Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers

Nathaniel I. Durlach,[a)] Christine R. Mason, Frederick J. Gallun,
Barbara Shinn-Cunningham, H. Steven Colburn, and Gerald Kidd, Jr.
Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 18 October 2004; revised 19 July 2005; accepted 20 July 2005)

Sensitivity d' and response bias β were measured as a function of target level for the detection of a 1000-Hz tone in multitone maskers using a one interval, two-alternative forced-choice (1I-2AFC) paradigm. Ten such maskers, each with eight randomly selected components in the region 200–5000 Hz, with 800–1250 Hz excluded to form a protected zone, were presented under two conditions: the fixed condition, in which the same eight-component masker is used throughout an experimental run, and the random condition, in which an eight-component masker is chosen randomly trial-to-trial from the given set of ten such maskers. Differences between the results obtained with these two conditions help characterize the listener's susceptibility to informational masking (IM). The d' results show great intersubject variability, but can be reasonably well fit by simple energy-detector models in which internal noise and filter bandwidth are used as fitting parameters. In contrast, the β results are not well fit by these models. In addition to presentation of new data and its relation to energy-detector models, this paper provides comments on a variety of issues, problems, and research needs in the IM area. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2032748]

PACS number(s): 43.66.Dc, 43.66.Lj, 43.66.Ba [RAL] Pages: 2482–2497

I. INTRODUCTION AND GENERAL BACKGROUND

Although there is still no consensus on how best to define informational masking (IM), it is generally agreed that it is central in origin; related to uncertainty, confusion, distraction, and/or misdirected attention; and clearly distinguished from masking that results from direct interaction between target and masker in the auditory periphery ("energetic" masking). Recent comments on definitional issues are available in Durlach *et al.* (2003a) and Watson (2005).

Substantial work has been done on both sequential IM and simultaneous IM. Background on the former is available in publications by Watson and his colleagues (e.g., Watson, 1987; Watson and Kelly, 1981; and Watson, 2005). Within the domain of simultaneous IM, work has been done on both speech and simplified nonspeech signals. Background on the work using speech signals (and on transformed versions of speech signals) is available in a number of publications (e.g., Freyman *et al.*, 1999, 2001; Brungart *et al.*, 2001; Arbogast *et al.*, 2002; and Kidd *et al.*, 2005).

This article is focused on the effects of randomizing the spectrum of the masker in situations where a tonal target is being masked by a simultaneously presented multitone complex (e.g., Neff and Green, 1987; Neff and Dethlefs, 1995; Oh and Lutfi, 1998; Wright and Saberi, 1999; Richards *et al.*, 2002).[1] Previous research in this area has shown that large amounts of masking occur (e.g., 40 dB) even when the masker is constrained to have no components in the vicinity of the target (e.g., Neff and Green, 1987; Neff *et al.*, 1993; Kidd *et al.*, 1994); that the slopes of the psychometric functions (in the few cases measured) are relatively shallow (e.g., Neff and Callaghan, 1987; Lutfi *et al.*, 2003a; Kidd *et al.*, 2003); and that the amount of masking depends strongly on the listener tested (e.g., Neff and Dethlefs, 1995; Durlach *et al.*, 2003b; Lutfi *et al.*, 2003b). In addition, it has been shown that the amount of masking varies nonmonotonically as a function of the density of the tonal maskers on the frequency axis (e.g., Neff and Green, 1987; Oh and Lutfi, 1998); and that when two-interval paradigms are used, the effect of randomizing the spectrum of the masker is greater when the changes occur between intervals rather than only between trials (e.g., Neff and Callaghan, 1988; Neff and Dethlefs, 1995; Wright and Saberi, 1999; Tang and Richards, 2003; Richards and Neff, 2004). Finally, it has been shown that a substantial decrease in IM can be achieved by decreasing the similarity between target and masker (e.g., Kidd *et al.*, 1994, 2002; Neff, 1995; Durlach *et al.*, 2003b) or by employing trial-by-trial masker cuing (e.g., Richards and Neff, 2004; Richards *et al.*, 2004).

Among the ideas that have been put forward in connection with the empirical findings on IM are the following. First, making use of the standard decision-theory framework, Lutfi and his associates have made detection-performance predictions under the assumption that the decision variable X consists of a weighted sum of the power outputs of various independent nonoverlapping auditory filters (e.g., Lutfi, 1993; Oh and Lutfi, 1998). In particular, they have predicted with considerable accuracy the nonmonotonic dependence of threshold level on the number of frequency components used in the maskers. Apart from choosing the filter shapes appropriately, a major challenge within this structure is to develop appropriate rules for determining the weighting functions

[a)]Electronic mail: durlach@mit.edu

and how these functions depend on the psychoacoustic task, the stimulus, and the listener (see also Richards *et al.*, 2002). Among other things, these weighting functions provide a means for capturing certain aspects of attentional phenomena. For example, a weighting function that includes substantial weights for a wide range of auditory filter outputs corresponds to the listener attending to a wide range of frequencies in the stimulus, i.e., employing a wide "attentional band" (e.g., Green, 1961; Oh and Lutfi, 1998).

Some ideas that have been put forward by Allen and Wightman (1994, 1995) and Lutfi *et al.* (2003a) concern the shape and interpretation of psychometric functions. For example, focusing on percent correct (PC) (as opposed to d'), these authors have stressed the importance of recognizing and taking account of the difference λ between ideal asymptotic performance (PC=100) as target-to-masker ratio becomes very large and observed asymptotic performance. The basic notion here is that the value of λ (referred to as the "lapse" parameter) is generally greater than zero, and that this parameter is important with respect both to the fitting of psychometric-function data and to the development of adequate theory.[2] Also of interest in this general domain is consideration of how the shape of the psychometric function is influenced by the density of the maskers (Lutfi *et al.*, 2003a).

Finally, an additional set of ideas relevant to interpretation of the empirical findings concerns the analysis of a given amount of masking into informational and energetic components (or, alternately, how the informational and energetic masking effects of a given masker should be combined). Material relevant to this issue is available in Lutfi (1990) and Neff and Jesteadt (1996), as well as in Kidd *et al.* (2005).

The research discussed in this paper is concerned with the detection of a tonal target in simultaneous multitone maskers with spectra that have relatively little energy in the vicinity of the target (i.e., in the "protected zone"). In particular, these experiments have been designed to study the effects of randomizing the masker spectrum (over a specified finite set of maskers) on how detection performance varies with target energy level. Among the characteristics of this research that distinguish it from most previous research on simultaneous IM for nonspeech stimuli are the following. First, it is concerned with psychometric functions, not only thresholds. Second, it compares psychometric functions obtained with a random masker to the psychometric functions obtained with the fixed constituents of this random masker (as well as to derived psychometric functions obtained by sorting the results obtained with the random masker and by pooling the results obtained with the fixed maskers). Third, apart from the uncertainty associated with the presence or absence of the target, this research restricts the stimulus uncertainty to the spectrum of the masker. It does not confound the effects of this uncertainty with uncertainty in the masker level (the masker level is held fixed rather than roved) or uncertainty in the target level (rather than use adaptive procedures or the method of constant stimuli, the target level is held fixed throughout each experimental run). Fourth, this research makes use of a one-interval paradigm which, aside from maximizing the number of trials per unit time that can be collected, enables one to examine the dependence of listener response bias on masker spectrum. Fifth, the empirical results are analyzed in terms of simple energy-detector models in which performance is characterized by the sensitivity index d' and the response bias β, and the decision variable is identified with the energy at the output of a linear filter centered on the target frequency. Sixth, rather than testing a large group of listeners and examining the performance statistics for this group, our study examines a few listeners in depth (using the "case-study" approach). Clearly, both approaches are needed to gain full understanding of IM and the large individual differences associated with IM.

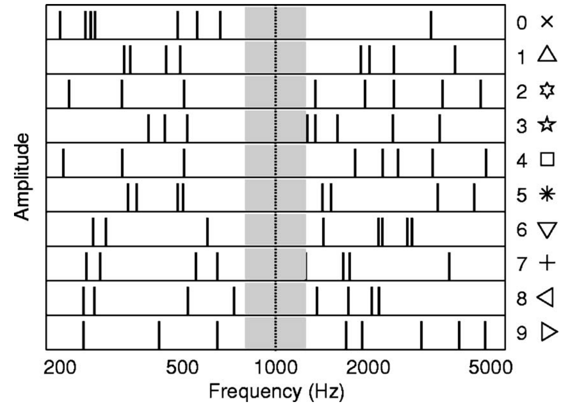

FIG. 1. Line spectra of the ten maskers F_i, $0 \leq i \leq 9$, together with identifying symbols. The shaded region shows the protected zone.

II. DETAILS OF THE EXPERIMENTAL DESIGN

Performance was measured for the detection of a 1000-Hz tone in multitone maskers consisting of eight tonal components. The level of each masker component was 60 dB SPL and the level of the overall masker was 69 dB SPL. The duration of the target and masker, which were gated on and off together, was 300 ms (with 20-ms cosine-squared ramps).

The experiment made use of ten different eight-tone maskers, $F_i, 0 \leq i \leq 9$, each of which was constructed by randomly selecting eight frequencies (with a random phase for each frequency) from the range 200–5000 Hz (on a logarithmic scale) with the subregion 800–1250 Hz around the 1000-Hz target omitted (to form a protected zone). In addition, the components were required to be at least 3.3 Hz (i.e., 1/300 ms) apart. The eight frequencies that comprised each of the ten maskers F_i are shown in Fig. 1. These same ten "frozen" masker samples were then fixed throughout the experiment. All stimuli were digitally generated in the time domain (addition of sine waves) with a sampling frequency of 20 000 Hz. The target and masker were played (using Tucker-Davis Technologies hardware) from separate D/A channels, low-pass filtered at 7500 Hz, attenuated, then summed and passed through a headphone buffer, the output of which was connected through the booth to the headphones. The desired levels were achieved by attenuating targets and maskers appropriately (and separately before summing) utilizing the reference levels obtained by measuring the sound-pressure levels for a known voltage in the TDH 50 headphones.

The experiment employed a one-interval, two-alternative, forced-choice (yes–no) procedure with a 0.5 *a priori* probability of target present and trial-by-trial correct-answer feedback. On each trial, the 300-ms observation interval was followed by a response period of at least 500 ms (additional time beyond 500 ms being available if the listener failed to enter a response during the 500-ms interval). The stimuli were presented monaurally over earphones (to the right ear) in a sound-treated (double-walled IAC) room, and responses were entered using a keyboard.

In the fixed condition, one of the ten eight-tone maskers F_i was held fixed from trial-to-trial. In the random condition, denoted R, the eight-tone masker was varied randomly from trial-to-trial (with uniform probability) over the set of ten maskers F_i specified in Fig. 1. In all cases, detection performance was measured at a variety of signal levels and, for each level, specified in terms of the sensitivity index d' and response bias β. Different sets of signal levels were chosen for the different F_i cases (as well as for the R case) in order to focus on the portions of the d'-vs-level curves that showed significant changes in d' with level.[3] Overall, we obtained 22 empirical functions of level (ten F_i functions and one R function for each of d' and β) for each listener.

The five listeners (three males, two females) tested in these experiments (L1–L5), all of whom were college students with normal hearing (thresholds of 15-dB HL or better at octave frequencies from 250 to 8000 Hz in both ears), had participated in previous IM experiments in this laboratory. For any given condition, data were gathered in 50-trial runs with signal level varied only between runs (decreasing in successive runs). The various F_i cases were tested in random order (different for each listener) and blocks of F_i runs were alternated with blocks of R runs. Also, after the whole sequence of tests was completed, it was repeated to increase the reliability of the results. Each point on the data graphs for the ten F_i functions (for each listener) is based on 100 trials, and each data point for the single R function (for each listener) is based on 1000 trials. Because of the way in which the various test blocks were ordered (as well as the relevant previous experience of all the listeners in IM), we believe that the differences among the various F_i functions and between these functions and the R functions were not influenced strongly by learning or fatigue factors.

In addition to the empirically determined F_i and R functions, the data were processed to obtain derived functions R_p and $F_{s,i}$, $0 \leq i \leq 9$ (for both d' and β). The derived function R_p was obtained by *pooling* the 2×2 stimulus–response matrices obtained in the F_i runs over all these runs, and the derived functions $F_{s,i}$, $0 \leq i \leq 9$ were obtained by *sorting* the matrices obtained in the R runs according to which F_i was presented, where each of the pooling and sorting processes was conducted once for each target level (so that the derived functions, like the empirical ones, are functions of target level). By definition, the pooling and sorting processes are inversely related. In other words, sorting the results obtained for R_p leads back to the results for F_i, $0 \leq i \leq 9$, from which R_p was derived, and pooling the results obtained for $F_{s,i}$, $0 \leq i \leq 9$, leads back to the results for R from which $F_{s,i}$, $0 \leq i \leq 9$ was derived.

In terms of these four types of functions of target level, the primary issues of interest, i.e., the IM issues, concern the way in which the function R differs from the function R_p and the function vector $(F_0,\ldots,F_9)$ differs from the function vector $(F_{s,0},\ldots,F_{s,9})$. The two equalities

$$\begin{aligned} R_p &= R, \\ F_{s,i} &= F_i, \quad 0 \leq i \leq 9, \end{aligned} \tag{1}$$

hold if and only if the listener's detection behavior for the target in the presence of the masker F_i is independent of whether that behavior occurs in an F_i run with that masker or on trials of an R run in which that masker occurs. In other words, these two equalities hold if and only if there is no IM.[4]

For all of the functions $R, R_p, F_i, F_{s,i}$, $0 \leq i \leq 9$, the quantities d' and β were obtained from the appropriate 2×2 matrix by using the entries to estimate the probabilities of detection (P_D) and false alarm (P_{FA}), converting to normal deviates $z(P_D)$ and $z(P_{\mathrm{FA}})$, and then computing $d' = z(P_D) - z(P_{\mathrm{FA}})$ and $\beta = -[z(P_D) + z(P_{\mathrm{FA}})]/2$. Whereas for the F_i case the target levels at which d' and β were measured depended on i, for the $F_{s,i}$ case these levels were independent of i (because all the d' and β results for the $F_{s,i}$ case were obtained by sorting the R data and thus made use of the same levels as those used in the R tests).

III. EXPERIMENTAL RESULTS ON SENSITIVITY *d′*

The dependence of d' on target level for both the empirical functions F_i and R and the derived functions $F_{s,i}$ (obtained by sorting the R data) and R_p (obtained by pooling the F_i data) is shown in Fig. 2. We note the following features of these psychometric functions:

(1) For each listener, the F_i cases have d' functions (within the measured range) that are roughly linear in the given coordinates, share a roughly common slope, and cover a lateral range (e.g., at the $d'=2$ level) of approximately 15−30 dB. Although not evident in the figure, all these functions must converge to $d'=0$ as the target level goes to zero (i.e., to $-\infty$ in dB SPL) and must plateau at some high level of d' as a consequence of occasional errors having nothing to do with hearing, e.g., caused by lapses in attention.[2] (In our computation of the R_p curves shown in Fig. 2, the d' plateau was set at $d'=4.7$.)

(2) Compared to the d' functions for the F_i cases, the d' functions for the R case tend to have a shallower slope and to be laterally displaced to the right. The tendency of these d' functions to converge with those for the F_i cases at the lower signal levels occurs of necessity because of the existence of the $d'=0$ (chance) asymptote for all the functions.

(3) The R_p curve tends to deviate from the R curve by being steeper and displaced to the left. Correspondingly, the $F_{s,i}$ curves tend to deviate from the F_i curves by being shallower and displaced to the right. The extent to which

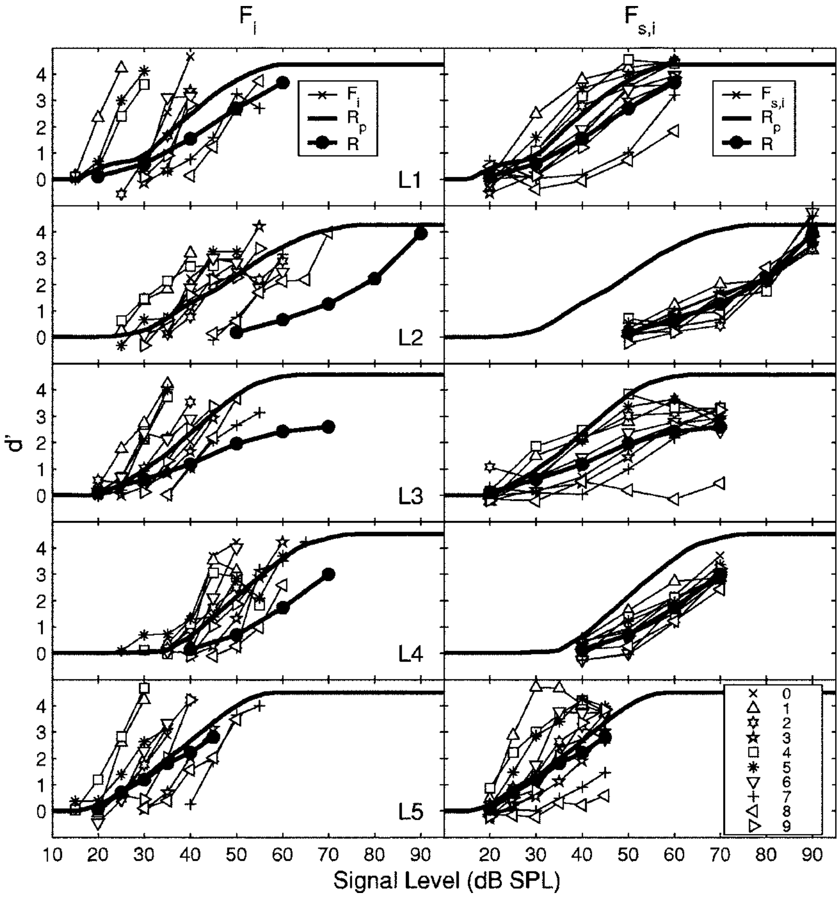

FIG. 2. Results on sensitivity d' as a function of target level for each of the cases F_i, R, $F_{s,i}$, and R_p. The left column gives the empirical functions for the individual maskers ($F_i, 0 \leqslant i \leqslant 9$). The right column gives the derived functions for the individual maskers ($F_{s,i}, 0 \leqslant i \leqslant 9$, obtained by sorting the R data). The empirical R function is shown in both the left and right columns. Also shown in both columns is the derived function R_p (obtained by pooling the F_i data). The correspondence between masker symbol and masker for $i=0$ to 9 is shown in the lower right corner. For details concerning the choice of target levels tested, see the text and footnotes 3 and 5.

these deviations are greater than zero is a measure of IM. That the slope of the R_p curve is shallower than the mean slope of the F_i curves (referred to as the "reduced-slope artifact") is a consequence of the pooling process used to construct the R_p curve combined with the decrease in slope of the F_i curves at small and large values of d'.

(4) The results differ substantially among subjects. For example, L5 exhibits F_i curves and $F_{s,i}$ curves that are relatively steep, quite far to the left, and strongly overlapping, as well as an R curve and R_p curve that are relatively steep, far to the left, and roughly coincident. In contrast, L2 is distinguished by a very large lateral shift to the right of the R curve and the $F_{s,i}$ curves, a very large (almost 30 dB) difference between the R curve and the R_p curve, and relatively shallow slopes for the F_i curves. Listener L1 appears somewhat similar to L5 except that the divergence between the R and R_p curves [and between the F_i and $F_{s,i}$ curves] is greater for L1. These divergences are even greater for L3 and L4 (although not as great as for L2). Note also the differences between L2 and L4 on the one hand and L1, L3, and L5 on the other hand with respect to the relatively tight bunching of the $F_{s,i}$ curves for L2 and L4 (relative both to the F_i curves of L2 and L4 and to the $F_{s,i}$ curves of L1, L3, and L5).

(5) The $F_{s,i}$ results for L3 show a monstrous outlier—a psychometric function that remains close to $d' \approx 0$ over the whole range tested! The psychometric function for this same masker (masker #8) is also the slowest rising function of all those measured in the $F_{s,i}$ data for L1 and L5. As shown in Fig. 1, masker #8 is distinguished by containing frequency components relatively close to the target frequency both above and below it. However, this factor by itself does not explain the strange results for masker #8 in the $F_{s,i}$ case for L3. For this listener, we do not understand why the function is so flat, why the difference in flatness for the F_i case and the $F_{s,i}$ case is so much larger for masker #8 than for all the other maskers, and why the results for masker #8 are so different than the results for masker #7, which also has components close to the target on both sides. (It should be noted, however, that this similarity between maskers #7 and #8 is reflected in the results for listeners L1 and L5.) Finally, it should be noted that based on L2's $F_{s,i}$ data for masker #2 (and, to a lesser extent, for a few other maskers as well), which are basically flat in the range 50–70 dB SPL and only begin to rise at 70 dB SPL, one might expect masker #8 to begin its rise for L3 just at the level at which the data terminate (i.e., 70 dB SPL).

(6) The tendency for the results on the functions F_i and $F_{s,i}$ to deviate from linearity and to evidence reduced slopes at both high and low values of d' (a tendency that must occur for sufficiently extreme values of d') varies considerably with the choice of masker and listener. These deviations are greater for the $F_{s,i}$ results than the F_i results and increase with the lateral spread of the $F_{s,i}$ results.[5]

In general, it is clear from the results in Fig. 2 that the amount of IM varies strongly among the five listeners.

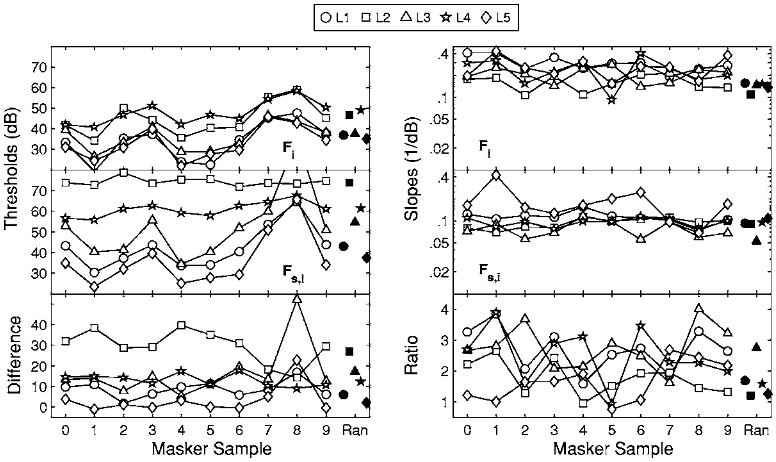

FIG. 3. Thresholds and slopes from straight-line fits to all psychometric functions for all listeners. The top two panels show the thresholds and slopes for the F_i case and the middle two panels for the $F_{s,i}$ case. The bottom two panels show the relation of the F_i results to the $F_{s,i}$ results (differences for thresholds and ratios for slopes). The corresponding results for the R and R_p cases are shown at the right edge of each panel (above abscissa label "Ran") using filled symbols (the R_p results in the top two panels, the R results in the middle panels, and the relation of the two in the bottom panels). The points in all panels are joined by straight-line segments to aid in following the results for individual listeners. For further details, see the text and footnotes 6 and 7.

Whereas L2 exhibits very large IM, listener L5 exhibits hardly any IM. Listeners L1, L3, and L4 appear to fall between these two extremes, with L1 showing the least IM (not much more than that for L5) and L3 and L4 somewhat more (although not as much as L2). Obviously, any satisfactory theory of IM will have to be capable of handling a wide range of listener susceptibilities to IM. The reduced-slope artifact is evidenced in the results for all listeners by the shallowness of the R_p function relative to the F_i functions.

To facilitate quantitative analysis, straight lines were fit to all the psychometric functions for each masker and each listener.[6] The threshold (defined as the level for which $d' = 2$) and the slope were chosen individually for each line to minimize the rms deviation in d' of the points from this line. The thresholds and slopes obtained with these straight-line fits are shown in Fig. 3 for each masker, each condition, and each listener (open symbols for the F_i and $F_{s,i}$ results and filled symbols for the R and R_p results). The extreme behavior associated with masker #8 in the $F_{s,i}$ case is clearly evident in the threshold data but not the slope data (because in the fitting process asymptotic values were ignored and the emphasis was on the points at which d' showed an orderly change with target level[6]). Also shown in this figure are the relationships of the F_i results to the $F_{s,i}$ results and the relationships of the R results to the R_p results (differences for thresholds and ratios for slopes). The strength of the IM in this experiment, i.e., the extent to which the thresholds and slopes of the psychometric functions for the various maskers depend on the context in which the masker was presented (alone in the fixed case or mixed together with the other maskers in the random case) is clearly evident in these relationships. If there were no IM, all the data in the bottom left panel of Fig. 3 would be at 0 dB and all the data in the bottom right panel would be at unity. For each of the five listeners, the means and standard deviations (across maskers) for each of the cases F_i and $F_{s,i}$, as well as the results for the R case, are shown in Table I. An indication of the overall quality of the straight-line fits used to obtain the results shown in Fig. 3 and Table I is given by the rms deviations of the points from the straight lines along the y axis, Y_{dev} (d'), and along the x axis, X_{dev} (target level in dB). These rms deviations are shown in Table II. According to these results, the worst fits occur for L1 and L2 in the $F_{s,i}$ case and L2 in the R case.

TABLE I. Thresholds and slopes derived from straight-line fits to the data. For the F_i and $F_{s,i}$ columns, the means and standard deviations are given across the ten masker samples for each listener. The R and R_p columns give the thresholds and slopes for the one relevant function per listener. The last two rows give the average and standard deviation of the means across listeners. Mean slopes are geometric means and standard deviations for slopes are given in percent. In keeping with our treatment of the outlier (masker #8 for the $F_{s,i}$ case), both its threshold and slope were excluded from the mean and standard deviation over maskers for L1, L3, and L5.

		Thresholds (dB SPL)				Slopes (1/dB)			
		F_i	$F_{s,i}$	R	R_p	F_i	$F_{s,i}$	R	R_p
L1	Mean	33.7	40.1	43.0	37.0	0.29	0.12	0.09	0.16
	sd	9.3	7.1			27	14		
L2	Mean	44.7	74.3	73.8	46.8	0.16	0.09	0.09	0.11
	sd	8.0	2.0			29	18		
L3	Mean	35.9	47.5	54.7	37.5	0.20	0.08	0.05	0.15
	sd	6.8	8.6			28	29		
L4	Mean	47.8	61.0	61.3	49.0	0.22	0.10	0.10	0.15
	sd	5.8	3.7			53	15		
L5	Mean	32.9	33.0	37.4	35.2	0.25	0.18	0.11	0.14
	sd	7.9	8.4			39	52		
Average		39.0	51.2	54.0	41.1	0.22	0.11	0.09	0.14
sd		6.8	16.6	14.5	6.3	26	37	32	15

TABLE II. Rms deviations between straight-line fits to the data and the data themselves in both y-(Y_{dev}) and x (X_{dev}) directions. The mean and standard deviation for each listener is across the masker samples for the F_i and $F_{s,i}$ columns (masker #8 was excluded in the $F_{s,i}$ case for L1, L3, and L5). The last two rows gives the average and standard deviation of the means across listeners.

		F_i		$F_{s,i}$		R	
		Y_{dev} (d')	X_{dev} (dB)	Y_{dev} (d')	X_{dev} (dB)	Y_{dev} (d')	X_{dev} (dB)
L1	Mean	0.29	1.0	0.37	3.2	0.17	1.8
	sd	0.22	0.7	0.16	1.6		
L2	Mean	0.27	1.8	0.34	3.6	0.35	3.9
	sd	0.11	0.9	0.19	1.7		
L3	Mean	0.21	1.1	0.25	3.5	0.14	2.7
	sd	0.13	0.8	0.11	2.2		
L4	Mean	0.30	1.5	0.26	2.7	0.19	2.0
	sd	0.23	1.3	0.08	0.8		
L5	Mean	0.20	0.9	0.22	1.2	0.05	0.5
	sd	0.09	0.5	0.11	0.6		
Average		0.25	1.2	0.29	2.9	0.18	2.2
sd		0.04	0.4	0.06	1.0	0.11	1.3

As part of our effort to compare thresholds for the F_i case to thresholds for the $F_{s,i}$ case, scattergrams and correlation coefficients r were examined for $F_{s,i}$ vs F_i across the ten maskers (separately for each listener). With the exception of L2, the results of this analysis (excluding masker #8 for L1, L3, and L5) showed scattergrams with slopes between 0.6 and 1.2 and values of r in the range 0.89 to 0.97. In other words, for subjects L1, L3, L4, and L5, the variation of threshold with masker was roughly the same for F_i and $F_{s,i}$. For L2, as a consequence of the tiny variation in threshold for the $F_{s,i}$ case, the value of r was found to be close to zero ($r=0.13$). Furthermore, correlations of thresholds between listeners over the set of maskers were generally high for all pairs of listeners in the F_i case (0.78 to 0.94, including all ten masker samples) and all pairs of listeners in the $F_{s,i}$ case (excluding masker #8) in which L2 did not appear (0.60 to 0.95). If the pair included L2, the correlation in the $F_{s,i}$ case for the pair was close to zero or negative (−0.47 to 0.20). In other words, the dependence of threshold on masker was similar across all listeners in the F_i case and across listeners L1, L3, L4, and L5 in the $F_{s,i}$ case (with masker #8 excluded). The main slope result to be noted is the general tendency for the $F_{s,i}$ slopes to be smaller than the F_i slopes (particularly for L1, L3, and L4).

To examine the statistical significance of these results, two separate two-factor repeated-measures analyses of variance (ANOVA) were done, one for threshold data and one for slope data. The two factors in each ANOVA were the uncertainty condition (F_i vs $F_{s,i}$) and the masker sample (always excluding masker #8). In the case of thresholds, the uncertainty condition (F_i vs $F_{s,i}$) alone did not quite reach significance [$F(1,4)=7.1$, $p=0.056$], the masker condition (nine samples) was significant [$F(8,32)=15.1$, $p<0.001$], and the interaction of those two factors was not significant [$F(8,32)=1.07$, $p=0.41$]. Clearly, the variance across masker samples and listeners obscured the difference between the F_i and $F_{s,i}$ thresholds in this ANOVA. When five separate paired-comparisons t-tests were done (one for each listener) on the set of F_i and $F_{s,i}$ thresholds, the result was highly significant for all listeners except L5. The results were $t=-7.7$, df=8, $p<0.001$; $t=-11.7$, df=9, $p<0.001$; $t=-9.3$, df=8, $p<0.001$; $t=-13.7$, df=8, $p<0.001$, and $t=-1.8$, df=8, $p=0.12$ for L1 to L5, respectively. The results of the ANOVA on log-slope values revealed a significant effect of uncertainty condition [$F(1,4)=41.7$, $p=0.003$], but no significant effect of masker sample [$F(8,32)=0.86$, $p=0.56$]. This supports the result that the slopes get shallower in going from F_i to $F_{s,i}$, but that the slope does not depend on masker sample.

IV. ANALYSIS OF THE d' DATA IN TERMS OF SIMPLE ENERGY-DETECTOR MODELS

A. Description of the models

The basic framework used in our analysis of the data shown in Fig. 2 is that of a simple energy-detector model. It is assumed that (1) the decision variable X is stimulus energy or power (measured in dB) at the output of a linear filter with transfer function $H(f)$ centered on the target frequency; (2) the probability density functions of X for the masker-alone case (M) and the target-plus-masker case ($T+M$) are Gaussian with means $\mathcal{M}_M=10\log E_M$ and $\mathcal{M}_{T+M}=10\log E_{T+M} \approx 10\log(E_T+E_M)$ and common variance σ^2, where E_T and E_M are the target and masker energies at the output of the filter; and (3) the listener responds "target present" if and only if $X>C$, where C is the listener's decision criterion. As usual, the sensitivity d' is given by

$$d' = \frac{\mathcal{M}_{T+M} - \mathcal{M}_M}{\sigma}. \tag{2}$$

For the given $\mathcal{M}_{T+M}$ and $\mathcal{M}_M$, one has

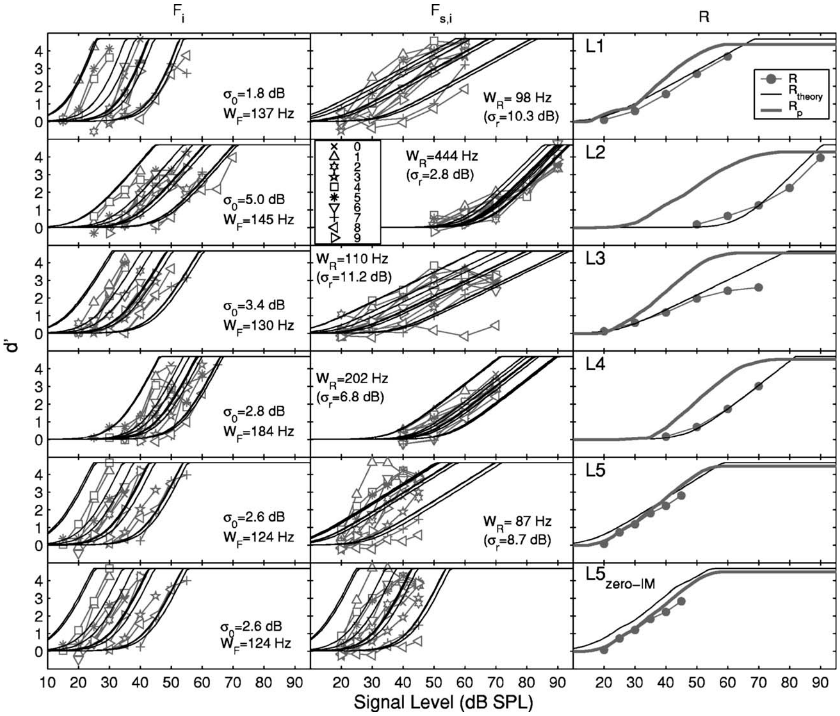

FIG. 4. The results of fitting theory to data. Column 1 is for the F_i case, column 2 for the $F_{s,i}$ case, and column 3 for the R and R_p cases. The F_i data, the R data, and the derived $F_{s,i}$ and R_p results (obtained by sorting the R data and pooling the F_i data, respectively) are identical to those shown in Fig. 2. The upper five rows show the results of applying the simple energy-detector model to each of the five listeners L1–L5. The thin, continuous curves in the upper five rows are determined from Eq. (3) using the parameter values specified in the figure. In the F_i case, both σ_0 and W_F were used as fitting parameters. In the $F_{s,i}$ case, it was required that the value of σ_0 be identical to that used for the F_i case, and thus only W_R was used as a fitting parameter (the value of σ_r being determined by the value of W_R). In the R case, no new fitting was performed (the values of σ_0 and W_R being identical to those obtained from the fitting of the F_i and $F_{s,i}$ results). For further details concerning the fitting of this model to listeners L1–L5, see the text and footnote 7. The bottom row in this figure shows the results of applying the zero-IM model to listener L5. In this model, the predictions for the $F_{s,i}$ case are identical to those for the F_i case and there are only two fitting parameters, σ_0 and W_F. Also, the value of W_R is the same as that of W_F and there is no σ_r term (see the text for details).

$$d' = \frac{10 \log(1 + E_T/E_M)}{\sigma}, \tag{3}$$

a formulation that is consistent with Weber's law for intensity discrimination.

It will be assumed that $H(f)$ is chosen to increase the T/M ratio by focusing on T rather than by nulling out M; that $H(f)$ represents the integrated frequency-selectivity behavior of the whole auditory system (including attentional phenomena), not only that associated with the periphery (i.e., the "critical band"); and that $H(f)$ is independent of the identity of the masker, but may depend on the listener and/or the experimental condition to which the listener is exposed (specifically, whether the test is an F_i test or an R test). It will further be assumed that in the R tests the listener has zero knowledge of which masker is presented on a given trial and thus sets the decision criterion C in a manner that is independent of the masker and designed to distinguish between the broad probability densities on the decision axis that characterize the energy at the output of the filter $H(f)$ over all of the maskers F_i (one density for the case of target absent and one for the case of target present). Although such a simple model (namely, one in which detection is assumed to depend only on energy and, furthermore, only on energy at the output of a single filter) is clearly inadequate to explain all results on IM, we believe it is instructive to examine how it relates to the data reported in this study.

The variance σ^2 will be assumed to consist of two components

$$\sigma^2 = \sigma_0^2 + \sigma_r^2, \tag{4}$$

where σ_0 represents internal noise (arising centrally as well as peripherally) and σ_r represents the increase in the variance associated with randomization of the masker. The term σ_0 will be allowed to vary with the listener, but will be held fixed otherwise. This term appears in all the F_i cases and the R case, and is the only variance term in the F_i cases. The quantity σ_r will be computed as the rms deviation in energy level (in dB) at the output of the filter $H(f)$ resulting from the masker-spectrum randomization (and will be referred to as the "roving-level" component of the variance). Note that the σ_r is completely determined by the filter $H(f)$ and the masker spectra, and thus is not a free parameter. Note also that as the filter $H(f)$ becomes increasingly broad in the R case, both the numerator and denominator of the equation for d' [Eq. (3)] decrease. The numerator decreases because the masker energy that gets through the filter (E_M) increases, and the denominator decreases because the roving-level variance (σ_r^2) decreases. For very broad filters, σ_r approaches zero because all the maskers F_i were constructed to have the same total energy.

B. Application of the model to the data

In applying this model to the data, the filter $H(f)$ was assumed to be a symmetric roex(p,r) filter (Patterson and Moore, 1986) centered at 1000 Hz with equivalent rectangular bandwidth (ERB) W, and in those cases in which σ_r was relevant, σ_r was computed as the rms deviation of the masker energy about its mean at the output of this filter (where the energy was measured in dB). Although no rigorous study was made, based on a few informal probes it is believed that modest modifications of the assumed filter shape would lead to only modest changes in the results presented in this section. The results of fitting this model to the data are shown in the upper five rows of Fig. 4 (the bottom row is considered below).

Consider first the fitting of the F_i functions. Because the masker was fixed in this case, one has σ_r=0 and the predictions are given by Eq. (3) with σ=σ_0 (the internal noise) and

TABLE III. rms deviations between model and data for threshold (T_{dev}) and slope (S_{dev}). The bottom row gives the results for the "zero-IM" model. Masker #8 was excluded in the $F_{s,i}$ case for L1, L3, and L5.

	F_i		$F_{s,i}$		R	
	T_{dev} (dB)	S_{dev} (%)	T_{dev} (dB)	S_{dev} (%)	T_{dev} (dB)	S_{dev} (%)
L1	4.1	21	4.8	23	2.3	2
L2	5.0	24	3.5	71	1.0	75
L3	4.8	22	5.4	26	6.7	59
L4	4.3	38	4.7	36	0.8	36
L5	4.1	29	2.9	71	3.4	0
Average	4.5	27	4.2	44	2.9	31
sd	0.4	5	1.0	17	2.4	29
$L5_{zero\text{-}IM}$	4.1	29	4.7	61	6.0	12

with E_T and E_M given by the target and masker energies at the output of the roex filter. The fitting in this case was achieved by choosing (for each subject) σ_0 and W_F to match the theoretical F_i thresholds and slopes to the empirical F_i thresholds and slopes.[7] The results of this fitting and the values of σ_0 and W_F used to achieve these fits are shown in the left column of Fig. 4. Whereas the theoretical slopes are determined by σ_0, the theoretical thresholds are determined by W_F as well as σ_0. Further computations showed that the mean of the correlation coefficient r between theoretical and empirical thresholds across maskers (over the five listeners) was 0.89 (range 0.81–0.95). The relatively high value of this correlation, combined with the assumptions of the model, necessarily implies that the correlation between the empirical thresholds and the amounts of energy through the assumed filter is relatively high.

Consider next the $F_{s,i}$ case. The fitting here was performed in exactly the same manner as the F_i case except that (1) the parameter σ_0, rather than being treated as a fitting parameter, was chosen to be the same as in the F_i case and (2) the variance term σ^2 now included a σ_r^2 term as well as the σ_0^2 term. The fits in the $F_{s,i}$ case, shown in the middle column of Fig. 4, were then achieved by choosing W_R to best fit the empirical $F_{s,i}$ thresholds and slopes (the slopes in this case are determined by the fixed σ_0 term and by W_R which determines σ_r).[7] Excluding L2, the correlation results (between theoretical and empirical thresholds) showed a mean of 0.85 and a range of 0.72 to 0.94. For L2, whose thresholds (both empirical and theoretical) were roughly independent of the masker, r was essentially zero ($r=-0.16$).

Consider, finally, the results for the random-masker case and the two theoretical curves shown in the right column of Fig. 4. One of the these curves, namely, the curve for R_p, is the same curve as that shown in Fig. 2. The second curve is obtained from Eq. (3) using the same values of σ_0, σ_r, and W_R that were used to generate the theoretical curves for the $F_{s,i}$ functions. In other words, no further fitting was done to match the theoretical R curves to the empirical R curves.

The quality of the fits of the theoretical results to the empirical results can be judged not only by visual inspection of Fig. 4, but also by examining the deviations in thresholds and slopes derived from the straight-line representations of the theory and of the data (denoted T_{dev} and S_{dev}). These deviations are shown in Table III. As one might expect from an examination of Fig. 4, the largest deviations occur with the $F_{s,i}$ slopes for L2 and L5, the R slopes for L2 and L3, and the R threshold for L3.

Beyond the issue of consistency between theoretical and empirical results, of course, is the issue of plausibility for the values of the constants obtained in the fitting process, i.e., the values of σ_0, W_F, and W_R shown in Fig. 4. Overall, these values appear quite reasonable. The values of σ_0 vary over the range 1.8–5 dB, corresponding to a range of 0.5–2.2 for the value of the ratio T/M at which d' is unity; and the values of W_F vary from 124 to 184 Hz, values roughly consistent with those encountered in various independent studies of the auditory filter bandwidth (e.g., Neff *et al.*, 1993). The values of W_R, which also determine in this case the values of σ_r, show relatively modest changes from W_F with the notable exception of L2, who is modeled best by assuming a dramatically wider bandwidth (roughly a factor of 3) for W_R. More surprising, perhaps, are the decreases in W for L1, L3, and L5 in going from W_F to W_R.[8]

For certain listeners, such as L5, there is relatively little vulnerability to IM and the above model may be inappropriate. Thus, we briefly examined an alternate model in which no IM is predicted (the "zero-IM" model). In this model, it is assumed that the listener not only uses the same decision variable in the R tests as in the F_i tests [the energy at the output of the same $H(f)$ filter], but also the same decision criteria. In other words, it is assumed that the listener can identify the masker F_i presented on each trial in an R run (by listening to the stimulus at frequencies away from the target frequency) and set the criterion on that trial in the same manner that it would be set in an F_i run with that masker. With this zero-IM model, not only is $H(f)$ the same for the F_i runs and the R run, thus reducing the number of free parameters in the model, but one has $\sigma=\sigma_0$ (there is no σ_r term), $F_{s,i}=F_i$, and $R_p=R$. Because in our experiment the total energy in the masker is independent of the masker, and only the spectrum of the masker is randomized, potential cues for identifying the masker by examining the stimulus at frequencies away from the target include masker energy as well as spectral shape at these frequencies. It should also be noted that in this zero-IM model the trial-to-trial identification of the masker in an R run is required only to the extent that it facilitates the choice of criterion for that trial, i.e., provides knowledge of E_M on that trial.

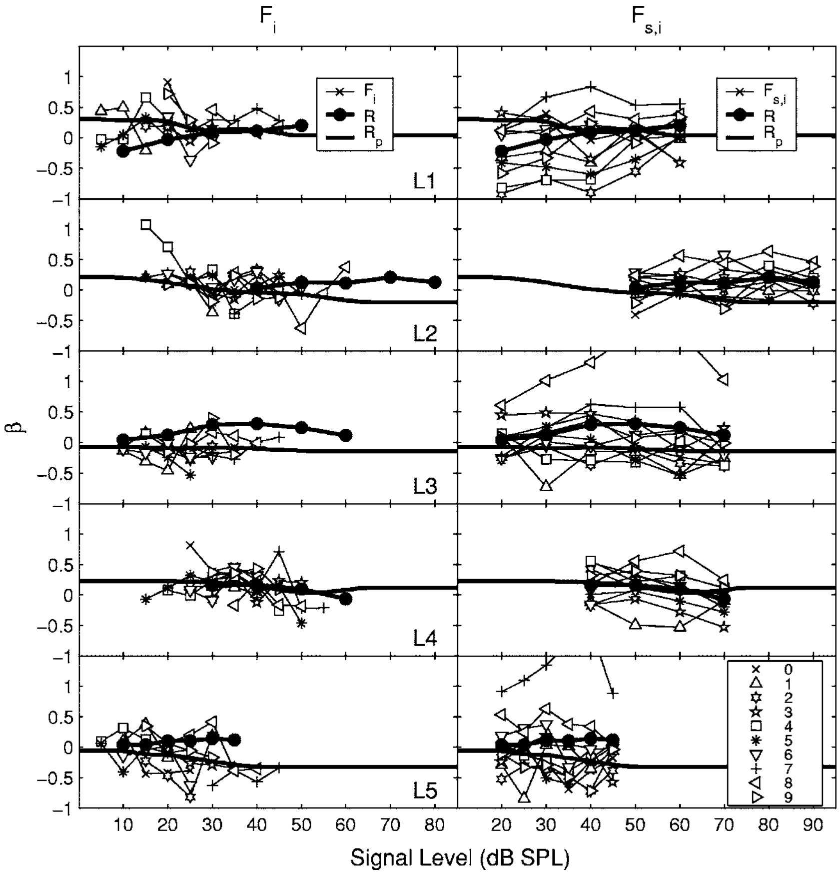

FIG. 5. Data on response bias β corresponding to data on d' shown in Fig. 2.

The results of comparing this zero-IM model to the data for L5 (using the same values of σ_0 and W_F as previously used for the F_i cases) are shown in the bottom row of Fig. 4 along with the fits obtained using the previous model. The theoretical curves for this model are the same as those derived for the previous model for F_i and R_p, but not for $F_{s,i}$ or R (with this zero-IM model, the curves for $F_{s,i}$ are identical to those for F_i). The values of T_{dev} and S_{dev} obtained with this model are included in Table III along with the values obtained with the previous model. According to these results, the change in model causes little change in S_{dev} but increases T_{dev} for both $F_{s,i}$ and R.

Finally, in concluding this section, three points should be noted. First, the results would not have been substantially altered if in the fitting process the data themselves rather than the thresholds and slopes determined from the straight-line fits to the data had been used. Second, the two models considered above reflect behavior extremes. Models of intermediate behaviors (which are likely to be appropriate to a substantial number of listeners) can be constructed by appropriate blending of these two models (e.g., by assuming some, but not perfect, knowledge of E_M on each trial). Third, the discrepancies between the simple energy detector model and the d' data are relatively small compared to the variations in the data across subjects, but relatively large compared to the variations in the data associated with the use of a finite number of trials in the experiments. If one determines the deviations T_{dev} and S_{dev} that arise when the results for one listener are used as a model for another listener, one obtains the following mean results (across all subject pairs): $T_{\mathrm{dev}}=9.3$, 21.4, and 18.2 dB and $S_{\mathrm{dev}}=53\%$, 50%, and 34% for F_i, $F_{s,i}$, and R, respectively. A comparison of either of these sets of numbers with the corresponding numbers in Table III ($T_{\mathrm{dev}}=4.5$, 4.2, and 2.9 dB and $S_{\mathrm{dev}}=27\%$, 44%, and 31%) provides support for the first of the above-mentioned statements, namely, that differences between theory and data are relatively small compared to differences among subjects. The validity of the second of these statements is proved by simulations we have done using 100 trials per data point (roughly equivalent to the number of trials per data point used in our tests). According to the results of these simulations, if the only source of deviation were the finite number of trials per data point, then T_{dev} should lie in the range 1–3 dB for the F_i and R cases and 2–6 dB for the $F_{s,i}$ case, and S_{dev} should lie in the range 11%–36% for all cases. All of these comparisons reflect the emphasis on thresholds relative to slopes in the process used to fit the model to the data.[7]

V. RESULTS ON RESPONSE BIAS

The results on response bias β (obtained according to the procedures outlined at the end of Sec. II) are shown in Fig. 5 in a manner similar to that used to display the d' results in Fig. 2. As with d', the β results for $F_{s,i}$ are obtained from the R results using the sorting process, and the β results for R_p are obtained from the F_i results using the pooling process. In general, a positive value of β corresponds to a bias to respond "no signal" too frequently, i.e., an overly conservative criterion.

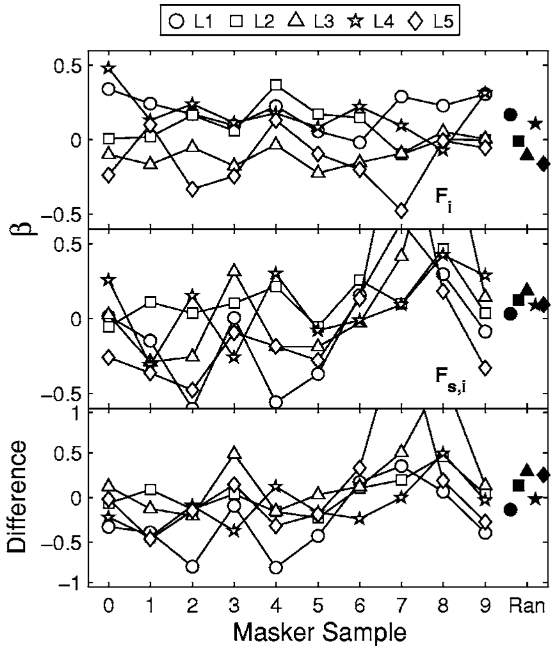

FIG. 6. Mean values of β over level as a function of masker, test condition, and listener. The top panel gives the results for the F_i case, the middle panel for the $F_{s,i}$ case, and the bottom panel for the difference. The filled symbols on the right (abscissa label "Ran") show the corresponding results for the R_p (top panel) and R (middle panel) cases. As in Fig. 3, the points are joined by straight-line segments to aid in following the results for individual listeners.

TABLE IV. Results of horizontal-straight-line fits to β-vs-level data. For each of the F_i and $F_{s,i}$ cases, the entries give the means and standard deviations over the set of maskers of the rms deviation (Y_{dev}) of the data points from the horizontal lines. For the R case, the entries give the single rms deviation of the data points from the single horizontal line. The row on the bottom gives the means and standard deviations of the results in the rest of the table across listeners. (In computing the entries in this table, the two outliers seen in Fig. 5 were omitted.)

		F_i	$F_{s,I}$	R
L1	Mean	0.22	0.24	0.15
	sd	0.12	0.10	
L2	Mean	0.21	0.16	0.06
	sd	0.12	0.06	
L3	Mean	0.14	0.22	0.10
	sd	0.08	0.07	
L4	Mean	0.18	0.17	0.09
	sd	0.10	0.04	
L5	Mean	0.21	0.22	0.04
	sd	0.10	0.07	
Average		0.19	0.20	0.09
sd		0.03	0.04	0.04

Visual inspection of the results shown in Fig. 5 indicate (1) the existence of two "outliers" in the $F_{s,i}$ case (generated by masker #8 for L3 and masker #7 for L5); (2) very weak (if any) dependence of β on signal level; and (3) excluding the two outliers, a general tendency for the values of β to cluster around $\beta=0$. In addition, it appears that (4) there is a slight tendency for β to decrease with level (become less conservative as signal level increases) in the R_p case (for all listeners but L3) and to be lower in the R_p case than in the R case (at least for listeners L2, L3, and L5).

As will be seen below, the most important result obtained from measurement and examination of the β data concerns the correlation between the value of β and the value of the threshold (over the set of maskers) for the $F_{s,i}$ case. With the goal of examining this correlation in mind, and the knowledge that β is at least roughly independent of level (as seen in Fig. 5), our initial processing of the β data consisted of eliminating the level parameter by averaging over it, i.e., of representing the β-vs-level data shown in Fig. 5 by best-fit horizontal straight lines. The results of this processing are shown in Fig. 6 and Table IV. Whereas Fig. 6 shows how β, averaged over level, depends on all maskers, all cases, and all listeners, Table IV provides information on the horizontal-straight-line fits. Even excluding the two outliers in the $F_{s,i}$ case, the results in Fig. 6 show a somewhat greater variation with masker in the $F_{s,i}$ case than in the F_i case. Examination of β scattergrams and correlation coefficients r for F_i vs $F_{s,i}$ (over the maskers) showed no relationship between the β values for these two cases; for all listeners, the value of r was close to zero ($0.06 \leqslant r \leqslant 0.18$).

As expected, the β results for the R_p case (solid symbols in top panel of Fig. 6) for each listener tend to be near zero and look much like an average of the β values shown for each masker sample in the F_i case in the same panel. The range of values across listeners is −0.16 to 0.17. The β values shown for the R case (solid symbols in the middle panel of Fig. 6) appear more similar across listeners, all having slightly positive β values in the range 0.03 to 0.19. This is due in part to the inclusion of all data in the R case (including the outliers that are seen in the $F_{s,i}$ case). The differences between the β values for the R_p case and for the R case [and also for F_i vs $F_{s,i}$] are shown in the bottom panel. As can be seen more readily in Fig. 5, listeners L2, L3, and L5 all have more positive β values in the R case than in the R_p case, whereas L1 shows a slight negative difference and L4 shows virtually no difference. A two-way repeated-measures ANOVA on the β values with uncertainty (F_i vs $F_{s,i}$) and masker sample (excluding maskers #7 and #8) as the factors showed that neither main effect is significant [($F(1,4)=3.24$, $p=0.15$ for F_i vs $F_{s,i}$; $F(7,28)=1.46$, $p=0.22$ for masker sample)].

Consider finally (and most importantly) the relationship of response bias β for a particular masker to the target threshold for that masker. Scattergrams and correlation coefficients r for β vs threshold over the set of maskers for each of the five listeners are shown in Fig. 7. The results for the F_i case (column 1) show no pronounced structure, with the values of r varying over the range −0.60 to 0.41, depending on the listener. In contrast, the results for the $F_{s,i}$ case (column 2) show a tendency to produce correlations that are positive. Not only do L1 and L3 show the relatively high correlations of 0.80 and 0.90 for the $F_{s,i}$ case, but for all listeners the value of r for the $F_{s,i}$ case is more positive than for the F_i case. Specifically, the values of $r(F_{s,i})-r(F_i)$ are given by 0.60, 0.15, 0.49, 0.97, and 0.71 for L1–L5, respectively. Note, moreover, the relationship for the outliers: both maskers #8 and #7 tend to combine a very high threshold with a very high value of β (see the results for both of these maskers in Figs. 3 and 6). Overall, with the exception of

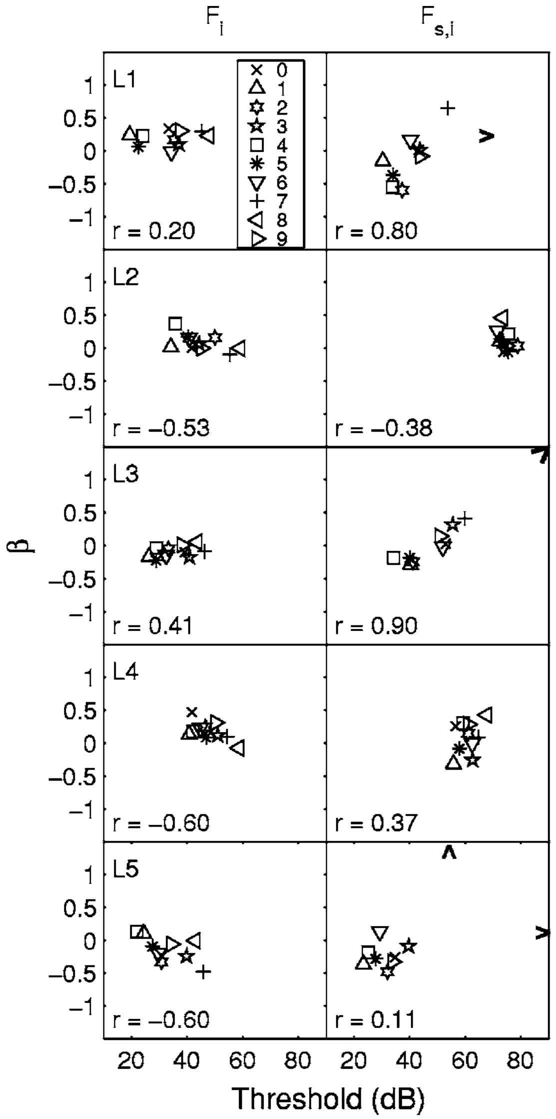

FIG. 7. Scattergrams and correlation coefficients r for response bias β vs threshold (across maskers). The first column gives results for the F_i case and the second column for the $F_{s,i}$ case. Arrowheads indicate the existence of outliers (not included in the computations of r).

listeners L2 and L5, whose d' results have already been shown to be relatively special, it appears that in the $F_{s,i}$ case, β and threshold tend to be positively correlated (with this positive correlation being smaller for the intermediate listener L4 than for L1 and L3).

Turning attention now to theoretical interpretations, we note the following. First, in the F_i case, where the same masker is presented trial after trial, there is no reason to anticipate large systematic deviations of β from zero and strong correlations (either positive or negative) with threshold because of the opportunity provided to each listener (via the trial-by-trial correct-answer feedback) to eliminate any initial bias the listener might have had based on the particular characteristics of the masker. Second, in the $F_{s,i}$ case, we would expect the β results to depend on the type of listener (as discussed previously in connection with the d' results). For the "normal" case, in which the listener (a) maintains a relatively narrow filter and (b) cannot identify the masker on individual trials of an R run, we would expect the listener to choose a single response criterion more or less in the middle of the whole range of energies coming out of the filter, and thereby produce results that would show a significant dependence of β on the energy of the masker coming through the filter and hence on the threshold of the target for the specific masker in question. In contrast, no such dependence would be expected if the listener either (a) opens up the filter to the extent that essentially all maskers produce roughly the same energy at the output of the filter (so that a fixed response criterion will not lead to a substantial dependence of β on the threshold) or (b) identifies the masker on each trial and locates the criterion in a manner that is appropriate for that masker. With these comments and the previous discussion of the data on d' for listeners L1–L5 in mind, the results showing the relation of β to the threshold shown in Fig. 7 look quite reasonable: whereas L1 and L3 show a clear dependence of β on threshold, L2 and L5 do not (with L4 lying somewhere in the middle of these two groups). Note also the extent to which the $F_{s,i}$ results shown in Fig. 7 for L5 support the zero-IM model rather than the original model for this listener.

There is, however, a big problem associated with these β results: according to our initial model, the predicted correlation between β and threshold that is seen in the results for L1 and L3 should be negative, not positive. For this model, the value of β is given by

$$\beta = \frac{C - \left(\dfrac{\mathcal{M}_{T+M} + \mathcal{M}_M}{2} \right)}{\sigma}, \tag{5}$$

where $\mathcal{M}_{T+M}$, $\mathcal{M}_M$, and σ are the same as considered previously in the analysis of the d' data, and the listener's decision criterion C is placed in the middle of the energy distribution as mentioned above. *This implies, however, that as the threshold increases the value of β should decrease* (corresponding to an increased tendency to respond signal present), *not increase* (corresponding to an increased tendency to respond signal absent). More specifically, if this model were correct, a masker that produced a high target threshold (i.e., a masker that produced a high level of energy at the output of the filter) would be associated with a high value of the quantity in parentheses in Eq. (5), which in turn would lead to a negative value of β. Further evidence of this model's inadequacy can be found in the range of values observed for β. If the response criterion were really the same for all maskers in the R case, the range of β would be much greater (because of the large range of masker energies at the output of the filter).

VI. SUMMARY AND DISCUSSION

A. Comparison with previous results

Previous work on psychometric functions in IM has been performed by AW (Allen and Wightman, 1994, 1995), WS (Wright and Saberi, 1999), and LKCW (Lutfi, Kistler, Callahan, and Wightman, 2003a). In principle, it should be useful to compare our results to theirs. In practice, however, meaningful comparisons are exceedingly difficult to make because of the major differences in stimuli, experimental procedures, methods of analysis, and groups of listeners. One such difference is that all the previous studies used a 2I-2AFC paradigm. Also, rather than using fixed constituents of the random masker for the "minimum uncertainty" case, AW and LKCW used broadband Gaussian noise for this case. Furthermore, AW did not use a notch in this noise to create a protected zone, and LKCW provided a target cue on each trial prior to the 2I-2AFC presentation. In addition, with the exception of one portion of the WS study, none of the previous studies held the target level constant during an experimental run. Finally, in the AW study, the analog to our R case was achieved by adding a single random-frequency tonal masker (referred to as a "distractor") to the Gaussian noise masker, and in the WS study the analog involved pairing each masker with only one other masker in the 2I-2AFC paradigm.

Based on our attempts to process the data presented in all the studies in a manner that permitted comparisons among all the results, we found that the mean slope for the minimum uncertainty condition in the WS study was outstandingly low, that the mean threshold for all cases considered by AW were outstandingly high (presumably due to the lack of a protected zone in their noise masker), and that the ranges of thresholds and slopes found in the LKCW study were outstandingly large (presumably due to their large and age-varied listener pool). Most surprising was the high degree of consistency between the mean thresholds and mean slopes obtained in the LKCW study and in our own study. Apparently, to a large extent, the effects of the main differences in how the two studies were conducted were surprisingly small or canceled each other out.

Comparisons of our data on response bias β with results of previous studies are even more problematic, in part because most of the previous IM studies did not report on response bias. In the one IM study that was concerned with bias, the WS study, the methodology for exploring bias was so radically different that comparisons are difficult to perform. Nevertheless, to the best of our understanding, the results in WS imply *a negative* correlation between threshold and bias, consistent with our model but inconsistent with our data. To add to the puzzle, examination of the correlation between response bias β and d' at one signal level (inversely related to threshold) in "frozen noise" tone-in-noise detection experiments (Siegel and Colburn, 1989; Gilkey *et al.*, 1985; Isabelle and Colburn, 1991; Evilsizer *et al.*, 2002; and Davidson and Carney, 2003) leads to the conclusion, like our own experimental work but not our model, that the correlation between bias and threshold is positive. Further information concerning these comparisons of response-bias results, as well as the comparisons of psychometric-function results, can be obtained directly from the authors.

In addition to examining how our data compare to previous data, we have explored the extent to which our data are consistent with and/or can be usefully interpreted with the help of the theoretical model in the LKCW study (related to the CoRE model, as discussed, for example, in Oh and Lutfi, 1998). This model is similar to the simple energy-detector model considered in the present paper in that it assumes that decisions are based on the values of energy variables, but it differs in a variety of other respects. Making use of the two equations in the LKCW paper concerned with the dependence of threshold and slope of a psychometric function on their model parameter n (see page 3280 in that article), we found, as did they when examining their own data, that the value of n needed to match the threshold data could differ dramatically from the value of n needed to match the slope data. In our data, this problem was most pronounced (by a large margin) for listener L2, the listener who was most susceptible (by a large margin) to IM.

B. Summary of our results

Even with a subject pool of only five listeners, a wide variety of listener behaviors in response to randomization of the masker was revealed. As shown in Fig. 2, the changes in both threshold and slope of the psychometric functions caused by randomization of the masker (as well as the thresholds and slopes of the psychometric functions for the fixed cases) varied considerably over the set of listeners.

Despite these variations, to at least a rough approximation all the psychometric functions for all cases and all listeners obtained using the given psychophysical procedures (one-interval paradigm, masker energy held constant, target energy held constant within each run, etc.) can be described by a simple energy-detector model in which the listener's detection decision is assumed to be based solely on the energy at the output of the filter centered on the target signal. The free parameters in the model, whose values were selected individually for each listener to fit that individual's d'-vs-level data (see Fig. 4), are the internal noise variance σ_0^2 (assumed to be independent of the stimulus), the bandwidth W_F (assumed to be applicable for fixed maskers), and the bandwidth W_R (assumed to be applicable for the random masker). Also found for all listeners, the relative effectiveness of the various maskers is roughly independent of whether one examines the F_i results or the $F_{s,i}$ results.

By far the largest differences across listeners in the d' results are those associated with listener L2. Whereas the values of σ_0, W_F, and W_R required to fit the energy-detector model to the d' data for L1, L3, L4, and L5 are all roughly what one would expect on the basis of previous work on energetic masking (taking account of both the reduced-slope artifact and past estimates of internal noise and critical bandwidth), the values of these parameters required to fit the d' data for L2 are quite different. Not only is the value of the internal noise term σ_0 for L2 unusually large, but the value of W_R for L2 is three times larger than the value of W_F for L2 (corresponding to the notion that L2, unlike the other listeners, responds to randomization of the masker spectrum by widening the effective acceptance filter).

Perhaps the most notable results of our study concern the data on response bias β (which were found to be essentially independent of target level). Of most importance is the correlation of response bias β with target threshold (over the maskers) for the $F_{s,i}$ case (see Fig. 7). For listeners L2 and L5, this correlation is relatively small, as would be expected on the basis of our interpretation of their d' data. In contrast, for L1 and L3 (and to a lesser extent L4), a positive correlation exists. Inasmuch as one would expect a negative correlation using the simple energy-detector model and the most obvious criterion-placement behavior by these listeners, i.e., placing a single decision criterion more or less in the middle of the roving-energy-level range at the output of the filter, it was concluded that the simple energy-detector model outlined in this paper was seriously inadequate for these listeners even for the simple detection tasks considered in this paper.

Finally, it is important to note with respect to listeners L1 and L3 (and to a lesser extent L4) that if it were not for the above-mentioned discrepancy between theory and data with respect to the correlation between detection threshold and response bias for these listeners, one could legitimately have concluded that the results of these experiments for these listeners are interpretable entirely within the context of energetic masking (even with the large "protected zone" employed). In other words, the detection data by themselves (i.e., ignoring the bias data) are totally consistent with a model in which the masked thresholds are determined solely by the "leakage" of masker energy into the target critical band (via the skirts of the masker spectra and the critical band). Thus, aside from the large value of W_R (and perhaps σ_0) observed for listener L2 in the d' data, the main results reported in this paper that appear to require ideas beyond those classically associated with energetic masking are those associated with response bias.

C. Comments on future research

In the remainder of this section, we comment briefly on some studies related to our results that we believe are important to pursue in the future.

First, focusing on intersubject differences, we note that even with only five listeners in the subject pool, our empirical results evidence a wide variety of behaviors, ranging from strong IM (L2) to essentially no IM (L5). Extension to a larger group of listeners would not only provide better estimates of the distribution of the observed behavior types, but also possibly lead to the discovery of still further types. For example, we would not be surprised to discover listeners who exhibit weak (or covert) IM in the sense of producing data that satisfy the equation $R=R_p$ but not $F_{s,i}=F_i$ $(0 \leq i \leq 9)$. Inasmuch as different sets of functions for the F_i case can result in the same function for the R_p case, there is no reason to believe that $R=R_p$ necessarily implies $F_{s,i}=F_i$ $(0 \leq i \leq 9)$. In fact, the extent to which L5 in our study should be viewed as exhibiting no IM rather than a small amount of covert IM depends on how seriously one regards the deviations between $F_{s,i}$ and F_i relative to the deviations between R and R_p.

Second, even within the given restricted listener domain of this study, the observed differences in performance and the associated differences in assumed strategies lead to differential predictions to be tested by further experiments with the same (or similar) subjects. For example, for a subject like L2, who appears to have addressed the R task by simply opening up the filter and sensing the overall loudness of the stimulus, one might expect d' to be relatively insensitive to the number of maskers but to increase substantially with training and/or the introduction of target or masker cuing (due to an alteration in strategy). In contrast, for a subject like L5 whose performance (in both d' and β) is roughly consistent with the notion of identifying the masker on each trial and establishing a criterion matched to that masker, one might expect performance to show only limited improvement as a consequence of cuing or training and to deteriorate as the number of maskers increased.

Third, it seems likely that our characterization and understanding of IM could be significantly improved by further analyzing existing data to (a) explore the extent to which previous studies have failed to take appropriate account of the reduced-slope artifact in assessing the existence and magnitude of IM and (b) examine trial-to-trial sequential effects (e.g., the dependence of filter and criterion settings on a given trial on the stimulus, response, and feedback associated with previous trials). Inasmuch as the phenomenon of IM is based on trial-to-trial changes in the stimulus, it seems very likely that considerable further insight into IM can be gained by detailed study of trial-to-trial sequential effects.

Fourth, there is a variety of additional theoretical studies that could play an important role in addressing the failure of the simple energy-detector model considered in this paper to explain the empirical findings on the sign of the correlation between response bias β and threshold (for subjects L1 and L3, to a lesser extent for L4, and for the results on frozen noise mentioned above). A revised energy-detector model that might be able to solve this bias-threshold correlation problem could be constructed by assuming that the listener estimates the identity of the masker on each trial and sets the criterion C according to that estimate, but, unlike listener L5, does this rather poorly. According to this model, not only is the ability to identify the masker somewhat degraded, but the setting of the criterion C suffers from both a tendency to overshoot its optimum value on the decision axis and a random jitter on this axis (criterion noise) that is proportional to the range of masker energy levels at the output of the filter (consistent with the context-coding model of intensity perception (e.g., Durlach and Braida, 1969; Braida *et al.*, 1984). An alternative, more attractive, model addressed to this problem would be one in which the masker energy outside the filter (in the "spectral fringe") plays a role in the formation of the decision variable as well as in the placement of the criterion C. Although such "background" can always play a useful role as a reference and reduce memory load, it could be of special significance in this case because of the constraint in masker construction that the sum of the masker energy in the filter plus the masker energy in the fringe is a constant independent of the choice of masker (plus the fact that, unlike the energy in the filter, which may or may not include target energy, the energy in the fringe could always be assigned unambiguously to the masker).

Fifth, there is a variety of further experiments to conduct that are likely to have important theoretical implications not only with respect to the β-threshold correlation problem, but also with respect to the assumption that variation in masker spectrum influences the listener solely via variation in energy level, and within the domain of energy-level models, the assumption that only the energy level at the output of a single filter containing the target is important.

In one set of such experiments, the basic stimuli would remain the same (a fixed-frequency target tone in multitone-complex maskers with a protected zone around the target), but the masker level variable would be treated differently. Among the experiments envisioned in this class are a "squeeze" experiment and two "roving-level" experiments. In the squeeze experiment, the levels of the various fixed maskers are adjusted prior to the R tests so that all the F_i psychometric functions are not only roughly parallel but also produce the same threshold. In one roving-level experiment, the level of the masker, as well as its spectrum, is randomly varied from trial-to-trial in an R run; in another, only the level of the masker is varied (the spectrum is held fixed, as in an F_i run). The results of such experiments should help separate out the effects of spectrum randomization into those effects that can be ascribed to the associated level randomization and those effects that cannot be so ascribed.[9]

In a second set of experiments, the basic stimuli would be altered and attention would focus not on determining the extent to which level is the key variable, but on results obtainable with this variable. In these experiments, both the target and the masker would consist of Gaussian noise. In the first of these experiments, the target would consist of narrow-band noise; the masker would consist of broadband noise; the central portion of the masker (i.e., that portion that overlaps the target spectrum's) would be roved in level; and the fringe portion of the spectrum would be roved in level in a complementary fashion (so that the total energy in the masker remains constant). In the second experiment, the fringe in the masker would be eliminated (so that the experiment degenerates into an intensity discrimination experiment with a roving reference level). Of particular interest in these experiments would be the β results obtained. Although it is relatively obvious that the correlation between β and threshold for the second experiment would be negative (consistent with our model but not with the results for L1 and L3), it is not obvious how this correlation would come out for the first of these experiments.

Sixth, and finally, we believe that the understanding of IM would benefit greatly from in-depth empirical studies of how performance in various tasks (both inside and outside the IM domain) is correlated, and how performance in these tasks is altered by various types of training. Although some previous work has been conducted along these lines (e.g., Watson *et al.*, 1976; Leek and Watson, 1984, 1988; Neff and Dethlefs, 1995), knowledge in this area remains extremely limited. Unfortunately, these types of studies are very difficult to perform because of the large amounts of data required. Nevertheless, deep understanding of IM cannot be gained without such studies.

ACKNOWLEDGMENTS

We are deeply indebted to our listeners for providing us with such interesting data, to Christine Carter and Sally Tressler for helping us collect this data, and to Dr. Tanya Arbogast, Dr. Laurel Carney, Dr. Robert Gilkey, Dr. G. Bruce Henning, Dr. Scott Isabelle, Dr. Robert Lutfi, Dr. Donna Neff, Dr. Virginia Richards, and Dr. Beverly Wright, as well as Sean Davidson and Yan (Felicia) Gai, for very useful communications during the preparation of this paper. We also wish to thank the associate editor and the reviewers for their substantial contributions to this effort. This work was supported by Grants NIH R01 DC04545 and DC00100, NIH F32 DC006526, and ONR N00014-04-1-0131.

[1]The exact counterpart to this situation in which a multitone masker is held fixed and the frequency of a tonal target is randomized has never been tested; however, the results obtained with approximate counterparts show relatively small randomization effects. Aside from noting that such results are consistent with a detection strategy that focuses on nulling out the masker rather than enhancing the target so that randomizing the masker disturbs detection performance more than randomizing the target (see the discussion of listener min vs listener max in Durlach *et al.*, 2003a), this issue is ignored throughout the rest of this paper. The focus in this article is exclusively on randomization of the masker.

[2]Apart from some subtle statistical issues associated with the definition and identification of performance asymptotes below perfect performance, serious study of the lapse parameter must take account of the differences between such asymptotes that reflect hearing phenomena (which are not likely to persist as asymptotes as the loudness of the target becomes much greater than that of the masker) and asymptotes that reflect nonauditory phenomena such as loss of general attentiveness and/or imperfect motor control in responding to the stimulus. Whereas in the limit, the auditory portion of the lapse parameter must go to zero (i.e., any apparent flattening in the psychometric function below perfect performance must be expected to disappear as the target becomes very loud), there is no reason to believe that the nonauditory portion goes to zero (i.e., that the listener is capable of performing perfectly over an infinite number of trials no matter how loud the target).

[3]Our detailed choice of the target levels to be tested was guided by knowledge about detection thresholds acquired in previous IM experiments for the same listeners. Based on this knowledge, tests were conducted that ensured coverage of the range 60% to 90% correct (the range in which the d'-vs-level function tends to be roughly linear) in 5-dB increments for the F_i tests and 10-dB increments for the R tests, the only exception being the R test for L5, where a 5-dB increment was used in response to what appeared to be an exceptionally steep slope for the R case.

[4]This definition of IM is obviously a very restrictive one. Although useful in the context of this paper, it is not intended to be applicable to all cases of interest. It should also be noted that, whereas the first of these two equations implies the second, the second does not imply the first. Unless explicitly stated otherwise, the term "IM" refers to the case in which neither equality holds. In Sec. VI, we briefly consider the case in which the second equation holds but not the first (which we refer to as "covert IM").

[5]This follows from the fact (discussed in footnote 3) that for the F_i curves the target test levels were chosen for each value of i to focus on the linear portion of F_i, whereas for the $F_{s,i}$ curves the relevant levels are those chosen to focus on the linear portion of R (and are thus independent of i). Had we measured d' in the R case over a much wider range of levels (so that we had a substantial number of data points at levels where $d' \approx 0$ and $d' > 3.0$ for the R case), the $F_{s,i}$ functions would have appeared more like horizontal translations of each other.

[6]In fitting these straight lines to the data, we ignored those points that were strongly influenced by the zero-slope asymptotic behavior of the data at very low or very high values of d'. More specifically, we ignored points outside the interval $0.25 \leq d' \leq 3.0$ that resulted in nonmonotonic behavior of the d'-vs-level function. The number of data points ignored constituted roughly 8% of the total number of data points obtained. As one would expect (see footnote 5), most of the data points ignored came from the $F_{s,i}$ case (the number of points ignored for each case constituted 5% of the F_i points, 12% of the $F_{s,i}$ points, and 0% of the R points).

[7]The fitting procedure we have used involved computing model psychometric functions for various values of the parameters σ_0, W_F, and W_R (about 500 000 such psychometric functions were computed) and determining thresholds and slopes for the linear portion of these functions (in the same manner as for the empirical psychometric functions). Root-mean-square

(rms) deviations were then calculated and denoted T_{dev} and S_{dev} (T_{dev} is the rms deviation between the set of theoretical and empirical thresholds, whereas S_{dev} is the rms deviation for the ratios of theoretical and empirical slopes). Determination of the values to be assigned to the parameters σ_0, W_F, and W_R was then carried out in two steps. First, in fitting the F_i results, the value of σ_0 was chosen to minimize S_{dev} (which depends almost exclusively on σ_0) and, given that value of σ_0, the value of the bandwidth W_F was chosen to minimize T_{dev} (the value of T_{dev} thus obtained being within 1 dB of the minimum T_{dev} obtainable over all σ_0). Second, with the same value of σ_0 being maintained, the value of W_R was chosen to fit the $F_{s,i}$ results by choosing this value to minimize T_{dev} (because T_{dev} depended much more strongly on W_R than S_{dev} depended on W_R). Although we cannot claim that the fits we have obtained using the given values of σ_0, W_F, and W_R are the best fits that can be obtained according to some explicit fitting criterion X, we can claim that all such optimum fits would be better than or equal to the relatively good fit demonstrated in Fig. 4 and Table III with respect to that criterion. Furthermore, based on the values of T_{dev} and S_{dev} shown in Table III for the values of σ_0, W_F, and W_R chosen (together with some further computations), we have shown that our results (in terms of the values of σ_0, W_F, W_R, T_{dev}, and S_{dev} obtained) are roughly comparable to the results that would be obtained using any linear fitting criterion in which the weights assigned to T_{dev} and S_{dev} result in a trading ratio such that 1 dB in T_{dev} corresponds to roughly 6%–12% in S_{dev}.

[8]Note, however, that this result is surprising only to the extent that one believes that the main effect of the trial-to-trial variation in masker spectrum in regions away from the target frequency is to cause the listener to attend to these nontarget spectral regions (i.e., to distract the listener from the task at hand). One might equally well believe that the difficulty associated with these trial-to-trial variations stimulates highly motivated listeners to focus more intensely on the target region.

[9]It is amusing to note that to the extent the increased masking caused by uncertainty in the masker spectrum can be explained in terms of uncertainty in the energy level at the output of a peripheral filter with critical-band-like filter characteristics, it may not be appropriate to use the term "IM" to refer to this increased masking. Although the increased masking would still be the result of stimulus uncertainty, it would not necessarily be the result of imperfections in central processing. In other words, it is possible that in certain cases the detection performance would be consistent with a model in which the central processing is ideal and the only defect in the processing results from the peripheral limitation on frequency resolution associated with the critical band (and therefore that the masking should be thought of as energetic rather than informational). For a discussion of definitional issues, see Durlach *et al.*, 2003b and Watson, 2005.

Allen, P., and Wightman, F. (**1994**). "Psychometric functions for children's detection of tones in noise," J. Speech Hear. Res. **37**, 205–215.

Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Davidson, S. A., and Carney, L. H. (**2003**). "Monaural and diotic detection of tones in narrow-band and wideband reproducible noise maskers," J. Acoust. Soc. Am. **113**(4), 2197.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003a**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003b**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (**2002**). "Binaural detection with narrow-band and wideband reproducible noise maskers. I. Results for human," J. Acoust. Soc. Am. **111**, 336–345.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207–1219.

Green, D. M. (**1961**). "Detection of auditory sinusoids of uncertain frequency," J. Acoust. Soc. Am. **33**, 897–903.

Isabelle, S. K., and Colburn, H. S. (**1991**). "Detection of tones in reproducible narrow-band noise," J. Acoust. Soc. Am. **89**, 352–359.

Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (**2002**). "Similarity, uncertainty and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.

Kidd, G., Jr., Mason, C. R., and Richards, V. M. (**2003**). "Multiple bursts, multiple looks and stream coherence in the release from informational masking," J. Acoust. Soc. Am. **114**, 2835–2845.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Leek, M. R., and Watson, C. S. (**1984**). "Learning to detect auditory pattern components," J. Acoust. Soc. Am. **76**, 1037–1044.

Leek, M. R., and Watson, C. S. (**1988**). "Auditory perceptual learning of tonal patterns," Percept. Psychophys. **43**, 389–394.

Lutfi, R. A. (**1990**). "How much masking is informational masking?" J. Acoust. Soc. Am. **80**, 2607–2610.

Lutfi, R. A. (**1993**). "A model of auditory pattern analysis based on component-relative-entropy," J. Acoust. Soc. Am. **94**, 748–758.

Lutfi, R. A., Kistler, D. J., Callahan, M. R., and Wightman, F. L. (**2003a**). "Psychometric functions for informational masking," J. Acoust. Soc. Am. **114**, 3273–3282.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003b**). "One factor underlies individual differences in auditory informational masking within and across age groups," Percept. Psychophys. **65**, 396–406.

Neff, D. L. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.

Neff, D. L., and Callaghan, B. P. (**1987**). "Psychometric functions for multicomponent maskers with spectral uncertainty," J. Acoust. Soc. Am. **81**(S1), 53.

Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.

Neff, D. L., and Jesteadt, W. (**1996**). "Intensity discrimination in the presence of random-frequency, multicomponent maskers and broadband noise," J. Acoust. Soc. Am. **100**, 2289–2298.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). "Informational masking for multicomponent maskers with spectral gaps," J. Acoust. Soc. Am. **94**, 3112–3126.

Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.

Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).

Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.

Richards, V. M., Huang, R., and Kidd, G., Jr. (**2004**). "Masker-first advantage for cues in informational masking," J. Acoust. Soc. Am. **116**, 2278.

Richards, V. M., Tang, Z., and Kidd, G., Jr. (**2002**). "Informational masking with small set sizes," J. Acoust. Soc. Am. **111**, 1359–1366.

Siegel, R. A., and Colburn, H. S. (**1989**). "Binaural processing of noisy stimuli: Internal/external noise ratios for diotic and dichotic stimuli," J. Acoust. Soc. Am. **86**, 2122–2128.

Tang, Z., and Richards, V. M. (**2003**). "Examination of a linear model in an

informational masking study," J. Acoust. Soc. Am. **114**, 361–367.
Watson, C. S. (**1987**). "Uncertainty, informational masking and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.
Watson, C. S. (**2005**). "Some comments on informational masking," Acta. Acust. Acust. **91**, 502–512.
Watson, C. S., and Kelly, W. J. (**1981**). "The role of stimulus uncertainty in the discriminability of auditory patterns," in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard, Jr. (Erlbaum, Hillsdale, NJ), pp. 37–59.
Watson, C. S., Kelly, W. J., and Wroton, H. W., (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.
Wright, B. A., and Saberi, K., (**1999**). "Strategies used to detect auditory signals in small sets of random maskers," J. Acoust. Soc. Am. **105**, 1765–1775.