between the physiology of normal and impaired auditory systems and perception, we can determine what information each system uses to detect and classify sounds. With this knowledge, new devices can be designed to enhance the signal prior to delivering it to the damaged system, thus compensating for impairments.

Signal detection theory has been used to study the auditory system by calculating the detection performance achieved by "ideal observers" on various psychophysical tasks [3]. Often, the theoretic principles were applied directly to the stimuli, ignoring any transformations that occur as the signal passes through the auditory system. Realizing that some processing does occur in the peripheral auditory system, researchers began applying signal detection theory to the output of simplified auditory models (e.g., an energy-detector) [7], [8]. However, these functional models mimicked little of the physiological detail of the auditory system. Subsequently, basic physiologically-based computational models were developed to simulate neurophysiological data such as neural firing rates. To link physiological data to psychophysical behavior, signal detection theory was applied to these physiologically-based models to obtain predictions of detection performance on several psychophysical tasks [9], [10].

Modern computational models, such as the Auditory Image Model [6], represent the physiology of the auditory system more accurately than their basic forerunners. Our recent work has demonstrated that analyzing such models using signal detection theory allows us not only to predict detection performance for any generic input signal, but also to study what information is lost and where this loss occurs as the signals propagate through the auditory system, an important factor when studying impairments [4]. However, like methods used in the past, theoretical predictions exceed experimental performance on the detection of a 500 Hz tone in broadband, Gaussian noise. Here, we first investigate this discrepancy by comparing predictions from both functional and physiological models under signal-known-exactly conditions. Secondly, the *a priori* assumptions are modified such that the input phase is uncertain and the results are again compared across models.

In the following sections, the models we used are described in detail. We then discuss the procedure used to derive the "optimal" detector and explain how the performance of each model was evaluated. Finally, the results are presented and some conclusions are drawn regarding where the discrepancy between theoretical predictions and experimental performance might originate.

2. COMPUTATIONAL AUDITORY MODELS

Typically, computational auditory models process auditory stimuli in multiple "stages" designed to simulate two components of peripheral auditory processing: spectral analysis and neural encoding. Both the Auditory Image Model (AIM) [6] and the Carney model [1] were designed in this way, however, they differ in their implementation. Broadly, the models can be classified as either functional, simulating hypothesized signal processing, or physiological, simulating actual structures such as inner hair cells. AIM contains both a functional and a physiological module whereas the Carney model only has the latter. AIM also provides an additional stage of processing that converts the neural firing rate into an "auditory image." The following sections describe the specific implementation each model uses to simulate auditory processing.

2.1. Auditory Image Model

2.1.1. Functional Module

The functional modules of AIM (AIM-F) are based solely on the hypothesized signal processing performed by the auditory system

Support for this work was provided by the National Institutes of Health (1R03-DC-03136-01), the Lord Foundation, the Whitaker Foundation, and the American Association of University Women.

rather than specific physiological mechanisms. The first stage of processing, spectral analysis, is performed by a linear gammatone filterbank. The filters are level-independent and have a time-invariant, frequency-dependent bandwidth. Each channel, corresponding to a filter with a particular center frequency, is processed independently. Therefore, non-linearities such as distortion products are not present in the output signal. Filtering also creates a general rightward skew in the output from high to low-frequency channels corresponding to propagation delay in the cochlea.

Neural encoding is simulated by rectifying, compressing, and applying two-dimensional adaptation to the outputs of the filterbank. Frequency adaptation, which is performed across channels, sharpens features such as formant frequencies, whereas temporal adaptation sharpens features within a single channel. The phase differences introduced in the previous stage are maintained.

Finally, "strobed" temporal integration is applied to the neural activity pattern. Once again, the output of each channel is processed independently with integration in a particular channel being triggered by a peak in the activity pattern of that channel. This results in phase alignment across each of the channels. The ultimate output of AIM is an "image" that, in the case of a periodic stimulus, is static.

2.1.2. Physiological Module

The physiological modules of AIM (AIM-P), as well as the Carney model (described below), are designed to simulate actual physiological structures and mechanisms of the auditory system, not simply the signal processing. As such, spectral processing is performed using transmission line filtering designed to approximate cochlear hydrodynamics. Level-dependence is incorporated using a feedback circuit that simulates outer haircell responses. In addition, channel interactions produce non-linearities such as distortion products and two-tone suppression. A phase lag, similar to that observed in the functional case, is also present.

The neural firing pattern is generated by simulating one inner haircell per channel using the Meddis haircell model [5]. All haircells were assumed to be synapsed with medium spontaneous rate fibers. In contrast to the functional version, there is no frequency sharpening due to cross-channel coupling. In addition, the compression that occurs at this stage of the functional model is instead performed as part of the transmission line filtering.

In the final stage of the physiological model, the neural activity pattern is processed by a bank of autocorrelators. The resulting "image" is known as a correlogram. Since each channel is processed independently, there are no cross-frequency effects. As in the functional model, the phase lag also disappears.

2.2. Carney Model

Like the physiological modules of AIM, the Carney model attempts to simulate physiological mechanisms in the auditory system to produce accurate representations of neural data. Parameters of the model were chosen so that accurate responses (PST histograms, rate-level functions, tuning curves) to simple and complex stimuli were obtained.

The Carney model, like the functional version of AIM, first performs spectral analysis using a linear gammatone filterbank. However, in contrast to AIM-F, Carney introduces a compressive nonlinearity through a feedback loop designed to simulate outer haircell response. The effect of the feedback is to create time-varying filters whose responses are level-dependent. However, unlike the physiological version of AIM, no channel interactions are modeled. Thus, nonlinearities such as distortion products and two-tone suppression cannot be represented. The latency of the response of auditory nerve fibers, which results in phase lags, is simulated by introducing a frequency-dependent time delay to the waveforms at the output of the filterbank.

The delayed output of the filterbank is then passed through an inner haircell simulator modeled as a saturating non-linear function followed by two low-pass filters. In contrast to AIM-P which simulates medium spontaneous rate fibers, this model simulates the responses of only high spontaneous rate fibers. In addition, nonlinear adaptation effects are incorporated into the model. At this point, the signal is comparable to the neural activity pattern produced by AIM. However, from this stage on, the processing performed by the two models differs substantially. Rather than perform temporal integration, as AIM does, the Carney model applies the neural activity pattern to a Poisson spike generator. In this way, some uncertainty or "internal noise" is added to the output data.

3. DERIVING THE OPTIMAL DETECTOR

One of the advantages of using computational auditory models is that optimal detectors can be derived from simulated physiological responses to various stimuli. The auditory stimuli used in our simultaneous masking task were a tone at 500 Hz and broadband, Gaussian noise. Under the null hypothesis (H_0), the noise alone was used as the input to the model; under H_1, the tone and noise were added together with $E/N_0 = 14$. Multiple independent realizations of the noise and tone plus noise were propagated through each model. The result was an ensemble of output waveforms representing the model's responses to the two different inputs. From these ensembles, histograms were generated from the output responses at the i^{th} point in time across all waveforms. This procedure yielded estimates of the true density functions, $p(r_i|H_0)$ and $p(r_i|H_1)$.

Despite some temporal correlation evident in the data, the density functions were assumed to be independent across time to simplify computations. Thus, the likelihood ratio, λ, which is defined as the ratio of the density functions under each hypothesis, can be expressed as:

$$\lambda = \frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)} = \prod_i \frac{p(r_i|H_1)}{p(r_i|H_0)}, \tag{1}$$

where i is a temporal index into the vector of received data, r. The optimal detector is simply a monotonic function of this likelihood ratio. Thus, by using the output data from the computational models to derive the optimal detector, it is possible to incorporate processing performed by the auditory system into this analysis, thereby providing an advantage over traditional methods which analyze only the input signals.

The previous equation was derived under the assumption that all parameters of the input signal are known exactly. However, this is not necessarily a reasonable assumption since the auditory system does not know *a priori* which signals will impinge upon it. It is more likely that some uncertainty regarding the signal exists in the auditory system. This uncertainty can be incorporated into the optimal processor by integrating over the uncertain parameters.

$$\lambda = \frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)} = \frac{\int_\psi p(\mathbf{r}|H_1,\psi)p(\psi|H_1)d\psi}{p(\mathbf{r}|H_0)}, \tag{2}$$

where ψ is the set of all uncertain parameters and $p(\psi|H_1)$ describes the *a priori* knowledge the system has about the distribution of the input parameters. For the case presented in this paper, ψ was the phase of the tone, θ, which was assumed to be uniformly distributed on the interval [0:2π]. This parameter was chosen because it is reasonable to assume that the auditory system has no *a priori* knowledge of the phase of the signal impinging on the ear. Assuming temporal independence, Equation 2 can be rewritten as,

$$\lambda = \frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)} = \frac{1}{2\pi}\int_\theta \prod_i \frac{p(r_i|H_1,\theta)}{p(r_i|H_0)} d\theta. \qquad (3)$$

The conditional density functions, $p(r_i|H_1,\theta)$, are estimated as before using input tones with all possible values of θ rather than a single, known value.

4. ANALYSIS PROCEDURE

In order to evaluate the detection performance of various auditory models, the Receiver Operating Characteristic (ROC) curves for each model are determined. Such curves are derived by comparing the likelihood ratio to a threshold, β. The probability of correctly detecting the tone when it is present, P_d, and the probability of giving a false alarm when the stimulus is noise alone, P_f, can be uniquely determined for each threshold value. Each pair of probabilities constitutes a single point on the ROC curve. The value of the threshold can be systematically varied so that a complete ROC curve is generated.

Many complete curves are generated for each model and averaged together resulting in a single performance curve. These curves, depicting the theoretical detection performance associated with each model, can then be compared visually, by plotting them on normal-normal plots, and statistically, by deriving the detectability index, d_s, from the raw data and determining the slope of the ROC curve, equal to the ratio of standard deviations, σ_n/σ_{sn} [2].

The differences between the models were investigated by studying the performance of each on a simple psychophysical task – the detection of a 500 Hz tone in broadband, white Gaussian noise. In order to compare AIM and the Carney model, the spontaneous rate of the fibers in the Carney model was set to zero since AIM does not include this parameter. In addition, the version of the Carney model used in our analyses did not include the Poisson discharge generator. Finally, the ROC curves presented in this paper only provide a comparison between the performance of the Carney model (which ends with the production of a neural firing rate) and the performance obtained at the comparable stage of AIM. A complete quantitative comparison is included in tabular form.

5. RESULTS

A prediction of detection performance was generated for each model under the assumption that all parameters of the input signal, including the phase, were known *a priori*. Figure 1 depicts the ROC curves generated by applying the optimal detector to the average neural firing data of each model as well as the ROC depicting the performance obtained when the "optimal" processor operates on the original input signal. Since, in the latter case, the signal has not been degraded by being processed by a model, it contains all of the original information. Therefore, the corresponding ROC curve

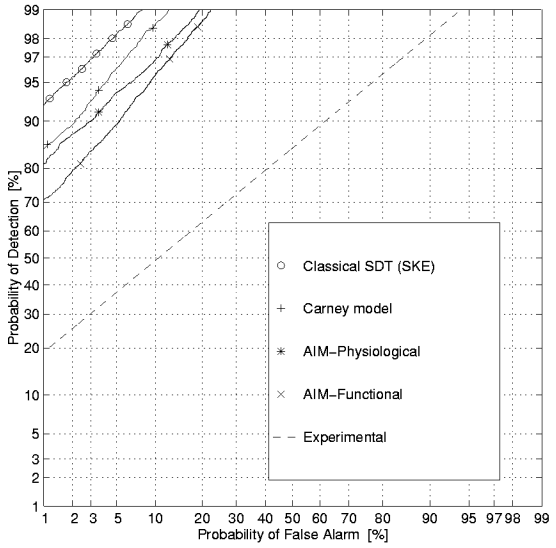

Figure 1: ROCs for the known phase case: (○), Classical SDT (SKE); (+), Carney model; (∗), AIM-Physiological; (×), AIM-Functional; (- -), Experimental (adapted from Watson *et al.* [11]).

represents the best performance achievable for the given task. Although this measure is not a realistic predictor of actual auditory performance, it is useful as a reference.

All of the models predict performance worse than the "optimal" case, with the functional model showing the greatest decrease. Table 1 quantifies the ROC curves in terms of a detectability index, d_s, and a slope, σ_n/σ_{sn}. The decrease in performance suggests that information is being lost as the signal propagates through each model. Discrepancies between the models can be attributed to the fact that different implementations of auditory processing result in different amounts of information being lost. Finally, all predictions can be compared to actual human performance. The dashed line represents the ROC derived from experimental data obtained by Watson and his colleagues [11] for the same detection task.

The performance curves in Figure 2 were obtained when the phase of the input signal was assumed *a priori* to be uniformly distributed. The optimal processor for each model was derived by integrating over the phase as shown previously in Equation 2. As in the first figure, the experimental and the classical signal-known-statistically (SKS) case ROCs are included for reference.

Comparing the classical SDT ROC curves for both the known and unknown phase cases reveals that phase uncertainty in itself causes a decrease in performance. It follows that, even if no information were lost as the signal propagated through the auditory system, one would expect to observe at least the same drop in detectability between the known and uncertain cases when processed by any auditory model. Examination of the model-predicted ROC curves reveals that the detection performance has indeed decreased for all models as a result of the added phase uncertainty.

According to Table 1, the expected change in detectability is $d_{s,SKE} - d_{s,SKS} = 0.6$. Performing similar calculations for the Carney, AIM-P, and AIM-F models, values of 0.9, 0.7, and 0.6 are obtained, respectively. We can conclude, therefore, that the change observed for the functional model (AIM-F) can be explained by

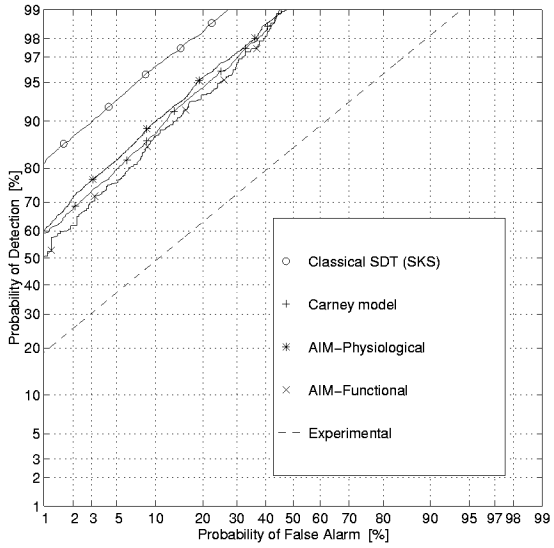

Figure 2: ROCs for the uncertain phase case: (∘), Classical SDT (SKS); (+), Carney model; (∗), AIM-Physiological; (×), AIM-Functional; (- -), Experimental (adapted from Watson *et al.* [11]).

the addition of phase uncertainty alone. The two physiological models, however, show a greater decrease in d_s suggesting that there has been additional information lost.

Table 1 also provides data quantifying the detection performance of the energy-detector model and the performance achieved by AIM at its final output, the auditory image. The results show that the detectability index of the energy-detector model is greater than indices obtained using either AIM or the Carney model under the unknown phase assumption. However, the slope is less than that of the models and the experimental value. Performance measurements for the auditory image stage of AIM indicate that there is a slight decrease in both d_s and the slope for both versions. However, for any single module, there is essentially no difference between the known and unknown phase cases.

6. SUMMARY

The data show that, for a simultaneous masking task, both the Carney model and AIM yield similar predictions of detectability at the neural firing stage when the same *a priori* assumptions are made. This suggests that both models are equally good at predicting detection performance for a simple task which requires only a single channel analysis. However, more complex tasks, such as frequency discrimination, create channel interactions, a feature not included in the Carney model. Differences in implementation, therefore, may determine which model should be used.

As the results show, phase information aids in the detection of the signal at the level of the neural firing. However, since this information is removed by temporal integration, uncertainty in this parameter does not affect detection performance at the level of the auditory image. In order to explain the remaining discrepancy between theoretical predictions and experimental data, future work will use the average firing rate as the input to a Poisson discharge generator, thereby adding physiologically-based "internal noise" to the system which will decrease detection performance.

Location	Model	Assumptions	d_s	σ_n/σ_{sn}
Outside the ear	SKE	Known θ	3.7	1.00
	SKS	Unknown θ	3.1	0.82
	ED	Unknown θ	2.8	0.76
Neural Firing	Carney	Known θ	3.4	1.13
	AIM-P	Known θ	3.2	0.98
	AIM-F	Known θ	3.0	1.17
	Carney	Unknown θ	2.5	0.95
	AIM-P	Unknown θ	2.5	0.90
	AIM-F	Unknown θ	2.4	1.00
Auditory Image	AIM-P	Known θ	2.4	0.75
	AIM-P	Unknown θ	2.4	0.76
	AIM-F	Known θ	2.2	0.92
	AIM-F	Unknown θ	2.2	1.04
Experimental			1.1	0.82

Table 1: Summary of ROC curves.

7. REFERENCES

[1] Carney, L.H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cats," J. Acoust. Soc. Am. **93**, 401-417.

[2] Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (**1961**). "Interval of time uncertainty in auditory detection," J. Acoust. Soc. Am. **33**, 771-778.

[3] Green and Swets (**1974**). *Signal Detection Theory and Psychophysics* (Kreiger, New York).

[4] Gresham, L. C. and Collins, L. M. (**1998**). "Analysis of the performance of a model-based optimal auditory signal processor," J. Acoust. Soc. Am. **103**, 2520-2529.

[5] Meddis, R. (**1988**). "Simulation of auditory-neural transduction: Further studies," J. Acoust. Soc. Am. **83**, 1056-1063.

[6] Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890-1894.

[7] Pfafflin, S. M., Mathews, M.V. (**1962**). "Energy-detection model for monaural auditory detection," J. Acoust. Soc. Am. **34**, 1842-1853.

[8] Sherwin, C. W., Kodman, F., Jr., Kovaly, J. J., Prothe, W. C., and Melrose, J. (**1956**). "Detection of signals in noise: A comparison between the human detector and an electronic detector," J. Acoust. Soc. Am. **28**, 617-622.

[9] Siebert, W. M. (**1968**). "Stimulus Transformations in the Peripheral Auditory System" in *Recognizing Patterns*, ed. Kolers and Eden, (MIT Press, Cambridge, MA).

[10] Siebert, W. M. (**1970**). "Frequency discrimination in the auditory system: Place or periodicity mechanisms?," IEEE Proc. **58**, 723-730.

[11] Watson, C. S., Rilling, M. E., and Bourbon, W. T. (**1964**). "Receiver-operating characteristics determined by a mechanical analog to the rating scale," J. Acoust. Soc. Am. **36**, 283-288.