Effect of decision weights and internal noise on the growth of d' with N

Robert A. Lutfi

*Auditory Behavioral Research Lab, Department of Communicative Disorders, University of Wisconsin, Madison, Wisconsin 53706
ralutfi@wisc.edu*

Lynn Gilbertson

*Auditory Behavioral Research Lab, University of Wisconsin, Madison, Wisconsin 53706
lrjohnson1@wisc.edu*

Abstract: A general finding of psychoacoustic studies is that detectability d' of a noisy signal grows less than optimally with the number N of independent observations of the signal. Competing accounts implicate internal noise common to all observations or nonoptimal decision weights given to observations. A discriminant analysis of listeners' trial-by-trial responses in a multitone level-discrimination task favored the latter account.

© 2011 Acoustical Society of America

PACS numbers: 43.66.Ba, 43.66.Fe [QJF]

Date Received: July 18, 2011 **Date Accepted:** August 31, 2011

1. Introduction

This paper is concerned with the question of how listeners accumulate information from multiple observations of a signal to improve their ability to detect and discriminate sounds. The question is of fundamental importance in psychoacoustics as it addresses the larger issue of how redundancy in signals, inherent in speech and everyday sounds, can serve to benefit detection and recognition in noise. Past studies have, indeed, shown an improvement in detectability with repeated observations of the signal.¹⁻⁴ However, the rate of improvement is found to be less than ideal for independent observations in excess of $N=4$. These studies employed a sample-discrimination procedure wherein the frequencies of N tones, presented in sequence, were drawn independently and at random on each trial from one of two normal distributions differing in mean. The listeners' task was to identify on each trial whether the tone frequencies were drawn from the distribution with high or low mean. Expressed as d' , ideal performance in this task grows as the square-root of N ; however, the obtained performance was found to grow more nearly as the cube-root of N . This basic result has been replicated for the sample-discrimination of tone intensity^{3,4} and tone duration⁴ and for the tones presented simultaneously as well as sequentially.^{3,4}

Two competing accounts have been given for the less than optimal growth of d' with N . The first attributes the suboptimal growth to an additive source of central internal noise on each trial that is common to all observations.^{1,2,5} In this model, the standard deviation of the central noise σ_C appears as an additive term in the prediction for d' , effectively slowing the rate at which performance grows with N . The alternative account attributes the suboptimal growth to an effective limit on the number of tones that listeners can process simultaneously.⁶ The limit is represented as a nonoptimal weighting of the information provided by the different tones, where some of the decision weights, w_j , are close to or equal to zero. Both factors can be incorporated in a general model for which predicted performance is given by

$$d'_N = \Delta \left(\frac{\sigma^2 + \sigma_{pj}^2}{\sum_j w_j^2} + \sigma_C^2 \right)^{-1/2}, \quad (1)$$

where Δ and σ are, respectively, the difference in means and the standard deviation of the stimulus distributions given in the experiment, σ_{pj} represents the influence of peripheral internal noise, which is independent for each observation, and $0 \leq w_j \leq 1$. Both weighting and internal noise accounts make equally good predictions for d' , given freedom in choosing values for σ_C , σ_{pj} , and w_j . However, the two accounts make quite different predictions for how the listener's trial-by-trial response correlates with individual observations. Zero weight given to some observations requires that some of these correlations be disproportionately large. Internal noise common to all observations requires these correlations to be near equal. We use the general model of Eq. (1) to test these predictions in what follows.

2. Method

The stimuli were N -tone complexes, where N ranged from 1 to 13 in steps of 2. The number of tones was varied in two ways. In the first, the tones (100 ms in duration) were presented simultaneously with frequencies spaced at equilog intervals from 250 to 7890 Hz ($f_{j+1}/f_j = 1.33$). This spacing ensured negligible mutual masking among tones.⁷ The number of tones was increased beginning with the lowest frequency tone and ending with the highest. In the second condition, the tones (30 ms in duration) were presented sequentially without silent gaps and with a fixed frequency of 2000 Hz. Individual tones were gated on and off with 5-ms, cosine-squared ramps. In both conditions, the individual levels of the tones on each trial comprised a random sample of size N from one of two, equivariate, normal level distributions differing in mean ($\mu_{\text{LOW}} = 65$ and $\mu_{\text{HIGH}} = 68$ dB SPL, $\sigma = 3$ dB). On each day of the experiment, a total of 50 different N -tone complexes were generated representing 50 random samples of levels from each distribution; we will refer to these as the high- and low-level set. All tone complexes were played over a 16-bit, audio-quality, digital-to-analogue converter at a 20-kHz rate and were low-pass filtered (8-kHz cutoff, 120 dB/octave) to remove aliased components.

A fixed, two-interval, forced-choice (2IFC) procedure was used to measure d' in each condition. On each trial, the listener heard a pair of the N -tone complexes separated by approximately 350 ms. One complex of the pair was randomly selected from the low-level set, the other from the high-level set of tone complexes. The listener's task was to identify the interval containing the complex drawn from the high-level set. The value of N was fixed for each block of 50 trials, and feedback was given after each trial.

Conditions were run in pseudorandom order and replications were made across days. A single percent correct score was obtained from a total of at least 1000 trials per subject per condition. The mean percent correct scores were then converted to d' values for the 2IFC procedure.⁸ Three female students from the University of Wisconsin–Madison, ranging in age from 18 to 26 years, participated as paid listeners. They ran a single 2-h session per day, five days of each week. Sessions included frequent breaks. All listeners received two or more weeks of training in the 2IFC task before data collection began.

3. Results

Figure 1 gives the d' results (symbols) for the three listeners (S1–3) for both simultaneous and sequential presentation of tones (panel columns). The values of d'_N are normalized with respect to d'_1 ($N = 1$) to facilitate comparison of the growth rate of d' with N across conditions. The dashed curve gives the optimal $\sqrt{N}$ rate of growth. The data replicate earlier studies showing closer to cube-root of N growth. The red curve gives the prediction of the internal noise model (σ_C ranging from 0.13 to 0.65 dB across listeners and conditions). The blue curve represents the weighting model where the weights are treated as free parameters. The weighting model, of course, describes the data perfectly as the data are overdetermined by the number of free parameters in the model. The comparison is made, however, to underscore that the d' data do not adequately distinguish the predictions of the two models.

Fig. 1. Open symbols give obtained values of d'_N , expressed relative to d'_1 for the three listeners (S1–3) for both simultaneous and sequential presentation of tones (columns). The optimal $\sqrt{N}$ growth rate of d'_N is given by the black curve. The predictions of the internal noise model and the weighting model are given, respectively, by the red and blue curves. The d' data do not distinguish the predictions of the two models.

To distinguish the predictions of the models, it is necessary to measure the strength of the relation between the listener's response and the individual observations from trial to trial. This was done by means of a discriminant analysis of the trial-by-trial data based on a logistic model. Let $P_{R=2}$ denote the probability of responding that the second interval contained a signal from the high-level set, and let ΔL_j represent the difference in the levels of the j th tone across the two observation intervals (second minus first). The logistic model is given by

$$\text{logit}(P_{R=2}) = b_0 + \sum_{j=1}^N b_j \Delta L_j \quad (2)$$

where $\text{logit}(\cdot)$ is the inverse logistic function. The strength of the relation between the listener's response and the individual observations is given by the normalized regression coefficients b_j in Eq. (2); $\sum b_j = 1$. These values were estimated using the GLMFIT routine of the software application MATLAB v.7.0 and are shown in Fig. 2. (Note: symbols and error bars have been omitted for clarity of presentation.) The regression weights reveal a tendency of listener responses to be most strongly influenced by the highest and lowest frequency tones for simultaneous presentation of tones and the first and last tones for the sequential presentation. This result is consistent with the predictions of weighting model. However, it remains to be determined whether the effect can account entirely for less than optimal growth rate of d' with N . To this end we undertook an analysis of weighting efficiency.^{6,9} Let d'_{wgt} represent the performance of each listener predicted exclusively from their relative decision weights in Eq. (1); that is, assuming no other limits imposed by internal noise ($\sigma_C = \sigma_{Pj} = 0$). Weighting efficiency is then defined as

Fig. 2. The normalized regression coefficients b_i obtained from the logistic regression model of Eq. (2) are plotted for each listener (panel rows) and condition (panel columns) as a function of N . Symbols have been omitted for clarity of presentation.

Fig. 3. For each condition (panel columns) and listener (panel rows), the obtained values of η_{obt} (symbols), η_{wgt} (blue curve) and η_{nos} (red curve) are plotted as a function of N . The horizontal black curve gives ideal performance. See text for further details.

$$\eta_{\text{wgt}} = \left(d'_{\text{wgt}} / d'_{\text{ideal}} \right)^2, \quad (3)$$

where d'_{ideal} is the performance of an optimal detector yielding $\sqrt{N}$ growth of d' . We next define an overall performance efficiency η_{obt} representing the combined influence of the weighting efficiency and internal noise,

$$\eta_{\text{obt}} = \left(d'_N / d'_{\text{ideal}} \right)^2 = \eta_{\text{wgt}} \times \eta_{\text{nos}}, \quad (4)$$

where the influence of internal noise $\eta_{\text{nos}} = (d'_N / d'_{\text{wgt}})^2$ is simply that component of performance not accounted for by the decision weights. The weighting model now predicts that η_{wgt} should decrease with N at the same rate as η_{obt} , while η_{nos} remains constant. Conversely, the internal noise model predicts that η_{nos} should decrease with N at the same rate as η_{obt} , while η_{wgt} remains constant.

To test these predictions, we computed η_{wgt} and η_{nos} taking the normalized regression coefficients as estimates of the listeners decision weights, $w_j = b_j$. The results are shown in Fig. 3; here η_{wgt} and η_{nos} are given by the blue and red curves, respectively, η_{obt} is given by the open symbols and the black line represents ideal performance. The results clearly support the weighting model. While the effect of internal noise, as measured by η_{nos} , serves to degrade overall performance, the less than optimal growth rate of performance with increasing N corresponds to the reduction in weighting efficiency with increasing N .

4. Conclusion

The results of the present study replicate those of past studies showing suboptimal growth of d' with N in a multitone level discrimination task. In past studies, competing accounts in terms of decision weights and internal noise are not meaningfully distinguished by their predictions for d' . In the present study, however, the weighting model is clearly supported by a discriminant analysis showing listener responses to depend largely on the “bracketing” tones within a stimulus; an effect that accounts entirely for the suboptimal growth of d' with N .

References and links

- ¹B. G. Berg, “Internal noise in auditory decision tasks,” Ph.D. dissertation, Indiana University, Bloomington, IN, 1987.
- ²B. G. Berg and D. E. Robinson, “Multiple observations and internal noise,” *J. Acoust. Soc. Am.* **81**, S33 (1987).
- ³R. A. Lutfi, “Informational processing of complex sound. I. Intensity discrimination,” *J. Acoust. Soc. Am.* **86**, 934–944 (1989).
- ⁴R. A. Lutfi, “Informational processing of complex sound. III. Cross-dimensional analysis,” *J. Acoust. Soc. Am.* **87**, 2141–2148 (1990).
- ⁵N. I. Durlach, L. D. Braida, and Y. Ito, “Towards a model for discrimination of broadband signals,” *J. Acoust. Soc. Am.* **80**, 63–72 (1986).
- ⁶R. A. Lutfi and C. J. Liu, “Individual differences in source identification from synthesized impact sounds,” *J. Acoust. Soc. Am.* **122**, 1017–1028 (2007).
- ⁷B. Scharf, “Critical bands,” in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York, 1970), pp. 159–202.
- ⁸J. A. Swets, *Signal Detection and Recognition by Human Observers* (Wiley, New York, 1964), pp. 147–198.
- ⁹B. G. Berg, “Observer efficiency and weights in a multiple observation task,” *J. Acoust. Soc. Am.* **88**, 149–158 (1990).