Evidence of stochastic resonance in an auditory discrimination task may reflect response bias

DANIEL SHEPHERD AND MICHAEL J. HAUTUS
University of Auckland, Auckland, New Zealand

Stochastic resonance is a mechanism that improves the detection of weak periodic signals when an optimal level of background noise is present. This counterintuitive concept is thought to involve a cooperative effect that arises out of a physical coupling between deterministic and random dynamics in nonlinear systems. We report the results of three psychophysical experiments designed to probe the auditory system for evidence of stochastic resonance. In all three experiments, participants judged which of two 1000-Hz tones was greater in intensity, with one of the tones embedded in a background of Gaussian noise. Experiment 1 used a two-alternative forced choice task and provided evidence of enhanced discriminability at an optimal level of noise. The aim of Experiment 2 was to determine whether the apparent enhancement observed in Experiment 1 was due to sensory processes or to a response bias. Receiver operating characteristic analysis supported an explanation based on decision processes, rather than a genuine sensory enhancement. An additional experiment (Experiment 3) was used to discount experimental artifacts.

The extraction of meaningful signals from a noise background is the concern of both electrical engineers and participants relying on their biological sensory apparatus. Noise in dynamical systems has traditionally been viewed as counterproductive. In information theory, the presence of noise in a dynamical system has been thought to exert only negative effects, and since the invention of the telegraph, engineers have sought to expunge the random processes of noise that plague information transmission. However, studies have shown that in certain nonlinear systems, a weak signal can be made more apparent by the presence of noise (Gammaitoni, Hanggi, Jung, & Marchesoni, 1998; McNamara & Wiesenfeld, 1989). Theorists have begun to rethink the role of noise in information transfer, and one phenomenon that has been the recipient of much attention is the contrarily named *stochastic resonance* (SR; Benzi, Sutera, & Vulpiani, 1981).

SR generally concerns a cooperative effect that arises out of the coupling between deterministic and random dynamics in nonlinear systems. To play a constructive role in the elucidation of weak (i.e., subthreshold) signals, the noise must be of an optimal level, although not necessarily random (Maddox, 1994). Study into the nature of SR has proceeded with vigor within physics (for a review, see Gammaitoni et al., 1998; Wiesenfeld & Moss, 1995), and evidence for the existence of SR is accumulating in the biological and psychological sciences. Generally, biological sensory systems can be modeled as intrinsically nonlinear processes that operate in noisy environments; hence, they

are candidates for SR. In biological sensory systems, an optimal level of noise is thought to exert two different effects. First, a gross physical gain in the amplitude of the stimulus, termed Type E (for *energy*) SR by Moss, Ward, and Sannita (2004). Second, an increase in the fidelity of signal encoding (Longtin, Bulsara, & Moss, 1991), known as Type I (for *information*) SR (Moss et al., 2004).

Strong evidence for SR in nonprimate sensory systems has been reported (Douglas, Wilkens, Pantazelou, & Moss, 1993; Levin & Miller, 1996; Moss et al., 2004; Ries, 2007), prompting experiments to be undertaken on humans (Collins, Imhoff, & Grigg, 1996; Moss et al., 2004; Riani & Simonotto, 1994; Simonotto et al., 1997). To date, evidence suggests that SR may be at work in both human sensory and perceptual processes (Moss et al., 2004), and to a degree that manifests itself in psychophysical data.

The research we report here is a search for evidence of a constructive effect of noise in the auditory system. It has been argued elsewhere that the auditory system fulfills all the requirements necessary for SR to occur (Benham & Zeng, 2003; Jaramillo & Wiesenfeld, 1998; Morse & Evans, 1996; Moss et al., 2004; Parnas, 1996; Zeng, Fu, & Morse, 2000)—notably, its being a nonlinear system processing deterministic signals embedded in a background of random noise. Furthermore, psychoacoustic experiments have indicated that optimal nonzero noise levels maximize the sensitivity of humans participating in some detection tasks (Ries, 2007; Ward, Desai, Rootman, Tata, & Moss, 2001; Zeng et al., 2000).

D. Shepherd, daniel.shepherd@auckland.ac.nz

Given that the auditory system possesses the characteristics to exhibit SR and that the phenomenon has been implicated as a possible mechanism in the detection of auditory stimuli, we asked whether or not SR could be induced in discrimination tasks. Specifically, can the addition of an optimal level of noise enhance a participant's capacity to correctly judge whether two tones of unequal intensity are different?

EXPERIMENT 1

A two-alternative forced choice (2AFC) task was employed, with participants being presented two tones: a standard tone and a comparison tone. The standard tone was always presented in quiet, whereas the comparison tone was embedded in a gated burst of Gaussian noise. Pilot studies identified a range of optimal noise levels and tone intensities that induced a nonmonotonic relationship between proportion correct and noise level. It was hypothesized that, in a 2AFC task in which a participant must judge two pure tones on the basis of intensity, the addition of an optimal level of noise to one, and only one, of the stimulus intervals would affect the participant's sensitivity in a specific manner. If, for example, the tone in noise was more intense than the tone in quiet, enhanced discriminability would be expected. If, on the other hand, the tone in noise was less intense than the tone in quiet, we predicted that the less intense tone would become amplified and, therefore, be less discriminable from the tone without noise. If this were the case, performance would settle on chance or less than chance. These hypotheses rested on the presumed existence of an optimal level of noise.

Method

Participants

Four adults with normal audiometric thresholds served as participants. There were 3 males (D.S. [the first author], 25 years of age; D.W., 29 years of age; and G.B., 32 years of age) and 1 female (M.W., 31 years of age). All listened monaurally with the left ear.

Materials

Tones. A synthesizer (Hewlett-Packard 8904A) generated 1000-Hz tones that were gated by an electronic switch (TTES, SW1) to a duration of 200 msec and were presented in a Hanning ($\cos^2$) window with 50-msec rise and fall times. An attenuator (Tucker Davis Technologies [TDT], PA4) then scaled the tone to 67, 68, or 69 dB SPL ($re\ 2 \times 10^{-5} \text{ N/m}^2$).

Noise. Narrowband Gaussian noise was generated (TDT, AP1) with a 3-dB bandwidth of 200 Hz and a center frequency of 1000 Hz. The noise was generated on a 16-bit converter (TDT, DA1) prior to anti-aliasing (TDT, FT5), was gated to 200 msec (TDT, SW1), and was presented in a Hanning window with 50-msec rise and fall times. The noise was then attenuated to the required spectrum level (TDT, PA4).

Other materials. The operations of the synthesizer, electronic switches, and attenuators were all controlled by a personal computer. Once the tone and, for the observation interval containing the comparison tone, noise had been generated, they were combined in a signal mixer (TDT, SM5) and then routed to a headphone buffer (TDT, HB6) and then to the participant's earpiece (TDH, 49P). Responses were made on a standard keyboard from within a sound-attenuating chamber (Amplaid, Model E).

Procedure

The participants partook in a 2AFC difference discrimination task and were required to discriminate between a standard and a comparison tone. Each trial contained two observation intervals, with the standard and comparison tones each randomly assigned to one of the intervals. The comparison tone was presented in Gaussian noise, whereas the standard tone was presented in quiet and was always 68 dB SPL. The comparison tone was embedded in one of five levels of Gaussian noise (40, 50, 60, 70, or 80 dB SPL, noise power density) and was of greater (69 dB SPL) or lesser (67 dB SPL) intensity than the standard.

An experimental trial began with a 250-msec warning light. Both observation intervals were of 200-msec duration, were separated by an interstimulus interval of 250 msec, and were indicated by an LED. The participants were required to report, via a keypad, which of the observation intervals contained the most intense tone. Feedback was provided, and the next trial began after 500 msec.

After 500 practice trials, the participants undertook 13 blocks of 90 trials across 6 weeks. The first 10 trials of any block were discarded, leaving 80 trials for subsequent analysis. Each combination of comparison tone and noise was presented randomly across a block of trials eight times. Thus, on any 1 trial, there was a 50% chance that the comparison tone would have a higher or a lower intensity than the standard tone. In all, 13 blocks were undertaken, yielding 104 (13×8) trials per comparison stimulus, per participant.

Results and Discussion

Figure 1 presents proportion correct as a function of noise level for the two intensities of comparison tone (67 and 69 dB SPL). Each point is based on 104 trials across 13 experimental blocks. Confidence intervals (95%) are included. Note that the functions are nonmonotonic and are remarkably consistent in shape across the 4 participants. Classical auditory masking experiments would predict, on the basis that participants do not compensate for the effects of noise when judging intensity, that as masker level increases, the performance of the participant should converge on chance (Green & Swets, 1966). When the comparison tone was greater than the standard tone, the effect of the added noise was to initially increase proportion correct. Note that as the noise level increases, the functions for 3 of the 4 participants converge on chance.

In psychophysical theory, it is generally accepted that the relationship between signal and noise is monotonic in form, with performance decreasing as a function of noise level (Green & Swets, 1966). Figure 1 demonstrates that this linear relationship has not been realized. If SR is indeed operating, a general linear decline will occur, but at some point, corresponding to an optimal noise level, there will be an increase in performance. When the comparison tone was more intense than the standard tone, the proportion of correct responses, for all 4 participants, initially increased, then plateaued, and then decreased. Such behavior is consistent with an SR process in which an optimal level of noise coincides with peak performance, with noise intensities extending beyond the optimal level producing masking. When the comparison tone was less intense than the standard tone, a nonlinear relationship between noise level and proportion correct was likewise observed. However, rather than performance improving, it steadily degraded until, at the highest noise level, it started to converge on chance. Such behavior can again

Figure 1. Proportion correct as a function of background noise level for two comparison tone intensities: 67 (□) and 69 (○) dB SPL. The dashed horizontal line indicates chance performance, and 95% confidence intervals are included. Data are for 4 participants.

be explained by Type E SR. When a physical coupling occurs between an optimal level of noise and the comparison tone, participants will judge that comparison tone to be of similar intensity as, or greater intensity than, the standard. As a consequence, participants will incorrectly indicate that the interval containing noise also contained the most intense tone. Thus, for both intensities of comparison tone, an SR process predicts that the decrease in sensitivity will not be a linear function with respect to increasing levels of noise, and this is reflected in the data. For 3 of the 4 participants (Figure 1), the minima of the lower functions (i.e., comparison tone less than the standard) correspond to the maxima of the upper functions (i.e., comparison tone greater than the standard). Such a relationship is consistent with an optimal noise level, which for all 4 participants appears between 60 and 70 dB SPL.

However, although SR is one explanation for these non-linear functions, it is not the sole explanation. It is conceivable that a response bias, rather than a true sensory enhancement, could account for the data. For example, if participants had a bias to respond to the interval contain-

ing the noise burst, we would expect data similar to those obtained. An explanation based on response bias cannot, however, explain why proportion correct would converge upon chance as noise level increased. Note also that both SR and response bias can effectively account for the individual differences evident in the data. The decisional processes underlying response biases are themselves grounded in epistemic or cognitive processes (Marks, 1997), and so individual differences would be expected. In terms of SR processes, there is no reason to assume that a universal level of optimal noise would exist, identical in quantity and effect, between all participants. Experiment 1 did not directly test between the two competing explanations—sensory enhancement or response bias.

To further explore the possibility that bias determines the shape of the functions in Figure 1, additional analyses were undertaken.¹ Figure 2A presents group data showing percentage of trials on which the noise interval was selected as a function of noise level for two intensities of comparison tone: 67 and 69 dB SPL. The function for the 67-dB SPL comparison tone in Figure 2A indicates

Figure 2. In the left panel (A) Group data showing percentage of trials on which the noise interval was selected as a function of noise level for two intensities of comparison tone: 67 (○) and 69 dB SPL (□). (B) Percentages of overall preference to the noise interval (Δ) and discriminability between the two comparison tones relative to the standard (▽) are plotted as a function of noise level for group data. The dashed line in both panels equals 50% correct.

a rapid change from responses favoring the interval not containing the noise to that containing the noise. Although this could represent a response bias toward the interval containing the noise, it could also indicate an enhancement of perceptual magnitude due to stochastic resonance, because stochastic resonance would initially make the perception of the 67-dB SPL tone more akin to that of the 68-dB SPL standard tone and could conceivably make the 67-dB SPL tone appear even larger in magnitude than the control. This, in essence, would ensure a plethora of incorrect responses, which in this case is synonymous with selecting the interval containing the noise.

For the 69-dB SPL comparison tone, a similar, but less dramatic, shift is observed in response tendency. The shift must necessarily be less dramatic because responses are primarily directed to the noise interval, even for low levels of noise; this interval is always the correct interval to select for the 69-dB SPL comparison tone. For this case, an increase in preference to respond to the noise interval is also consistent with stochastic resonance. For both comparison tones, the observation that this preference is reduced for the highest level of noise is also consistent with stochastic resonance. There should be an optimal level of noise that produces a maximum increase in perceptual magnitude, beyond which a genuine masking effect takes over. Figure 2A shows this level to be around 60 or 70 dB SPL.

Unfortunately, these results are also consistent with a response bias that is related in a nonmonotonic way with the level of the noise. It is even conceivable that both processes (bias and stochastic resonance) operate in a manner that makes the disentanglement of their effects extremely difficult.

Figure 2B may help clarify this issue. The “preference” function in Figure 2B is the average of the two

functions in Figure 2A and represents an overall preference for responses to the interval that contains the noise. The function labeled “discriminability” in Figure 2B is the difference between the upper and lower functions presented in Figure 2A and represents an indirect index of the perceptual distance between the two comparison tones. Discriminability clearly drops away as the level of the noise increases. The 67- and 69-dB SPL tones become increasingly less discriminable, which may be consistent with a masking effect that grows monotonically with noise level. However, this too is consistent with stochastic resonance. If the perceptual magnitude of the 67-dB SPL tone increases with noise level at a faster rate than does the 69 dB SPL tone—the latter already “requires” responses to the interval containing noise and is, therefore, constrained in much movement toward greater preference for this “correct” interval—the result will be a reduced ability to perceptually discriminate between the two tones. This is exactly what Figure 2B illustrates.

For the highest noise level (80 dB SPL), masking clearly has an effect. The distributions of responses for both comparison tones converge on 50%. This reflects, as compared with immediately lower noise levels, getting more incorrect responses for the 69-dB SPL tone and more correct responses for the 67-dB SPL tone. This convergence toward chance performance is consistent with a falloff in the effect of stochastic resonance as the noise level passes its optimal level.

EXPERIMENT 2

Experiment 1 suggests that the addition of an optimal noise background may aid the discrimination of two tones differing slightly in intensity. However, the data cannot tell

us which of the two, sensory enhancement or decisional processes, is the more likely explanation, although the simpler explanation is response bias. Odgaard, Arieh, and Marks (2003) argued that signal detection theory (SDT) can determine whether a genuine sensory enhancement is occurring or whether some other process is at work. The ability of SDT to control for response bias may help determine whether the results from Experiment 1 are attributable to an SR-like mechanism, and so Experiment 2 was conducted within the framework of SDT.

The general design of Experiment 2 was taken from Irwin and Whitehead (1991), Irwin, Hautus, Dawson, Welch, and Bayly (1994), and Wang, Irwin, and Hautus (1998). Ten combinations of noise (five levels) and comparison tone (two levels) were compared with a standard tone embedded in the lowest level of noise (the *reference*). If there were to be a genuine enhancement of early sensory processes (i.e., SR), the data should retain the nonlinearities evident in Figure 1. Conversely, if the form of the data in Experiment 1 was a product of response bias, then, as the noise level increased, a systemic decrease in sensitivity should be observed.

We note, as have others (e.g., Inchiosa & Bulsara, 1996; Moss et al., 2004; Tougaard, 2000; Ward, 2004; Ward, Neiman, & Moss, 2002) that although models of SR explicitly involve bistable systems possessing threshold characteristics, the notion of a sensory threshold is in conflict with the assumptions of SDT. Instead of a sensory threshold, SDT posits a decision-related nonlinearity (i.e., a response criterion) that is constantly changing. Thus, at a theoretical level, the two approaches to psychophysical processes are incompatible. However, at a practical level, a signal detection experiment can still expose processes such as SR, which would manifest in sensitivity estimates peaking at some optimal level of noise (Moss et al., 2004).

Method

Participants

Four males (D.S., D.W., G.B., and M.H. [the second author]) and 1 female (M.W.) served as participants. All but M.H. (male, 32 years of age) had participated in Experiment 1.

Materials

The equipment and methods used to generate tones and noise bursts were identical to those used in Experiment 1.

Procedure

The procedure used in Experiment 2 was similar to that used in Experiment 1, but with the following differences. First, the method of constant stimuli was employed, as opposed to the 2AFC difference discrimination task utilized in Experiment 1. Each trial consisted of two observation intervals that contained a standard and a comparison tone, but the standard tone was always presented in the first interval. Second, the binary response regime was replaced with a 6-point rating scale, with participants rating their a posteriori confidence about the intensity of the comparison tone, relative to the standard. If the participant was *confident that the comparison was of a lower intensity than the standard*, he or she responded "1." A rating of "6" indicated *confidence that the comparison was of a greater intensity*. Third, 11 possible combinations of signal and noise were present in the second interval, as opposed to the 10 used in Experiment 1. The use of receiver-operating characteristic (ROC) analysis made it necessary to create a new combination of comparison tone

and noise to use as a reference stimulus. Specifically, the reference stimulus consisted of a new comparison tone equal in intensity to the standard stimulus (68 dB SPL) and embedded in the lowest level of noise (40 dB SPL).²

Experimental sessions consisted of 2 or 3 blocks of 121 trials. In any 1 block, the first 11 trials were discarded from analysis, leaving 110 trials (10 per comparison level) per block to be analyzed. Analysis was based on 20 blocks, giving 200 trials per comparison type.

Data Analysis

The framework of SDT was used to analyze the rating scale data. In this context, responses to *noise* trials were defined as participant ratings to the reference stimulus. Responses to *signal* trials were defined as participant ratings to the comparison tones that were of greater or lesser magnitude than the standard. Hence, there were 10 sets of signal ratings being compared with one set of reference (i.e., noise) ratings, resulting in 10 values of d' .

A peculiarity of the analysis arises that some of the computed values of d' are negative. This is a product of the signal distribution's having a mean less than the noise distribution and participants' consistently judging this to be the case. A value of d' less than zero indicates that the participant has perceived the comparison tones to be less intense than the standard tone. All measures of d' reported here should, however, be treated as absolute values. To aid comparative analysis, the negative signs have been retained, although a negative d' does not imply a poorer degree of sensitivity, relative to a positive value. Figure 3 displays three ROC curves from participant M.W. and illustrates the form of a negative d' value. Note that the area between the ROC and the major diagonal is proportional to sensitivity.

Results and Discussion

Criterion-free indices of discrimination, d' , were calculated using maximum likelihood estimation (e.g., Dorf-

Figure 3. Three receiver-operating characteristic (ROC) curves, based on the normal-normal equal variance model, generated using data from M.W. The uppermost (circles) ROC ($d' = 1.04$) is for a 69-dB SPL comparison tone embedded in 40 dB SPL of noise, and the lowermost ROC ($d' = -1.22$) is for a 67-dB SPL tone embedded in the same noise. The middle ROC ($d' = -0.11$) is for a 69-dB SPL tone in 80-dB SPL noise.

man & Alf, 1968), where each value of d' indicates the sensitivity of the participant in discriminating between the standard tone embedded in 40-dB SPL noise and any 1 of the 10 combinations of comparison tones and background noise. Adequate fits, as defined by chi-square goodness-of-fit statistics ($p > .05$), of the normal-normal equal-variance model were obtained in all cases. Inspection of the points within the ROC space failed to uncover any evidence of participants' adopting either exclusively liberal (i.e., points clustered to the upper right) or exclusively conservative (i.e., points clustered to the lower left) response criteria. This suggests that the bias may be task specific and that, although the 6-point rating scale utilized in this experiment encouraged more measured patterns of responding, the forced choice nature of Experiment 1 may have led to a more cavalier response style.

Figure 4 displays the sensitivity index, d' , as a function of background noise level for 5 participants. Displayed also (Figure 4, lower right panel) are the group data. With respect to individual data, the plots presented in Figure 4 show, for the most part, a steady decline in sensitivity as background noise increases. One exception is participant G.B., who, for the case in which the comparison tone was more intense than the standard, exhibits a nonmonotonic function peaking at the 50-dB SPL noise level.

In regard to aggregated data, one must proceed cautiously (MacMillan & Kaplan, 1985) and care must be taken not to underestimate true sensitivity. Jackknifing (e.g., Dorfman & Berbaum, 1986) avoids the pitfalls normally associated with group data that come from participants with different sensitivities and response biases (Irwin, Hautus, & Stillman, 1992). The group data in Figure 4 show that the trends observed in the individual data have been preserved and that a quasilinear decline in sensitivity occurs as background noise increases.

The results of Experiment 2 fail to provide support for the notion that noise can play a constructive role in auditory discrimination processes. The individual data reveal that sensitivity toward the two tone bursts did not have a nonmonotonic form like that seen in Experiment 1. Instead, sensitivity systematically declined in a manner compatible with contemporary psychophysical theory (Green & Swets, 1966)—that is, a partial masking effect giving way to total masking as the noise background increases. The absence of a signature SR effect in Experiment 2 counters the argument that a sensory enhancement occurred in Experiment 1. If a process such as SR was in effect, a characteristic spike in the functions would be expected, and this feature was absent once response bias was controlled for. Participant G.B., it could be argued, has functions that are consistent with an explanation based on SR, although G.B.'s functions are different on two counts, when compared with the data from Experiment 1. First, G.B. exhibited a peak between 60- and 70-dB SPL noise level in Experiment 1, whereas in Experiment 2, the peak was at 50-dB SPL noise level. Second, the functions in Experiment 1 can be explained by a gross physical gain in the comparison tone (i.e., Type E SR), whereas those in Experiment 2 cannot.

EXPERIMENT 3

All the participants (excluding M.H.) undertook the 2AFC task first, followed by the rating scale procedure. Could the differences between the two studies be explained by an order or learning effect? To test this possible explanation, Experiment 1 was repeated by 3 of the participants.

Method

Participants

Two males (D.S. and D.W.) and 1 female (M.W.) served as participants. All the participants had completed Experiment 1.

Materials

The same stimuli as those employed in Experiment 1 were used in Experiment 3. The equipment and methods used to generate tones and noise bursts were identical to those used in Experiment 1.

Procedure

The procedure was identical to that employed in Experiment 1.

Results and Discussion

Figure 5 illustrates proportion correct as a function of background noise level for 67- and 69-dB comparison tones. As in Experiment 1, the functions are nonmonotonic and demonstrate a nonlinear relationship between discriminability and noise level. Again, the form of these functions could be explained by a nonlinear mechanism, such as SR, or by response bias.

For each individual, differences are apparent when the data from this experiment are compared with those from Experiment 1. For participant D.S., performance was greatly improved at the 40-, 70-, and 80-dB SPL noise levels, whereas for D.W., performance at the highest noise levels (i.e., 70 and 80 dB SPL) improved for the 67-dB SPL comparison but degraded for the 69-dB SPL comparison. For participant M.W., performance on both comparison tones improved at the lowest noise level (i.e., 40 dB SPL). There are two explanations to account for these differences, although these data cannot distinguish them: First, a practice effect arose for each participant undertaking Experiments 1 and 2; second, the nature of the bias changed across the two experiments. If the latter explanation is assumed, the differences between Experiments 1 and 3 serve as a powerful demonstration that bias, and not a genuine sensory enhancement, underlies the nonlinear functions obtained in the two experiments.

GENERAL DISCUSSION

The objective of this study was to probe the auditory system for evidence of SR; that is, background noise at optimal levels can interact with acoustic stimuli and consequently improve their discriminability. If, on face value, one were to accept the results from Experiment 1, it could be concluded that noise can facilitate the discrimination of two tones differing in intensity. The data from Experiment 2, however, did not reveal such an effect, with the key difference between the two experiments being that

Figure 4. Participant sensitivity (d') plotted as a function of noise level for two comparison tone intensities: 67 ($\square$) and 69 ($\circ$) dB SPL. Plots are for 5 participants and for jackknifed data (lower right panel). The dashed horizontal lines represent zero sensitivity, and error bars represent 95% confidence intervals.

response bias was controlled better in Experiment 2. In contrast to the predictions of SR, an explanation of the results based on decisional processes can, however, account for the differences obtained between Experiments 1 and 2. An additional experiment, Experiment 3, indicates

that the difference between Experiments 1 and 2 is not an artifact of order.

The stimuli utilized in both Experiments 1 and 2 were, aside from the addition of a reference stimulus in Experiment 2, identical. Consequently, an unbiased participant

Figure 5. Proportion correct as a function of background noise level for two comparison tone intensities: 67 (□) and 69 (○) dB SPL. The dashed horizontal line indicates chance performance. Data are for 3 participants, and 95% confidence intervals are included.

should produce similar data between the two experiments. A well-known theorem in psychophysics, Green's area theorem (Green & Swets, 1966), states that the area under a yes-no or rating scale ROC is equal to proportion correct in a 2AFC procedure for an unbiased participant. Therefore, for an unbiased participant, a transformation of the proportion correct judgments in Experiment 1 to their equivalent d' values would be expected to yield functions comparable to those obtained in Experiment 2. To further highlight the incompatibility of the results obtained in Experiments 1 and 2, the data obtained in Experiment 1 were converted to d' (Hacker & Ratcliff, 1979). The resulting functions were indistinguishable from those in Figure 1 and, hence, are in contrast to those displayed in Figure 4. This incompatibility is further evidence of response bias.

Several commentaries and studies in the literature have served warnings regarding the substantial role played by decisional processes in sensory judgments (Marks, 1997). Psychophysics itself underwent a major paradigm shift some 50 years ago in response to the confounding role that nonsensory factors play in sensory measurements (Green & Swets, 1966). Measures of psychophysical performance based on thresholds or proportion correct were misrepresenting true sensitivity by failing to disassociate sensory factors from decisional factors. Contrary to these warnings, threshold and percentage correct measures still dominate experimental psychology, although, usually, as a consequence of convenience rather than conviction. The experiments performed here and by others (Odgaard et al., 2003; Wang et al., 1998) indicate that psychophysical procedures failing to control for response biases can potentially yield invalid data and lead to incorrect conclusions. Marks (1993, 1996, 1997) and collaborators (Aylor & Marks, 1976; Ben-Artzi & Marks, 1995; Odgaard et al., 2003; Rankin & Marks, 1991) and others (Kuze, 1995; Ward, 1979) have repeatedly warned of the interaction between decisional processes and stimulus context in sensory data, and the results reported here speak in support of the widespread adoption of criterion-controlled methods such as ROC analysis.

The most likely form of response bias in Experiment 1 would be a bias toward the interval containing the noise burst. In this respect, participants are behaving as energy detectors, selecting the observation interval containing the greater amount of energy (Pfafflin & Mathews, 1962). Even this explanation, however, fails to account for the results obtained across the range of noise levels, and the nature of the bias is likely to be multifaceted. For the lower noise levels (i.e., 40 and 50 dB SPL), there is no evidence of a noise bias. This can be explained away by arguing that, at these levels, there is not a sufficient level of masking to induce the noise bias. The results from Experiment 2 lend credence to this account, suggesting that the noise begins to actively mask only from 60 dB SPL and above. Figure 4 shows that sensitivity stays relatively constant across the three weakest noise levels and may, in fact, exhibit a nonlinearity across the entire range. The expected linear decline in d' as noise level increases did not occur, and instead, there appears a critical masking point at the 60- to 70-dB noise spectrum level. Thus, in respect to the present experimental paradigm, the masking of the comparison tone may follow a threshold characteristic. Such an effect is evident in the scaling data presented by Scharf (1964), who, to the authors' knowledge, is the only other researcher to use the unique stimulus configuration (i.e., noise in one interval only) employed in this study. It follows from Experiment 2 that if the lowest levels of noise did not provide sufficient masking to induce a noise bias in Experiment 1, then, at noise levels where partial masking begins (i.e., about 70 dB SPL), the functions in Figure 1 should peak, which for 3 of the 4 participants they do.

As with most psychological experiments, shortcomings exist. First, the range of noises used could have been expanded to further examine the behavior of the functions presented in Figures 1 and 4. For example, in Figure 1, do both functions converge on chance as the noise level increases, or is there crossover as the bias switches from the noise interval to the quiet interval? In Figure 4, does sensitivity increase as noise levels are further decreased,

or do they stay constant, as suggested by the lowest three noise levels?

Although the results reported here fail to uphold the discovery of SR in auditory intensity resolution, we believe that further effort should still be expended. From these experiments, we have experienced the difficulties in psychophysically quantifying SR first hand. One barrier may be that, throughout human evolution, the auditory system has tuned itself into the endogenous noise, rendering the endogenous noise the optimal noise level (Ries, 2007; Ward, 2004). If this were the case, it would be extremely difficult, if not impossible, to construct nonmonotonic functions peaking at the optimal level. A second hindrance, relevant especially to detection experiments, is the current state of technology. Detection experiments probing the auditory system for SR demand noise levels that are below the noise floor of many signal-generating devices common in psychophysical laboratories. Third, in the case of difference discrimination experiments, another factor centers on the gain that SR would induce if it were indeed operating in the auditory system. If the gain is not proportional to the stimulus and is, instead, a constant value, the effect of SR may diminish as the stimulus (e.g., the comparison tone in this series of experiments) increases. If, on the other hand, the gain is proportional to the level of the stimulus embedded in noise, then, in light of Weber's law, that gain would have to be of sufficient value to exceed the difference threshold. Thus, a gain may indeed be occurring, but it would fail to manifest itself at the psychophysical level, even though it could be present at the neural level.

In summary, the method we present, in which only a single interval contains noise, is potentially an effective and valid method with which to probe the sensory modalities for Type E SR. However, as our data demonstrate, this approach can be confounded if the participant has a bias toward the interval containing the noise. We are not aware of any other discrimination data based on methods utilizing the same stimulus configuration that we use, or any SR studies dedicated to intensity discrimination, as opposed to detection. However, on the basis of our data, we recommend that any investigation of SR in discriminatory processes engage methods that permit a disentanglement of sensitivity and response bias.

AUTHOR NOTE

This research was supported by a grant from the University of Auckland's Research Committee. Correspondence concerning this article should be addressed to D. Shepherd, Psychophysics Laboratory, Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand (e-mail: daniel.shepherd@auckland.ac.nz).

REFERENCES

- AYLOR, L., & MARKS, L. E. (1976). Perception of noise through barriers. *Journal of the Acoustical Society of America*, **59**, 397-400.
- BEN-ARTZI, E., & MARKS, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, **57**, 1151-1162.
- BENHAM, S. E., & ZENG, F. G. (2003). Noise improves suprathreshold discrimination in cochlear-implant listeners. *Hearing Research*, **186**, 91-93.
- BENZI, R., SUTERA, S., & VULPIANI, A. (1981). The mechanism of stochastic resonance. *Journal of Physics A*, **14**, L453-L457.
- COLLINS, J. J., IMHOFF, T. T., & GRIGG, P. (1996). Noise-enhanced tactile sensation. *Nature*, **383**, 770.
- DORFMAN, D. D., & ALF, E. (1968). Maximum likelihood estimation of parameters of signal detection theory: A direct solution. *Psychometrika*, **33**, 117-124.
- DORFMAN, D. D., & BERBAUM, K. S. (1986). RSCORE-J: Pooled rating-method data. A computer program for analyzing pooled ROC curves. *Behavior Research Methods, Instruments, & Computers*, **18**, 452-462.
- DOUGLAS, J. K., WILKENS, L., PANTAZELOU, E., & MOSS, F. (1993). Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, **365**, 337-339.
- GAMMAITONI, L., HANGGI, P., JUNG, P., & MARCHESONI, F. (1998). Stochastic resonance. *Review of Modern Physics*, **70**, 223-288.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HACKER, M. J., & RATCLIFF, R. (1979). A revised table of d' for M -alternative forced choice. *Perception & Psychophysics*, **26**, 168-170.
- INCHIOSA, M. E., & BULSARA, A. R. (1996). Signal detection statistics of stochastic resonators. *Physical Review E*, **53**, R2021-R2023.
- IRWIN, R. J., HAUTUS, M. J., DAWSON, N. J., WELCH, D., & BAYLY, M. F. (1994). Discriminability of electrocutaneous stimuli after topical anesthesia: Detection-theory measurement of sensitivity to painful stimuli. *Perception & Psychophysics*, **55**, 125-132.
- IRWIN, R. J., HAUTUS, M. J., & STILLMAN, J. A. (1992). Use of the receiver operating characteristic in taste perception. *Journal of Sensory Studies*, **7**, 291-314.
- IRWIN, R. J., & WHITEHEAD, P. R. (1991). Toward an objective psychophysics of pain. *Psychological Science*, **2**, 230-235.
- JARAMILLO, F., & WIESENFELD, K. (1998). Mechanoelectrical transduction assisted by Brownian motion: A role for noise in the auditory system. *Nature Neuroscience*, **1**, 384-388.
- KUZE, J. (1995). The effect of tone frequency on successive comparison of brightness. *Psychologia*, **38**, 50-57.
- LEVIN, J. E., & MILLER, J. P. (1996). Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature*, **380**, 165-168.
- LONGTIN, A., BULSARA, A. R., & MOSS, M. (1991). Time-interval sequences in bistable systems and the noise-induced transmission of information by sensory neurons. *Physical Review Letters*, **67**, 656-659.
- MACMILLAN, N. A., & KAPLAN, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, **98**, 185-199.
- MADDOX, J. (1994). Bringing more order out of noisiness. *Nature*, **369**, 271.
- MARKS, E. L. (1993). Contextual processing of multidimensional and unidimensional auditory stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 227-249.
- MARKS, E. L. (1996). Recalibrating the perception of loudness: Interaural transfer. *Journal of the Acoustical Society of America*, **100**, 473-480.
- MARKS, E. L. (1997). Loudness perception: Sensory, decisional, and cognitive processes. In A. Preis & T. Hornowski (Eds.) *Fechner Day 97: Proceedings of the Thirteenth Annual Meeting of the International Society for Psychophysics* (pp. 79-84). Poznań: ISP.
- MCMANARA, B., & WIESENFELD, K. (1989). Theory of stochastic resonance. *Physical Review A*, **39**, 4854-4868.
- MORSE, R. P., & EVANS, E. F. (1996). Enhancement of vowel encoding for cochlear implants by the addition of noise. *Nature Medicine*, **2**, 928-932.
- MOSS, F., WARD, L. M., & SANNITA, W. G. (2004). Stochastic resonance and sensory information processing: A tutorial and review of application. *Clinical Neurophysiology*, **115**, 267-281.
- ODGAARD, E. C., ARIEH, Y., & MARKS, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, **65**, 123-132.
- PARNAS, B. R. (1996). Noise and neuronal populations conspire to encode simple waveforms reliably. *Transactions on Biomedical Engineering*, **43**, 313-317.
- PFÄFFLIN, S. M., & MATHEWS, M. V. (1962). Energy-detection model for monaural auditory detection. *Journal of the Acoustical Society of America*, **34**, 1842-1853.

- RANKIN, K. M., & MARKS, L. E. (1991). Differential context effects in taste perception. *Chemical Senses*, **16**, 617-629.
- RIANI, M., & SIMONOTTO, E. (1994). Network simulations of the Necker cube. *Physical Review*, **72**, 3120-3123.
- RIES, D. T. (2007). The influence of noise type and level upon stochastic resonance in human audition. *Hearing Research*, **228**, 136-143.
- SCHARF, B. (1964). Partial masking. *Acoustica*, **14**, 16-23.
- SIMONOTTO, E., RIANI, M., SEIFE, C., ROBERTS, M., TWITTY, J., & MOSS, M. (1997). Visual perception of stochastic resonance. *Physical Review Letters*, **78**, 1186-1189.
- TOUGAARD, J. (2000). Stochastic resonance and signal detection in an energy detector—Implications for biological receptor systems. *Biological Cybernetics*, **83**, 471-480.
- WANG, M.-F., IRWIN, R. J., & HAUTUS, M. J. (1998). Discriminability in length of lines in the Müller-Lyer figure. *Perception & Psychophysics*, **60**, 511-517.
- WARD, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 444-459.
- WARD, L. M. (2004). Psychophysics of stochastic resonance. *Fluctuation & Noise Letters*, **4**, L11-L21.
- WARD, L. M., DESAI, S., ROOTMAN, D., TATA, M., & MOSS, F. (2001). Noise can help as well as hinder seeing and hearing. *Bulletin of the American Physical Society*, **46**, N23.002.
- WARD, L. M., NEIMAN, A., & MOSS, F. (2002). Stochastic resonance in psychophysics and in animal behavior. *Biological Cybernetics*, **87**, 91-101.
- WIESENFIELD, K., & MOSS, F. (1995). Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDS. *Nature*, **373**, 33-36.
- ZENG, F., FU, Q., & MORSE, R. (2000). Human hearing enhanced by noise. *Brain Research*, **869**, 251-255.

NOTES

1. We are indebted to Neil Macmillan for drawing our attention to this alternative form of analysis.
2. Not having noise in the second interval would indicate that the reference stimulus was unique and could, potentially, bias responding.

(Manuscript received July 2, 2008;
revision accepted for publication June 8, 2009.)