JARO 02: 189–198 (2001)
DOI: 10.1007/s101620010074

JARO
Journal of the Association for Research in Otolaryngology

Informational Masking in Profile Analysis: Comparing Ideal and Human Observers

VIRGINIA M. RICHARDS[1] AND TAO ZENG[2]

[1]*Department of Psychology, University of Pennsylvania, Philadelphia PA, 19104, USA*
[2]*Department of Bioengineering, University of Pennsylvania, Philadelphia PA, 19104, USA*

Received: 30 August 2000; Accepted: 19 February 2001; Online publication: 1 August 2001

ABSTRACT

Predictions from an ideal observer model are compared with human thresholds for two profile analysis tasks. Past work has shown that ideal observer models reasonably account for human thresholds when the profile components are fixed in frequency and amplitude. Randomly varying the frequencies of the tones making up the profile leads to higher thresholds. Owing in part to large interobserver variation, the ideal observer model is not successful in accounting for the pattern of psychophysical thresholds associated with increases in frequency uncertainty. The ideal observer also fails to account for the results of a recent profile analysis experiment in which amplitude randomization was studied [Lentz JJ, Richards VM: J. Acoust. Soc. Am. 102: 535–541, 1998]. Overall, the ideal observer predicts *smaller* effects of uncertainty on thresholds than are observed in psychophysical experiments.

Keywords: auditory profile analysis, ideal observer

INTRODUCTION

For the detection of a tone added to multitone maskers, substantial masking can be induced by randomly choosing the frequency of the maskers on different trials, and there is even more when the frequencies are randomly chosen on each stimulus presentation (cf. Neff and Green 1987; Neff and Callaghan 1998). The additional masking associated with increases in masker uncertainty is referred to as informational masking. When the number of masker tones is fixed, increasing the range across which the frequencies of the masker components are drawn does not necessarily lead to increases in masking (Neff and Callaghan 1988). Oh and Lutfi (1998; see also Neff et al. 1993) argue that informational masking is reduced as more masker tones are introduced, but that reductions in informational masking associated with more masker tones are offset by increases in energetic masking associated with the increased likelihood that one or more masker tones will be close to the signal frequency. Uncertainty can also be manipulated by varying the number of different maskers in the set of potential maskers; Wright and Saberi (1999) found that informational masking increases as the number of potential maskers increases. Moreover, the magnitude of the informational masking varies substantially across observers (cf. Neff and Dethlefs 1995). In all, as suggested by Neff and Callaghan (1988, p. 1838), “the dynamic properties of the maskers appear to interfere with such detection, although listeners are instructed to ignore the interfering stimuli and focus on the signal.”

Informational masking has also been reported in profile analysis experiments. Spiegel et al. (1981) examined the detection of an increment in level to one of many simultaneously presented equal-amplitude tones (2–20 tones were tested). They examined the effects of randomly choosing the frequencies of the tones on a trial-by-trial basis (i.e., the frequencies were fixed across intervals) and found thresholds increased on average approximately 4 dB relative to threshold measured when the component frequencies

Correspondence to: Dr. Virginia M. Richards · Department of Psychology · University of Pennsylvania · 3815 Walnut St. · Philadelphia, PA, 19104. Telephone: (215) 898-4587; fax: (215) 898-7301; email: richards@psych.upenn.edu

were fixed. In contrast, randomly choosing the frequency of the signal did not, on average, lead to a change in threshold compared with thresholds measured when the signal frequency was fixed. Richards et al. (1989) examined various types of spectral shape discriminations and evaluated the effects of randomizing the digital-to-analog converter (DAC) output rate. Changing the DAC output rate scales all frequencies present, or translates the magnitude spectrum toward lower and higher frequencies on a logarithmic axis. When the range of frequency shift extended beyond an octave, discrimination thresholds were substantially increased relative to the fixed-frequency condition. Gockel and Colonius (1997) found that when the frequency shift extended over 3 octaves, thresholds for detecting differences in spectral shape were so large as to be unmeasurable.

To date there have been no profile analysis studies in which presentation-by-presentation frequency randomization of the individual components has been a manipulation of interest. Presumably this is because randomly choosing component frequencies leads to components that interact within single auditory filters. Under those circumstances, the excitation pattern is varied in an uncontrolled manner. If instead of varying the frequency of the components the magnitudes of the components are varied, the change in the excitation pattern is more easily appreciated. Kidd et al. (1986) and others (Berg and Green 1990; Kidd et al. 1991; Lentz and Richards, 1998) found that introducing amplitude perturbation leads to higher thresholds in profile analysis tasks. Likewise, introducing "interfering" tones with random amplitudes degrades sensitivity to changes in the relative magnitudes of target tones, provided the interfering and target tones are gated on and off together (Hill and Bailey 1998). In these cases, the effect of amplitude perturbation is thought to be mediated by substantially degraded long-term standards against which the test stimuli (or the target tones) can be compared.

Kidd et al. (1988) incorporated the intensity discrimination model set out by Durlach and Braida (1969) in arguing that the decision processes associated with the discrimination of changes in spectral shape might depend on both short-term (sensory trace) representations, which are volatile across time and useful for only short time epochs (e.g., across intervals), and long-term (context coding) representations, which develop with experience. By this approach, varying the spectral patterns on a presentation-by-presentation basis would lead to a dependence on variable long-term representations (Durlach and Braida 1969; Kidd et al. 1988). While Kidd et al. (1988) were primarily interested in studying the role of the short-term trace mode in profile analysis experiments, they also used stimuli composed of tones whose amplitudes were randomly chosen for each stimulus presentation. The results support the conclusion that discrimination depends primarily on context-coding mechanisms in these conditions.

In the present experiment, effects of frequency perturbations of individual components in profile analysis experiments are considered. Because frequency perturbation is present, substantial changes in excitation patterns are introduced. The primary empirical question addressed in this article is whether a Bayesian ideal observer analysis applied to excitation patterns can predict the pattern of human thresholds. Put another way, we were interested in determining whether uncertainty effects in profile analysis experiments could be accounted for in terms of the stimulus present at the periphery. Two types of profile analysis discriminations are tested: down–up vs. flat and 1-step vs. flat. The "flat" stimulus is composed of equal-amplitude tones. The down–up stimulus has amplitudes that vary low–high– . . . –low–high. For the 1-step stimulus, the low-frequency tones have low amplitudes and the high-frequency tones have high amplitudes.

Richards et al. (1989) used both discrimination types and found that modest DAC output rate randomization (sample periods ranging from 40 to 45 μs; sample frequencies ranging from 22,222 to 25,000 samples/s) led to threshold shifts of about 4 dB for both discrimination types. In the current study, the spacing between components is much wider than that tested in the Richards et al. study (8 rather than 21 components equidistant on a logarithmic frequency scale ranging from 200 to 5000 Hz; or the ratio between adjacent tone frequencies is 1.58 rather than 1.17). As a result, at least when the component tones are fixed in frequency, the auditory filters centered at the component tones may be treated as independent of one another (cf. Lentz et al., 1999). Additionally, in the present study, frequency randomization is not achieved by DAC randomization. Rather, the frequencies of the components are chosen from a (log) uniform distribution centered at the frequency each tone holds when frequencies are fixed. Note that when frequency uncertainty is introduced, for the 1-step stimulus all components below the geometric mean of 1000 Hz have lower magnitudes and the components above 1000 Hz have higher magnitudes. It seemed possible that this stimulus construction would "protect" the 1-step condition from effects of frequency randomization compared with the down–up condition. In the 1-step condition, the signal vs. no-signal decision may be made by comparing the level of any one (or more) of the tones with frequencies below 1000 Hz against the level of any one (or more) of the tones with frequencies higher than 1000 Hz. This is true regardless of the degree of frequency randomization. For the down–up condition, fine-tuned comparisons between adjacent tones are required.

There are several reasons to believe that an ideal observer will be at least partially successful in accounting for human profile analysis thresholds in the presence of frequency randomization. First, the Durlach et al. (1986) ideal observer channel model has been somewhat successful in accounting for profile analysis data (cf. Berg and Green, 1990; Green 1992; Lentz and Richards 1997). Second, profile analysis is thought to depend on comparisons of spectral shape; thus, randomization in the dimensions of frequency and amplitude are rationally incorporated into an ideal observer model of profile analysis. Third, it is reasonable to assume that, as has been suggested for amplitude perturbation (Kidd et al. 1988), frequency perturbation encourages observers to depend on context coding in making their decisions. While it is not assured, at face value it seems reasonable that a dependence on long-term representations is similar to comparisons with state-conditional probabilities (e.g., the probability of observing vector x given signal; the probability of x given no signal) used to form likelihood ratios in ideal observer analyses. In all, it seems that the effects of frequency randomization on profile analysis data is a problem well suited to ideal observer analysis.

There are, nonetheless, several reasons why an ideal observer analysis may fail to account for human thresholds when frequency randomization is introduced in profile analysis tasks. First, any or all of the assumptions listed above may be incorrect. Second, as described below, the ideal observer model depends on an analysis of excitation patterns. As a result, errors in the description of the excitation pattern may undermine the comparison between ideal and human thresholds. Third, effects of frequency uncertainty may depend on central rather than peripheral factors. If this is so, the ideal observer model will fail.

As suggested above, the predictions of an ideal observer model depend fundamentally on the assumed stimulus representation. For example, if the stimuli are represented as FFT-based magnitude spectra, no effect of frequency randomization is expected. Wherever the components fall, the obtained magnitudes relative to the mean are uniquely associated with either the signal (down–up or 1-step) or the no-signal (flat) stimuli. When frequency selectivity is imperfect, as for excitation pattern models, the excitation patterns do not reliably differentiate between stimuli unless the auditory filters are narrow relative to the smallest frequency separation between two components. Two sinusoidal components falling within the passband of a single auditory filter lead to local increments in the excitation pattern, as does a single relatively intense sinusoid. When frequency selectivity is degraded, frequency randomization and intensity variation are difficult to disentangle. For the detection of changes in spectral pattern, intensity variations are known to reduce sensitivity (cf. Kidd et al. 1986; Berg and Green 1990; Kidd et al. 1991; Lentz and Richards 1999). Thus, for excitation pattern models, frequency randomization should also reduce sensitivity relative to conditions in which there is no frequency randomization.

METHODS

Human observers

Two, 4, or 8 tones made up the stimuli (N, number of components). For the 8-component stimulus, the tonal components were equally spaced on a logarithmic frequency scale ranging from 200 to 5000 Hz. When fewer than 8 tones were used, the frequency spacing between tones was the same but the central 2 or 4 tones were the only ones presented.

Two spectral shape discriminations were tested: down–up vs. flat and 1-step vs. flat. For the former, the stimuli were composed of either equal-amplitude tones (flat) or tones with magnitudes that varied down–up . . . down–up. For 1-step stimulus, the low-frequency tones (one-half of all tones present) were of lower magnitude and the high-frequency tones were of higher magnitude. The phases of each tone were randomly chosen on each presentation from a uniform distribution with a range of 2π. Similarly, the overall level was randomly chosen on each presentation from a 30-dB range using 0.1-dB steps. Thresholds are reported as ΔL, the change in level, up or down, in dB relative to the mean. A threshold of 9.1 dB (as ΔL) is the minimum threshold that can be supported based on changes in the level of a single component.[1]

In the Fixed condition, the frequencies of the components did not change. In the three random conditions, the frequency of each component tone was chosen randomly on each stimulus presentation. This was achieved by multiplying the frequency of each tone by a randomly chosen scalar. The end result is that the distribution of possible frequencies was uniformly distributed on a logarithmic scale and the midpoint was the component's frequency in the Fixed condition.

[1] Fantini and Moore (1994; see Green 1988) describe the signal level required for the detection of an increment in level in a 3IFC procedure. Denoting the percent correct level used to define threshold as PC, the range of randomization as R, and the change in level as C, one obtains

$$PC = \frac{C}{R} + \frac{1}{3} - \frac{1}{3}\left(\frac{C}{R}\right)^3$$

For the most extreme degree of randomization, the Max Ran condition, the end points of the uniform distributions were halfway between (geometric mean) the frequencies of neighboring tones. For the other two random conditions, Mid Ran and Min Ran, the procedure was the same except that the range was smaller (achieved by taking the square root and 4th root of the scalar relative to the Max Ran condition, respectively). Consider as an example the component which, in the Fixed condition, was centered at 795 Hz. In the Max Ran condition, its frequency was chosen from a range of 630–1000 Hz. For the Mid Ran and Min Ran conditions, the range was 710–890 Hz and 750–840 Hz, respectively. Because the stimulus duration was 200 ms, including 10-ms raised cosine onset/offset ramps, the frequency gradation was 5 Hz.

The stimuli were digitally generated and presented through two channels of a 16-bit DAC using a sampling rate of 20,000 samples/s, lowpass filtered at 7 kHz using matched filters (KEMO VBF 8), and presented diotically by way of two channels of Sennheiser HD410SL headphones. The interstimulus interval was approximately 450 ms. The component tones each had a mean level of 50 dB SPL.

The stimuli were presented using a 3IFC procedure, with the signal being equally likely to be present in any of the three intervals. Thresholds were estimated using a 3-down, 1-up staircase procedure, which estimated the 79% correct performance level (Levitt 1971). Initial signal levels and step sizes varied depending on condition. In the Fixed condition, the initial step size was 0.6 dB, which was reduced to 0.3 dB following three reversals. In the random conditions, the initial and final step sizes were 2 and 1 dB, respectively. On four occasions (three for Obs 3 and one for Obs 4), the tracking procedure attempted to assign a negative ΔL, which was disallowed. The initial value of ΔL was approximately 2–4 large steps greater than the ultimate threshold. After observers practiced (see below), 15 threshold estimates were obtained for each condition tested. The final 10 were averaged to provide the final threshold estimate.

All observers practiced for at least 10 hours prior to data collection. Data collection was blocked. For Obs 3 and Obs 4, the discrimination type was blocked (down–up vs. fixed or 1-step vs. fixed). Then, for each type of discrimination, both observers ran the Fixed condition first and the other conditions were tested in random order. Within each condition, the order in which the different numbers of components were run was chosen quasirandomly for each observer. For Obs 1 and Obs 2, each discrimination type was run initially for the Fixed condition and then for the Max Ran conditions. Then, for each discrimination type, the Min Ran and Mid Ran conditions were run.

Before starting a new condition, observers practiced briefly in the new condition. If practice effects were apparent in the 15 threshold estimates obtained in any one condition, data collection was repeated. Reevaluations of thresholds were not common but did occur. Because the pattern of results varied across the four observers, it is difficult to evaluate the potential effect of long-term practice.

Observers had thresholds in quiet of 10 dB HL or better (for frequencies ranging from 250 to 8000 Hz), except that Obs 2's threshold at 500 Hz (right ear) was 25 dB HL. The observers ranged in age from 19 to 27 years and were paid for participation. One of the four observers had prior experience in psychoacoustic tasks. Tests were conducted with the observer seated in a double-walled sound-attenuated booth.

Quasi-ideal observer analysis

The quasi-ideal observer analysis relied on a general Bayesian pattern recognition approach and the assumption that the distributions of interest are multivariate normal (see Duda and Hart 1973). Because it is assumed that the task relies solely on differences in level at the output of auditory filters, and because the normal assumption is used (and, in restricted simulations, shown to provide reasonably accurate results), the ideal observer analysis is referred to as quasi-ideal. For example, effects of suppression, potential temporal/envelope cues, etc., are not evaluated. Computer simulations were used to evaluate the performance of the quasi-ideal observer using Matlab 5.3 (The Math Works, Inc. 1996). Initially 1000 "signal" and 1000 "no-signal" stimuli were generated. The stimuli were passed through a linear version of the single-parameter RoEx(p) filters [weighting function $W(g) = (1 + pg)e^{-pg}$ where g is the deviation from the center frequency relative to the center frequency; cf. Patterson and Moore 1986]. In most instances, the equivalent rectangular bandwidth (ERB) was set according to the recommendations of Glasberg and Moore (1990) and a stimulus level of 51 dB/ERB. Because Lentz et al. (1999) found little effect of including filter nonlinearities in their evaluation of the Glasberg and Moore (1990) filters applied to profile analysis stimuli [using RoEx(p,r) filters rather than the simpler RoEx(p) filters used here], level-dependent nonlinearities are not incorporated. In some exploratory simulations, ERBs approximately half and double those recommended by Glasberg and Moore (1990) were evaluated. The frequency axis was defined between 0 and 10,000 Hz using 2-Hz step sizes, meaning that the frequency gradation was finer than in the psychophysical experiment.

In most instances, 41-auditory filters were placed with center frequencies equidistant on a logarithmic scale, with the lowest and highest center frequencies taking on values of 126 and 8000 Hz, respectively.[2] It was assumed that only the levels at the outputs of the auditory filters contribute to the decision rule, and so filtering was achieved by multiplying the power spectra and the filter weighting function and then summing power to estimate the power passed by each filter. The output of each filter was expressed on a dB scale and set to threshold if the output of the filter was below threshold. Then, independent 4-dB zero-mean normal deviates were added to the output of each filter. The added "channel noise" prevents performance from being perfect when frequencies are fixed. It was set to 4 dB so that in the Fixed condition the model and averaged human data were about the same. Additional details of the filtering procedures are described in Lentz et al. (1999).

The computational procedure used to generate predictions is essentially equivalent to using the decision rule:

$$\text{Choose "signal" if } \frac{p(x/s)}{p(x/n)} > 1,$$

$$\text{otherwise, choose "no signal"}$$

where $p(x/s)$ and $p(x/n)$ are the state-conditional probability density functions for the "signal" and "no-signal" states, respectively. As indicated above, the computational method included an added assumption that the distributions are multivariate normal. Thus, computationally, the following steps were carried out. The initial 1000 signal and 1000 no-signal stimuli were passed through the filter bank yielding 2000 m-dimensional excitation patterns, where m is the number of filters used. Then, the summary statistics for the signal and no-signal excitation patterns were derived. Spot checks indicated the signal and no-signal covariance matrices were essentially the same, so they were averaged to provide a single estimate of the covariance matrix (Duda and Hart 1973). Next, 1000 signal and 1000 no-signal test stimuli were generated, and the Mahalanobis[3] distance between each test stimulus and the signal and no-signal mean vectors was computed. When the test stimulus was "nearer" the signal distribution, a signal response was assigned. Otherwise, a no-signal response was assigned.

The simulation led to hit and false alarm rates based on the 1000 signal and 1000 no-signal test stimuli. Then, treating the discrimination as a single-interval task, the hit and false alarm rates were converted to d' scores. The process was repeated anew for three different signal levels that led to d' values between approximately 0.5 and 1.5. Then, using a linear fit, the ΔL required for $d' = 1$ was estimated. This process was repeated for the different Ns tested and for the different conditions tested. The ultimate results are based on the average of two simulation replicates.

RESULTS AND DISCUSSION

Human observers

Figure 1 shows the results for the individual observers in different panels. Thresholds expressed as ΔL in dB are plotted as a function of N. The top panel of Figure 2 shows the results averaged across observers. In Figure 1, error bars indicate the standard errors of the mean across 10 threshold replicates whereas in Figure 2, error bars indicate the standard errors of the mean across the 4 observers. Filled symbols indicate the down–up vs. flat discrimination and open symbols indicate the 1-step vs. flat discrimination. Fixed, Min Ran, Mid Ran, and Max Ran conditions are indicated using squares, circles, triangles, and upside down triangles, respectively.

The most striking result is the variation in the pattern of the data across observers. An analysis of variance (ANOVA) revealed only one significant effect, a main effect of frequency randomization ($F(9,3) = 42.4$, $p < 0.001$). The effect of discrimination type neared statistical significance ($p \approx 0.07$) but the remaining comparisons did not. It is remarkable that there is an effect of frequency randomization when the stimuli are composed of only two components (Fig. 2). In this case, the lower-frequency tone is always associated with an intensity decrement relative to the higher-frequency tone; there is little for the observer to keep track of. Nonetheless, substantial increases in thresholds are obtained. Spiegel et al. (1981) likewise found effects of uncertainty when just two tones were tested, but it is more impressive in their experiment because the frequency of the more intense of the two tones was fixed.

Considering the individual data, it should be kept in mind that for signal levels above 9 dB or so, observers might, in principle, perform the task using level information available for just one component. Nine dB is

[2] When the Glasberg and Moore (1990) ERBs were adopted and when 30 or more filters were used, simulations indicated that increasing the number of filters had little impact on the pattern of results (although the amount of "channel noise" required to match thresholds increased with number of filters). The number of filters required for "stable" performance (e.g., results that did not depend critically upon the number of filters tested) depended upon filter bandwidths.

[3] For a sample vector x the squared Mahalanobis is given by $(x - \mu^t \Sigma^{-1}(x - \mu)$, where t denotes transpose μ and σ are the mean vector and covariance matrix for the multivariate normal distribution, for either the signal or the no-signal stimuli. For example, if the excitation pattern is based on the outputs of 41 filters, x and μ are vectors of length 41 and σ is a 41×41 matrix.

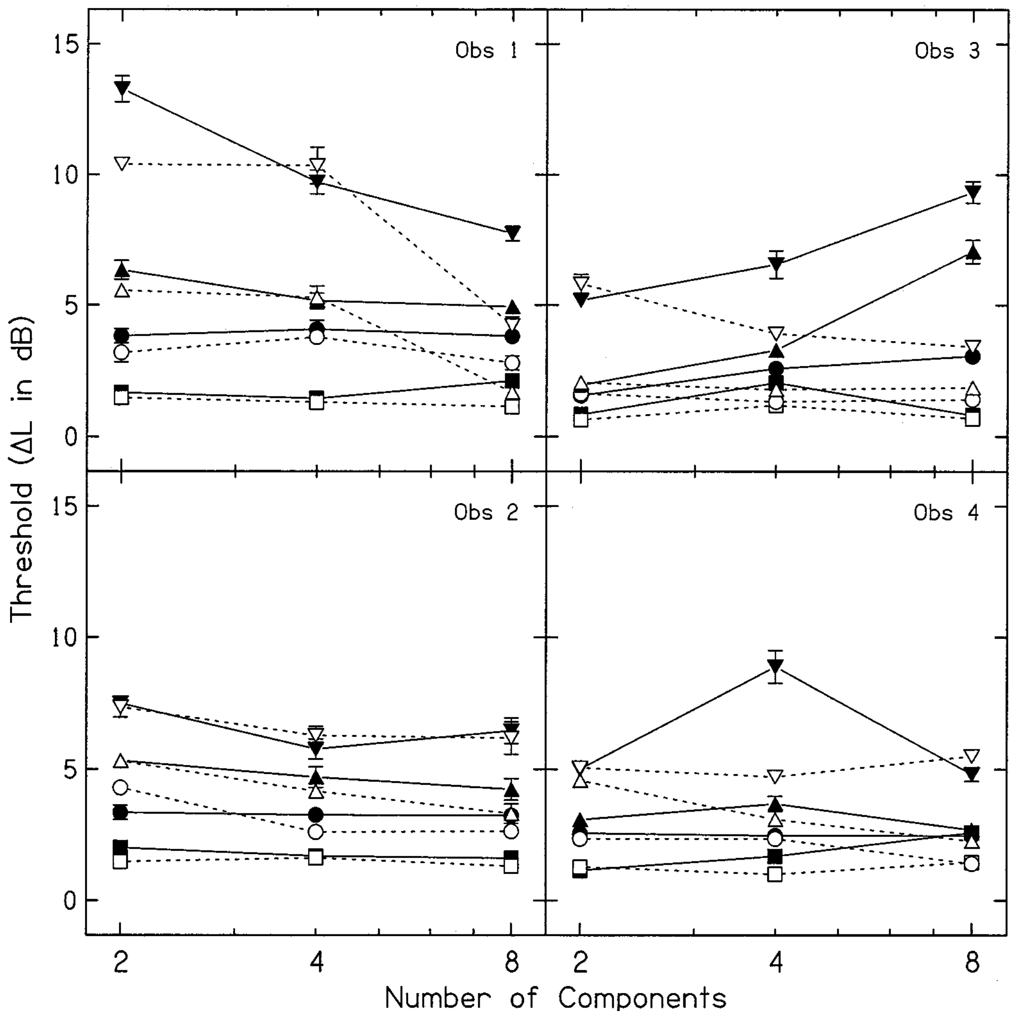

FIG. 1. The data for individual observers are plotted in separate panels. Error bars indicate the standard errors of the mean across 10 threshold replicates. The filled symbols are for the down–up vs. flat discrimination; the unfilled symbols are for the 1-step vs. flat discrimination. The parameter indicates the condition: Fixed (□, ■), Min Ran (○, ●), Mid Ran (△, ▲), and Max Ran (▽, ▼).

a conservative estimate; when frequency randomization is applied, the expected threshold at a single frequency locus is sure to be much higher than 9 dB. For Obs 1 in the Max Ran condition, $N = 2$ and $N = 4$, thresholds exceed this limit. This also holds for Obs 3 in the Max Ran condition in the down–up vs. flat discrimination. Even though there is no difference in the stimuli, on occasion thresholds measured with $N = 2$ vary depending on the discrimination type (e.g., Obs 1, Max Ran and Obs 2, Mid Ran), but in the main the thresholds are similar. This result suggests that long-term practice effects are small, if present at all.

Quasi-ideal observer

The bottom panel of Figure 2 shows the results for the quasi-ideal observer. The abscissa is the number of components and the ordinate is threshold expressed as ΔL in dB. Unfilled symbols are for the 1-step vs. flat discrimination and filled symbols indicate the down–up vs. flat discrimination. The parameter is the degree of frequency randomization: Fixed (squares), Min Ran (circles), Mid Ran (triangles) and Max Ran (upside down triangles). Note that the ordinate is expanded relative to the upper panel by a factor of 3 1/3. Thresholds fall with increasing numbers of components, and the impact of frequency randomization is modest except when the randomization is the maximum tested (Max Ran). For the Max Ran and Mid Ran conditions, there is an interaction between N and discrimination type such that thresholds in the 1-step vs. flat discrimination condition fall more rapidly than in the down–up vs. flat discrimination. For the Fixed and Min Ran conditions, there is little difference between thresholds for the two discrimination types and thus no interaction is apparent.

These results may be compared with expectations for an optimal model when there are no auditory fil-

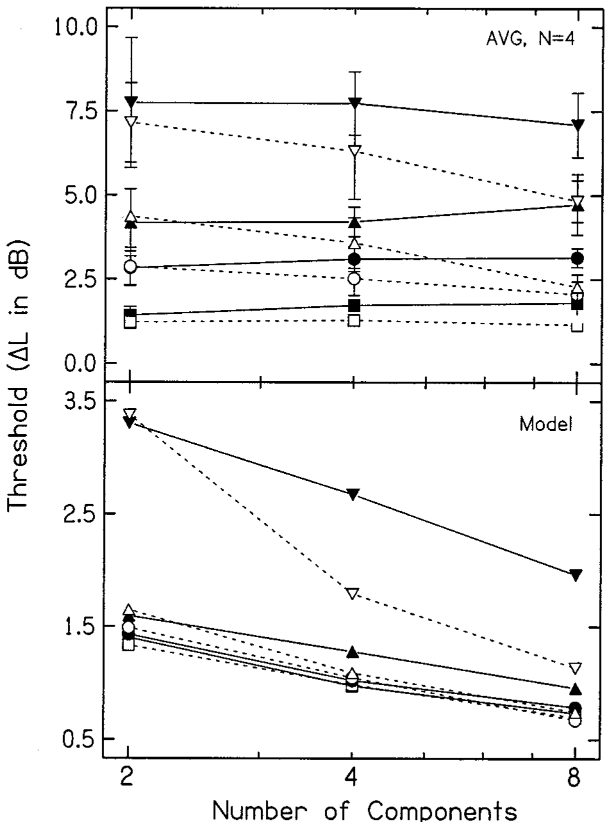

FIG. 2. The top panel shows the data averaged across 4 observers; error bars indicate one standard error of the mean across observers. The bottom panel plots the predictions of the quasi-ideal observer model. Note that scale for the lower panel is expanded relative to the top panel. Symbols are as in Fig. 1.

ters. For the fixed condition, the ideal channel model (cf. Durlach et al. 1986) predicts that thresholds should fall as $1/\sqrt{N}$. Moreover, thresholds should be the same for the two discrimination tasks ("balanced stimuli" in Durlach et al., 1986). The current model likewise generates thresholds that fall as $1/\sqrt{N}$ for the Fixed and Min Ran conditions. In the Max Ran and Mid Ran conditions, for the flat vs. down–up discrimination the function relating thresholds to N is shallower than the $1/\sqrt{N}$ prediction. In the flat vs. 1-step discrimination, the slope is steeper than $1/\sqrt{N}$.

In additional simulations, we examined the effects of changes in several aspects of the model. In none of the simulations were large changes in the effects of frequency randomization noted. Increasing the number of filters led to better overall performance (all other factors being fixed), but the pattern of results was similar to that shown in the bottom panel of Figure 2. Increasing the filter bandwidth led to lower thresholds, whereas decreasing the filter bandwidth led to higher thresholds. The magnitude of the shift depended slightly on the magnitude of the threshold. For example, halving the bandwidths led to larger increases for $N = 2$ than $N = 8$, but even for $N = 2$, the shift was restricted to the Max Ran conditions. Even in the Max Ran condition, the shift was modest. Decreasing the standard deviation of the added Gaussian deviate lowered thresholds overall but left the pattern of predictions largely unaltered.

Comparing human and ideal observers

The quasi-ideal observer model clearly fails to predict the magnitude of observed threshold shifts with increases in the magnitude of frequency randomization. The ideal observer model predicts a substantial rise in threshold only for the Max Ran condition. As indicated above, additional simulations indicate that changes in the variance of the added channel noise and changes in the number of filters do not lead to notable increases in the effects of uncertainty. When filter bandwidths are reduced, there is a slightly larger effect of frequency randomization, although the impact is mainly in the Max Ran condition. In contrast to the model predictions, the human data indicate effects of level randomization even when the shift is from Fixed to Min Ran.

The interaction predicted by the quasi-ideal observer was not found for all observers. As shown in Figure 2, the ideal observer model predicts that when effects of frequency randomization are obtained, an interaction between N and discrimination type is also obtained. For the human observers, only Obs 1 and Obs 3 tended to show this result (Fig. 1). For Obs 1, the divergence with N reflects the relatively low thresholds measured in the Max Ran and Mid Ran, $N = 8$ conditions in the 1-step vs. flat discrimination. For Obs 3, the interaction owes as much to the *rise* in threshold that occurs as N increases in the down–up vs. flat discrimination as to the reduction in thresholds obtained as N increases in the 1-step vs. flat discrimination.

RETROSPECTIVE ANALYSIS OF AMPLITUDE PERTURBATION DATA

When the frequencies of the individual components are randomly chosen, human data indicate a larger effect of uncertainty than the quasi-ideal observer. Past experiments concerning the impact of amplitude randomization on profile analysis studies have shown that human thresholds increase with increases in the magnitude of the amplitude randomization. Moreover, Berg and Green (1990) found that even though thresholds increase with increases in the magnitude of amplitude perturbation, the decision rule appears to remain stable and near optimal. Whether the increase in thresholds reflects a shift in efficiency in addition to a shift associated with increases in variability is not

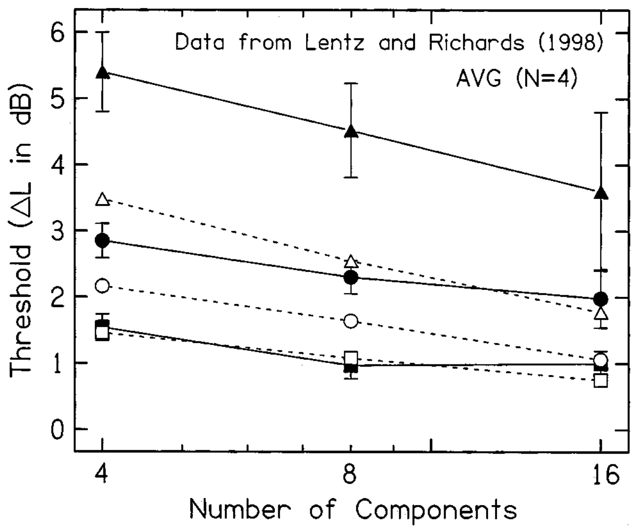

FIG. 3. Averaged data from Lentz and Richards (1998) are plotted using solid symbols and error bars show the standard errors of the mean across 4 observers. The open symbols plot the predictions of an ideal observer model. The parameters indicate the degree of amplitude variation: none (Fixed; □, ■), σ = 3 dB (○, ●), and σ = 6 dB (△, ▲).

addressed; a change in efficiency that is correlated with increases in variability would suggest that the ideal model falls short of accounting for the effects of uncertainty when amplitude perturbation is present. Here, data reported by Lentz and Richards (1998) are considered.

Observers discriminated between stimuli with a "tophat" vs. flat profile. For the tophat shape, the middle components were incremented in level relative to the mean and the outer components were decremented relative to the mean. For example, for an 8-component tophat stimulus, the two lowest- and two highest-frequency components had lower amplitudes and the four middle components had higher amplitudes. *N*s of 4, 8, and 16 were tested. The components were equidistant on a logarithmic frequency axis such that when 16 tones were tested they ranged in frequency from 200 to 5000 Hz. Thus, the individual components were nearer in frequency than in the current experiment. Three conditions were tested. In the Fixed condition, no amplitude variation was applied. In $\sigma = 3$ and $\sigma = 6$ conditions, a zero-mean, normally distributed deviate was independently chosen and added to each component of the profile (as ΔL in dB). The standard deviation of the normal distribution was either 3 or 6 dB.

The filled symbols in Figure 3 show the results of the Lentz and Richards' (1998) Experiment II, averaged across their 4 observers. The results for the Fixed, 3-, and 6-dB conditions are indicated by filled squares, circles, and triangles, respectively. Error bars indicated the standard errors of the mean across observers.

The results for an ideal observer are plotted using unfilled symbols. First, the standard deviation of the "channel noise" was set so that the model thresholds for $N = 4$, Fixed condition, were approximately the same as the averaged psychophysical data.[4] Then, the remaining thresholds were determined using computer simulations and the methods described above. Comparing the predicted and obtained effects of amplitude perturbation, it is apparent that the model underestimates the effect of amplitude uncertainty. This parallels the results obtained when uncertainty is introduced by randomizing the frequencies of the component tones. The magnitude of the model's error appears to be larger for frequency than amplitude uncertainty. When the model predicts a threshold of about 3.5 dB, the obtained thresholds are approximately 7.5 and 5.5 dB when frequency and amplitude randomization are applied, respectively.

It is interesting to note that nearly identical results are achieved if one forgoes auditory filtering and uses the ideal observer channel model (Durlach et al. 1986) prediction. Denoting thresholds as Δ, one form of the channel model is $\Delta \propto \sigma_C / \sqrt{N}$, where Δ_C is the standard deviation of the "channel noise" and N is the number of (independent) components. When amplitude randomization is present, the variance associated with the "channel noise" reflects both the impact of the perturbation applied to the stimulus amplitudes and the encoding noise in each channel. If these perturbations are independent and normally distributed, and referring to these two noise sources in terms of the standard deviations σ_S and σ_E for the stimulus (perturbation) and encoding noises respectively, the prediction becomes

$$\Delta \propto \frac{\sqrt{(\sigma_E^2 + \sigma_S^2)}}{\sqrt{N}}$$

If σ_E is estimated using Fixed thresholds, the predicted effects of uncertainty for this simpler model are approximately as shown in Figure 3. Note that for this simpler model, increasing stimulus variability (increasing σ_S) has the largest impact on threshold when the encoding noise is small.

SUMMARY AND DISCUSSION

To summarize, the quasi-ideal observer model fails to account for human data when the task is to detect changes in spectral shape and the frequencies of the component tones are randomly chosen. Depending

[4] Lentz and Richards (1998) noted that for the averaged data the ideal model predicts a larger effect of *N* than obtained. The standard deviation of the channel noise was set to 4.3 dB, compared with 4 dB in the primary experiment.

on one's confidence in the *linear* quasi-optimal model examined, this result may be taken to indicate that uncertainty effects in profile analysis data cannot be accounted for in terms of the stimulus properties as represented at the periphery. Despite substantial individual differences, it is apparent that the quasi-ideal observer *underpredicts* the impact that randomizing the amplitudes or frequencies has on thresholds. This failure is similar to the "informational" masking measured for the detection of a tone added to a multitone masker, where an energy-model observer would monitor the output of a single auditory filter near the signal frequency and thus is little affected by changes in the masker frequency composition.

How might the quasi-ideal model be altered to provide an increased effect of randomization in profile analysis experiments? Changes in the quasi-ideal model such as filter bandwidth, number of filters, channel noise, etc., had little or no impact on the predicted relative effect of uncertainty on thresholds. If one imagines that the effect of increases in uncertainty is to systematically increase a "decision" noise, the likely effect would be to increase the effects of uncertainty as well as to reduce the slope relating thresholds to N (cf. Lentz and Richards 1997). For amplitude randomization, Kidd et al. (1991) obtained such an interaction whereas Lentz and Richards (1998) did not. In the present experiment, the interaction between the degree of randomization and N did not approach significance.

If one rejects a primary assumption of the model used here, that discrimination relies on long-term context rather than short-term trace comparisons, larger effects of uncertainty might be obtained. The primary argument against short-term trace comparisons is that Kidd et al. (1988) found no effects of interstimulus interval on thresholds measured using stimuli with amplitudes that randomly varied on a presentation-by-presentation basis. In contrast, a detrimental effect of interstimulus interval was obtained when the stimuli had amplitudes that varied randomly across trials rather than across presentations (Kidd et al. 1988). The argument (Kidd et al. 1988; see also Durlach and Braida 1969) is that increases in the interstimulus interval should lead to increases in variance of short-term representations due to decay of memory. Thus, the absence of an effect of interstimulus interval measured when presentation-by-presentation uncertainty is applied points to a dependence on a long-term representation in making decisions regarding changes in spectral shape. If, however, the representations have large variances at the outset, as holds when there is presentation-by-presentation uncertainty, increases in variance associated with trace memory loss may not be measurable. The challenge in exploring such a model is to quantify the across-interval comparisons observers might incorporate in making a decision in the face of frequency and/or amplitude uncertainty. Alternatively, one might imagine that the long-term standard is not fixed but relies relatively heavily on recently heard stimulus samples.

Increases in the effects of uncertainty might be produced by vastly reducing the number and/or spacing of auditory filters whose outputs the observer is assumed to incorporate into their decisions. Oh and Lutfi (1998; see also Lutfi, 1993) modeled the combination of informational and energetic masking for the detection of a 1000-Hz tone added to multitone maskers by restricting the number of filters the observer was assumed to integrate across (sum of power) and varying the bandwidth across which those filters might reside. Two auditory filters situated in the 100–2500 Hz frequency range fit their averaged data set well. Moreover, by varying these two free parameters, number of filters and frequency range, individual differences were well described. It remains to be seen whether a similar scheme will be successful in capturing profile analysis data. For example, in Figure 3 it is apparent that thresholds fall with number of components, a result that seems unlikely to be captured using a very sparse number of auditory filters.

The quasi-ideal model considered here also fails in that an interaction between discrimination type and N is predicted, a result not supported by the individual data. Given the magnitude of the individual differences, it may be that a single model will not ultimately account for this type of data. Rather, subject-specific parameters associated with strategies, etc., may have to be invoked.

ACKNOWLEDGMENTS

This work was supported by grant RO1 DC 02012 from the National Institutes of Health. Dr. Gerald D. Kidd Jr., Christine R. Mason, and two reviewers generously provided very helpful suggestions on an early draft of this manuscript. Dr. Brian C.J. Moore provided additional comments that greatly improved the manuscript.

REFERENCES

BERG BG, GREEN DM. Spectral weights in profile listening. J. Acoust. Soc. Am. 88:758–766, 1990.

DUDA RO, HART PE. Pattern Classification and Scene Analysis. John Wiley and Sons, New York, 1973.

DURLACH NL, BRAIDA LD. Intensity perception I: Preliminary theory of intensity resolution. J. Acoust. Soc. Am. 46:372–383, 1969.

FANTINI DA, MOORE BCJ. Profile analysis and comodulation detection differences using narrow bands of noise and their relation to comodulation masking release. J. Acoust. Soc. Am. 95:2180–2191, 1994.

GLASBERG BR, MOORE BCJ. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47:103–138, 1990.

GOCKEL H, COLONIUS H. Auditory profile analysis: Is there perceptual constancy for spectral shape for stimuli roved in frequency? J. Acoust. Soc. Am. 102:2311–2315, 1997.

GREEN DM. Auditory Profile Analysis: Auditory Intensity Discrimination. Oxford University Press, New York, 1988.

GREEN DM. The number of components in profile analysis tasks. J. Acoust. Soc. Am. 91:1616–1623, 1992.

HILL NI, BAILEY PJ. Across-channel intensity discrimination in the presence of an interferer. J. Acoust. Soc. Am. 104:1008–1012, 1998.

KIDD G, MASON CR, GREEN DM. Auditory profile analysis of irregular sound spectra. J. Acoust. Soc. Am. 79:1045–1053, 1986.

KIDD G JR, MASON CR, HANNA TE, Evidence for sensory-trace comparisons in spectral shape discrimination. J. Acoust. Soc. Am. 84:144–149, 1988.

KIDD G, MASON CR, UCHANSKI RM, BRANTLEY MA, SHAH P. Evaluation of simple models of auditory profile analysis using random reference spectra. J. Acoust. Soc. Am. 90:1340–1354, 1991.

LENTZ JJ, RICHARDS VM. Sensitivity to changes in overall level and spectral shape: An evaluation of the channel model. J. Acoust. Soc. Am. 101:3625–3634, 1997.

LENTZ JJ, RICHARDS VM. The effects of amplitude perturbation and increasing number of components in profile analysis. J. Acoust. Soc. Am. 102:535–541, 1998.

LENTZ JJ, RICHARDS VM, MATIASEK MR. Different auditory filter bandwidth estimates based on profile analysis, notched noise, and hybrid tasks. J. Acoust. Soc. Am. 106:2779–2792, 1999.

LEVITT H. Transformed up–down methods in psychoacoustics. J. Acoust. Soc. Am. 49:167–177, 1971.

LUTFI RA. A model of auditory pattern analysis based on component-relative-entropy. J. Acoust. Soc. Am. 94:748–758, 1993.

The Math Works, Inc. Matlab: The Language of Technical Computing. The Math Works, Inc. Natick, MA, 1996.

NEFF DL, CALLAGHAN BP. Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. J. Acoust. Soc. Am. 83:1833–1838, 1986.

NEFF DL, DETHLEFS DL. Individual differences in simultaneous masking with random-frequency, multicomponent maskers. J. Acoust. Soc. Am. 98:125–134, 1995.

NEFF DL, GREEN DM. Masking produced by spectral uncertainty with multicomponent maskers. Percept. and Psychophys. 41:408–415, 1987.

NEFF DL, DETHLEFS TM, JESTEADT W. Informational masking for multicomponent maskers with spectral gaps. J. Acoust. Soc. Am. 94:3112–3126, 1993.

OH EL, LUTFI RA. Nonmonoticity of informational masking. J. Acoust. Soc. Am. 104:3489–3499, 1998.

PATTERSON RD, MOORE BCJ. Auditory filters and excitation patterns as representations of frequency resolution. In: Moore BCJ (ed) Frequency Selectivity in Hearing. Academic Press, New York, 1986, pp. 123–177.

RICHARDS VM, ONSAN ZA, GREEN DM. Auditory profile analysis: Potential pitch cues. Hear. Res. 39:27–36, 1989.

SPIEGEL MF, PICARDI MC, GREEN DM. Signal and masker uncertainty in intensity discrimination. J. Acoust. Soc. Am. 70:1015–1019, 1981.

WRIGHT BA, SABERI K. Strategies used to detect auditory signals in small sets of random maskers. J. Acoust. Soc. Am. 105:1765–1775, 1999.