

On the choice of adequate randomization ranges for limiting the use of unwanted cues in *same–different*, dual-pair, and oddity tasks

HUANPING DAI

University of Arizona, Tucson, Arizona

AND

CHRISTOPHE MICHEYL

University of Minnesota, Minneapolis, Minnesota

A major concern when designing a psychophysical experiment is that participants may use a stimulus feature (cue) other than that intended by the experimenter. One way to avoid this problem is to apply random variations to the corresponding feature across stimulus presentations to make the unwanted cue unreliable. An important question facing experimenters who use this randomization (roving) technique is how large the randomization range should be to ensure that the participants cannot achieve a certain proportion correct by using the unwanted cue, while at the same time avoiding unnecessary interference of the randomization with task performance. Previous researchers have provided formulas for the selection of adequate randomization ranges in *yes–no* and multiple-alternative forced choice tasks. In this article, we provide figures and tables that can be used to select randomization ranges that are better suited to experiments involving a *same–different*, dual-pair, or oddity task.

A common concern in designing a psychophysical experiment is the possibility that participants may perform the task on the basis of a cue (stimulus feature or dimension) other than that intended by the experimenter. For instance, in the field of auditory perception research known as *profile analysis* (for a review, see Green, 1988), the researcher is primarily interested in how well listeners can detect or discriminate features, such as peaks or troughs, in the spectral shape of sounds. However, unless special precautions are taken, the listeners may be able to perform the task correctly without even extracting spectral shape. For example, listeners can identify which of two successively presented sounds contains a spectral peak solely on the basis of differences in loudness if the sound containing the spectral peak has a higher intensity overall. Consequently, there is a risk that thresholds or performance in this type of experiment reflect loudness perception rather than spectral-shape perception. Another example is auditory frequency (subjectively, pitch) discrimination, where listeners can use differences in loudness between tones because of variation in equal-loudness contours across frequency (Dai, Nguyen, & Green, 1995; Emmerich, Ellermeier, & Butensky, 1989; Henning, 1966; Moore & Glasberg, 1989; Moore, Glasberg, Low, Cope, & Cope, 2006). As a result, performance or thresholds in an experiment that was originally designed to measure the perception of pitch may actually reflect or be contaminated by the perception of another sound attribute (loudness).

Two approaches have traditionally been used by experimenters to limit participants' ability to take advantage of unwanted cues in discrimination tasks. The first approach involves equalizing the values of the stimuli along the unwanted dimension. For instance, in a frequency-discrimination experiment, the experimenter can try to adjust the relative intensities of tones in a frequency-dependent manner in an attempt to ensure that loudness remains constant as frequency changes. Unfortunately, in general, precisely equating the perceived values of stimuli, which keep changing during the course of psychophysical measurements, is a challenging task that often requires detailed and time-consuming measurements beforehand.¹

A second approach to limiting participants' use of unwanted cues is to apply random stimulus variation along the unwanted dimension. In the auditory psychophysics literature, this is commonly referred to as *randomization* or *roving*. For instance, to prevent the listeners in the above-mentioned spectral-shape discrimination experiment from taking advantage of overall loudness cues, experimenters rove (i.e., vary randomly) the overall level of the stimuli (see Dai & Green, 1992; Drennan & Watson, 2001; Durlach, Braida, & Ito, 1986; Farrar et al., 1987; Green, 1988; Kidd & Dai, 1993; Kidd, Mason, Uchanski, Brantley, & Shah, 1991; Mason, Kidd, Hanna, & Green, 1984; Spiegel, Picardi, & Green, 1981; Versfeld & Houtsma, 1991). Similarly, to prevent listeners from taking advantage of loudness cues in an auditory frequency-discrimination

H. Dai, hdai@email.arizona.edu

experiment, experimenters can randomize the level of each tone, so that loudness differences no longer provide a reliable cue for task performance (Dai et al., 1995; Emmerich et al., 1989; Henning, 1966; Moore & Glasberg, 1989; Semal & Demany, 2006). The roving technique is very popular among auditory-perception researchers. It has been used in studies of intensity perception (Berliner & Durlach, 1973; Berliner, Durlach, & Braida, 1977; Oxenham & Buus, 2000), pitch discrimination with complex tones (see Houtsma & Smurzynski, 1990; Moore, Glasberg, Flanagan, & Adams, 2006; Oxenham, Micheyl, & Keebler, 2009), tone-in-noise detection (Hall & Fernandes, 1983; Kidd, Mason, Brantley, & Owen, 1989), binaural hearing (Bernstein & Trahiotis, 1994, 1997; Henning, Richards, & Lentz, 2005), speech perception (Macmillan, Goldberg, & Braida, 1988), temporal gap detection (Formby & Muir, 1989; Forrest & Green, 1987), and frequency- or amplitude-modulation perception (Furukawa & Moore, 1997; Moore & Sek, 1998; Stellmack, Viemeister, & Byrne, 2006), among others.

A practical question facing the experimenter who plans to use roving is how large the roving range should be. If the range is too small, participants may still be able to achieve a relatively high proportion of correct responses on the basis of the unwanted cue. On the other hand, if the range is too large, the participants' performance might be unnecessarily impacted by the random stimulus variations.² Therefore, experimenters must strive to find a good compromise between limiting contributions from the unwanted cue (which encourages the use of a wide roving range) and limiting potential side effects of roving on performance (which calls for the use of as small a roving range as is safely possible). In order to select a suitable roving range, the experimenters must know how the proportion of correct responses that can be achieved on the basis of the unwanted cue, $P_{c_{\text{unwanted}}}$, depends on the roving range, R . The latter is defined as the distance between the largest and smallest values that the stimulus can assume along the unwanted dimension because of roving. For instance, in a frequency-discrimination experiment in which the level of tones can vary randomly between 45 and 55 dB SPL across presentations, the roving range is 10 dB.

In addition to depending on R , $P_{c_{\text{unwanted}}}$ also depends on the size of the unwanted cue, Δ . The latter corresponds to the change along the unwanted stimulus dimension, which accompanies (and is correlated with) the change applied by the experimenter along the primary dimension. For instance, if the loudness of a tone changes by an amount corresponding to 1 dB when the tone frequency changes by 1%, the size of the unwanted loudness cue corresponding to a 1% change in frequency in a frequency-discrimination experiment is 1 dB. In the framework of signal detection theory (see Green & Swets, 1966), Δ can be identified with the distance along the unwanted physical dimension between the two stimuli that must be discriminated in a *yes–no* paradigm. If the physical-to-sensory mapping is linear, and the internal noise that contaminates the sensory observations evoked by the stimuli

is constant, Δ is directly proportional to the familiar index of sensitivity, d' . However, there are two important differences between Δ and d' . First, whereas d' usually denotes sensitivity to the primary cue, Δ refers to an *unwanted* cue. Second, whereas d' is dimensionless, Δ has the dimension of the stimulus attribute being randomized.

In most applications, the size of the unwanted cue either is known to the experimenter or can be estimated on the basis of data in the relevant literature—especially data from studies in which the corresponding cue, which is now the unwanted cue, was then the cue of primary interest. For instance, loudness cues associated with changes in the frequency of pure tones in a frequency-discrimination experiment can be estimated on the basis of data on equal-loudness contours and intensity discrimination. When relevant data for estimating the size of the unwanted cue are not available in the existing literature, such data must be collected. In some applications, the size of the unwanted cue is directly available. For instance, in spectral-profile analysis experiments, the overall loudness difference between the sounds that the listener must discriminate (i.e., the size of the unwanted cue) is, to a first approximation, proportional to the increment or decrement in level that the listener must detect (i.e., the size of the primary cue). In general, when the primary and unwanted cues share the same dimension—as in the profile analysis example—the size of the unwanted cue is known to the experimenter; in all other cases, the size must be estimated on the basis of existing data or measured.

Although the relationship among $P_{c_{\text{unwanted}}}$, R , and Δ can be studied empirically, measurements of this relationship are usually impractical in the context of experimental studies, the primary aim of which is not to characterize it. Therefore, experimenters do not usually choose R on the basis of empirical measurements. Instead, they rely on predictions derived from ideal-observer models from signal detection theory (Green & Swets, 1966). Because these models assume a noiseless observer who uses the information conveyed by the unwanted cue optimally, they provide an upper bound on the performance that can be achieved on the basis of that cue. In particular, Green (1988) provided a relatively simple formula relating $P_{c_{\text{unwanted}}}$, R , and Δ , in the two-interval, two-alternative forced choice (2I–2AFC) paradigm: $P_{c_{\text{unwanted}}} = .5 + \Delta/R - .5(\Delta/R)^2$. More recently, Dai and Kidd (2009) derived similar formulas for the *yes–no* and *m*-alternative forced choice (*m*AFC) paradigms. Specifically, they showed that, for the *yes–no* paradigm, $P_{c_{\text{unwanted}}} = .5 + .5(\Delta/R)$, whereas for the *m*AFC paradigm, $P_{c_{\text{unwanted}}} = \Delta/R + [1 - (\Delta/R)^m]/m$.

Although the *yes–no* and *m*AFC paradigms have been used in a large number of auditory- and visual-perception studies over the past 50 years, other paradigms are better suited for certain applications (overviews can be found in Creelman & Macmillan, 1979; Macmillan & Creelman, 2005; Macmillan, Kaplan, & Creelman, 1977; Noreen, 1981). For instance, the *same–different* paradigm provides a measure of basic stimulus discriminability, which does not involve an ability to identify the direction of

changes in, for example, sound intensity or frequency (Dai, Versfeld, & Green, 1996). The oddity paradigm, wherein participants are presented on each trial with m stimuli, one of which differs from the other $m-1$, is better suited than its m AFC counterpart in some situations (Versfeld, Dai, & Green, 1996). The dual-pair comparison paradigm (Creelman & Macmillan, 1979) allows researchers to measure stimulus-change detection and change-direction identification using the same stimulus structure (two pairs of stimuli, one containing a change, the other not) by simply changing the instructions given to the participant (Micheyl, Kaernbach, & Demany, 2008; Semal & Demany, 2006). In some contexts, it is necessary to use roving in a *same-different* (e.g., Jesteadt & Bilger, 1974), oddity (e.g., Lyzenga & Horst, 1995, 1997, 1998),³ or dual-pair (e.g., Micheyl & Messing, 2006; Semal & Demany, 2006) paradigm. Unfortunately, the above-cited formulas, which give the relationship between $P_{c_{\text{unwanted}}}$ and roving range for the *yes-no* and m AFC paradigms, do not apply to these other paradigms. In fact, as the results presented in this article reveal, using these formulas to determine the roving range required to keep $P_{c_{\text{unwanted}}}$ under a target level in any of the three paradigms mentioned above (*same-different*, dual-pair, and oddity) can lead to substantial errors in both experimental design and data interpretation.

Although no simple analytical formulas exist that can be used as guidelines for selecting suitable roving ranges in *same-different*, dual-pair, or oddity experiments, in this article, we provide figures and tables that experimenters can use to select an adequate roving range, R , given a target $P_{c_{\text{unwanted}}}$ and a known (or estimated) unwanted-cue size, Δ , in the *same-different* paradigm, two versions of the dual-pair paradigm (i.e., 4IAX and AB-versus-BA), and two versions of the oddity paradigm (i.e., three- and four-interval oddity). The information in the tables and figures can also be used, conversely, to find the $P_{c_{\text{unwanted}}}$ that can be (or could have been, in retrospect) achieved in an experiment using one of these paradigms, given the roving range and the unwanted-cue size.

METHOD

In order to derive the results presented below, we assumed a maximum-likelihood observer, who would make optimal use of the information conveyed by the unwanted cue that is being roved. Obviously, the information conveyed by the unwanted cue becomes less and less useful for correct task performance as the roving range increases. The general approach is similar to that described by Green (1988) for the 2I-2AFC task and more recently extended to *yes-no* and m AFC tasks by Dai and Kidd (2009). The basic idea of this approach is that the unwanted cue shifts the distribution of stimulus values, and the corresponding distribution of sensory observations, along the considered stimulus dimension. The distribution of stimulus values is produced by the application of stimulus roving. Here, as in Green (1988) and in Dai and Kidd (2009), we assume a rectangular—that is, continuous uniform—distribution. Near the end of the article, we show how the results can be corrected when the uniform distribution is discrete, instead of continuous. The uniform distribution is the most frequently used in studies of auditory perception. Of all continuous roving distributions having a fixed range, the uniform distribution is the one that minimizes the maximal proportion cor-

rect (P_c) that can be achieved (by an ideal, maximum-likelihood observer) on the basis of the unwanted cue (Dai, 2008).

Under these assumptions, the maximal P_c that can be achieved on the basis of the unwanted cue alone ($P_{c_{\text{unwanted}}}$) can be computed as the integral, over the observation space, of the probability density corresponding to the most likely a posteriori stimulus alternative at the current point in the observation space. Using Bayes's theorem, the latter probability can be determined on the basis of the (uniform) probability density function of the observations, given the stimulus alternative. Since our calculations are for an ideal observer—and in most experimental applications, the various stimulus alternatives are equally likely—the maximum a posteriori and maximum-likelihood solutions are equivalent.

For the 2I-2AFC, *yes-no*, and m AFC paradigms, the integral has a relatively simple analytical solution (see the above-mentioned equations by Green [1988] and Dai & Kidd [2009]). For other paradigms, analytical solutions are more difficult to obtain because of the greater dimensionality of the decision space or because of the presence of nonlinearities (e.g., an absolute-value or maximum-of operation) in the decision rule. Here, rather than attempt to provide analytical solutions, we resorted to a numerical-evaluation approach. We evaluated the integral, over the relevant observation space, of the (uniform) probability density corresponding to the most likely stimulus alternative. Several publications have already described the relevant observation spaces and maximum-likelihood decision rules for the various paradigms considered here: *same-different* (Dai et al., 1996; Irwin & Hautus, 1997; Irwin, Hautus, & Butcher, 1999; Macmillan & Creelman, 2005), dual-pair 4IAX (Micheyl et al., 2008; Micheyl & Messing, 2006; Noreen, 1981; Rousseau & Ennis, 2001, 2002), dual-pair AB-versus-BA (Micheyl & Dai, 2008, 2009), and oddity (Frijters, 1979a, 1979b; Geelhoed, MacRae, & Ennis, 1994; Versfeld et al., 1996). In the remainder of this article, we present the results of our calculations relating $P_{c_{\text{unwanted}}}$ to Δ/R for these different paradigms both in figures and in tables.

RESULTS AND DISCUSSION

Figure 1 shows $P_{c_{\text{unwanted}}}$ as a function of Δ/R for the *same-different* paradigm, the 4IAX paradigm, and the dual-pair AB-versus-BA paradigm. Similar functions are also shown for the *yes-no* paradigm and the 2I-2AFC paradigm for comparison. As was mentioned in the introduction, for these two latter paradigms, analytical solutions for the relationship between Δ/R and $P_{c_{\text{unwanted}}}$ have been provided in other publications (Dai & Kidd, 2009; Green, 1988). The five paradigms illustrated in this figure all have the same chance-performance level, corresponding to $P_{c_{\text{unwanted}}} = .5$. Tables 1, 2, and 3 list Δ/R values corresponding to $P_{c_{\text{unwanted}}}$ between .5 and 1 (in steps of .01) for the *same-different* paradigm and the two versions of the dual-pair paradigm (i.e., 4IAX and AB-versus-BA).

Figure 1 reveals that, of the five paradigms, the *same-different* paradigm generally yields the lowest $P_{c_{\text{unwanted}}}$ (given Δ/R). On the other hand, the 2I-2AFC paradigm generally yields higher $P_{c_{\text{unwanted}}}$ values than the other paradigms considered in this figure, with the exception of the 4IAX version of the dual-pair paradigm for relatively low P_c levels (below .65). The difference between the curves for the *same-different* and 2I-2AFC paradigms is considerable. For example, whereas a roving range five times the size of the unwanted cue, which corresponds to a Δ/R ratio of .2, is needed in order to limit $P_{c_{\text{unwanted}}}$ to just under .70 in a 2I-2AFC experiment, in

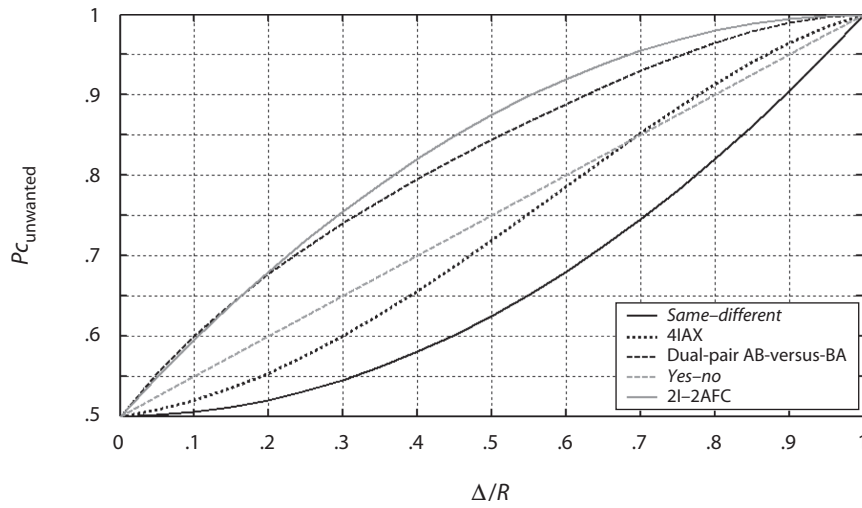


Figure 1. Pc_{unwanted} as a function of Δ/R for the *same-different*, 4IAX, dual-pair AB-versus-BA, *yes-no*, and 2I-2AFC paradigms. Each line type corresponds to a single paradigm, as indicated in the inset legend. Pc_{unwanted} refers to the maximal proportion of correct responses that can be achieved on the basis of an unwanted cue, the use of which is limited by the application of a random rove drawn for a uniform (or rectangular) distribution with a range (or width) of R . Δ denotes the size of the unwanted cue, as it is defined in the text.

Table 1
 Δ/R As a Function of Pc_{unwanted} for the *Same-Different* Paradigm

Pc_{unwanted}	Δ/R	Pc_{unwanted}	Δ/R	Pc_{unwanted}	Δ/R
.50	.00	.67	.58	.84	.82
.51	.14	.68	.60	.85	.84
.52	.20	.69	.62	.86	.85
.53	.24	.70	.63	.87	.86
.54	.28	.71	.65	.88	.87
.55	.32	.72	.66	.89	.88
.56	.35	.73	.68	.90	.89
.57	.37	.74	.69	.91	.91
.58	.40	.75	.71	.92	.92
.59	.42	.76	.72	.93	.93
.60	.45	.77	.73	.94	.94
.61	.47	.78	.75	.95	.95
.62	.49	.79	.76	.96	.96
.63	.51	.80	.77	.97	.97
.64	.53	.81	.79	.98	.98
.65	.55	.82	.80	.99	.99
.66	.57	.83	.81	1.00	1.00

Table 2
 Δ/R As a Function of Pc_{unwanted} for the Dual-Pair 4IAX Paradigm

Pc_{unwanted}	Δ/R	Pc_{unwanted}	Δ/R	Pc_{unwanted}	Δ/R
.50	.00	.67	.42	.84	.68
.51	.06	.68	.44	.85	.70
.52	.10	.69	.46	.86	.71
.53	.13	.70	.47	.87	.73
.54	.16	.71	.49	.88	.74
.55	.19	.72	.50	.89	.76
.56	.22	.73	.52	.90	.78
.57	.24	.74	.53	.91	.79
.58	.26	.75	.55	.92	.81
.59	.28	.76	.56	.93	.83
.60	.30	.77	.58	.94	.85
.61	.32	.78	.59	.95	.87
.62	.34	.79	.61	.96	.89
.63	.36	.80	.62	.97	.91
.64	.37	.81	.64	.98	.94
.65	.39	.82	.65	.99	.97
.66	.41	.83	.67	1.00	1.00

Table 3
 Δ/R As a Function of $P_{c_{unwanted}}$ for the
 Dual-Pair AB-Versus-BA Paradigm

$P_{c_{unwanted}}$	Δ/R	$P_{c_{unwanted}}$	Δ/R	$P_{c_{unwanted}}$	Δ/R
.50	.00	.67	.19	.84	.49
.51	.01	.68	.20	.85	.51
.52	.02	.69	.22	.86	.54
.53	.03	.70	.23	.87	.56
.54	.04	.71	.25	.88	.58
.55	.05	.72	.27	.89	.60
.56	.06	.73	.28	.90	.63
.57	.07	.74	.30	.91	.65
.58	.08	.75	.32	.92	.68
.59	.09	.76	.33	.93	.70
.60	.10	.77	.35	.94	.73
.61	.11	.78	.37	.95	.76
.62	.12	.79	.39	.96	.79
.63	.14	.80	.41	.97	.82
.64	.15	.81	.43	.98	.85
.65	.16	.82	.45	.99	.90
.66	.18	.83	.47	1.00	1.00

a *same-different* experiment, a roving range having this relative size limits $P_{c_{unwanted}}$ to less than .55. Another way of looking at the difference between the 2I-2AFC and *same-different* paradigms is that, for a given Δ , the smallest roving range required to ensure that $P_{c_{unwanted}}$ does not exceed .60 is about four times smaller for the *same-different* paradigm than for the 2I-2AFC paradigm. An experimenter who uses Green's (1988) formula (reproduced in the introduction) to determine the roving range needed to limit $P_{c_{unwanted}}$ to within .52-.70 is likely to overestimate the required roving range by as much as six times.

These observations should not be interpreted as implying that stimulus roving always reduces the influence of an unwanted cue more effectively in the *same-different* paradigm than in the 2I-2AFC paradigm. As was discussed

in Dai (2008), the relative contributions of a primary cue and of an unwanted cue in determining the P_c measured in an experiment depend on, among other things, the relative salience of each cue and on how these cues interact. However, the ideal-observer analysis described in this article provides an upper bound on the performance that can be achieved by a real observer on the basis of the unwanted cue.

Figure 2 shows how $P_{c_{unwanted}}$ depends on Δ/R in the three- and four-interval oddity paradigms. For comparison, the functions relating $P_{c_{unwanted}}$ to Δ/R in the 3AFC and 4AFC paradigms are also shown (as gray solid and dashed lines, respectively). The latter were computed using the formula provided by Dai and Kidd (2009), which can be found in the introduction. It can be seen that, for a given value of Δ/R , the three- and four-interval oddity para-

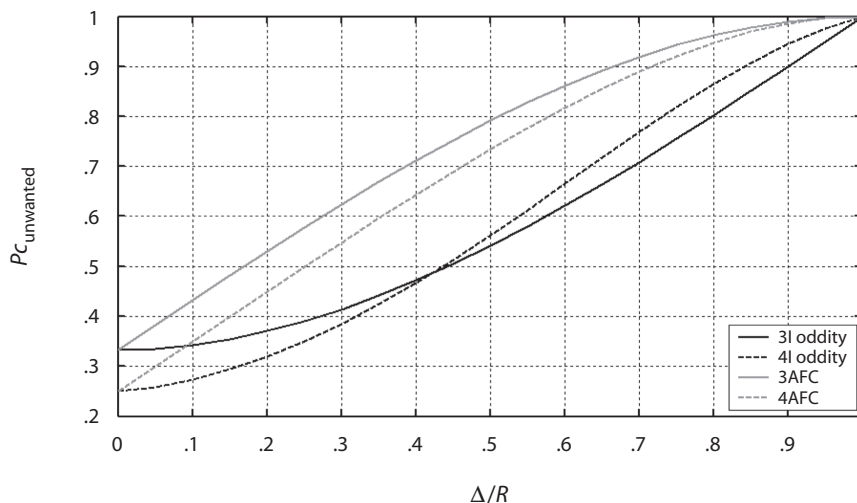


Figure 2. $P_{c_{unwanted}}$ as a function of Δ/R for the three- and four-interval oddity, 3AFC, and 4AFC paradigms. Each line type corresponds to a single paradigm, as is indicated in the inset legend. $P_{c_{unwanted}}$ refers to the maximal proportion of correct responses that can be achieved on the basis of an unwanted cue, the use of which is limited by the application of a random rove drawn for a uniform (or rectangular) distribution with a range (or width) of R . Δ denotes the size of the unwanted cue, as it is defined in the text.

Table 4
 Δ/R As a Function of $P_{c_{\text{unwanted}}}$ for the Three-Interval Oddity Paradigm

$P_{c_{\text{unwanted}}}$	Δ/R	$P_{c_{\text{unwanted}}}$	Δ/R	$P_{c_{\text{unwanted}}}$	Δ/R	$P_{c_{\text{unwanted}}}$	Δ/R
.33	.00	.50	.44	.67	.66	.84	.84
.34	.09	.51	.46	.68	.67	.85	.85
.35	.13	.52	.47	.69	.68	.86	.86
.36	.17	.53	.48	.70	.69	.87	.87
.37	.20	.54	.50	.71	.70	.88	.88
.38	.22	.55	.51	.72	.71	.89	.89
.39	.25	.56	.52	.73	.72	.90	.90
.40	.27	.57	.54	.74	.73	.91	.91
.41	.29	.58	.55	.75	.74	.92	.92
.42	.31	.59	.56	.76	.76	.93	.93
.43	.33	.60	.57	.77	.77	.94	.94
.44	.35	.61	.59	.78	.78	.95	.95
.45	.36	.62	.60	.79	.79	.96	.96
.46	.38	.63	.61	.80	.80	.97	.97
.47	.40	.64	.62	.81	.81	.98	.98
.48	.41	.65	.63	.82	.82	.99	.99
.49	.43	.66	.65	.83	.83	1.00	1.00

Table 5
 Δ/R As a Function of $P_{c_{\text{unwanted}}}$ for the Four-Interval Oddity Paradigm

$P_{c_{\text{unwanted}}}$	Δ/R	$P_{c_{\text{unwanted}}}$	Δ/R	$P_{c_{\text{unwanted}}}$	Δ/R	$P_{c_{\text{unwanted}}}$	Δ/R
.25	.00	.44	.37	.63	.57	.82	.75
.26	.05	.45	.38	.64	.58	.83	.76
.27	.09	.46	.39	.65	.59	.84	.77
.28	.12	.47	.40	.66	.60	.85	.78
.29	.14	.48	.41	.67	.60	.86	.79
.30	.16	.49	.43	.68	.61	.87	.81
.31	.18	.50	.44	.69	.62	.88	.82
.32	.20	.51	.45	.70	.63	.89	.83
.33	.22	.52	.46	.71	.64	.90	.84
.34	.24	.53	.47	.72	.65	.91	.85
.35	.25	.54	.48	.73	.66	.92	.87
.36	.27	.55	.49	.74	.67	.93	.88
.37	.28	.56	.50	.75	.68	.94	.89
.38	.29	.57	.51	.76	.69	.95	.91
.39	.31	.58	.52	.77	.70	.96	.92
.40	.32	.59	.53	.78	.71	.97	.94
.41	.33	.60	.54	.79	.72	.98	.96
.42	.35	.61	.55	.80	.73	.99	.98
.43	.36	.62	.56	.81	.74	1.00	1.00

digms yield lower values of $P_{c_{\text{unwanted}}}$ than their 3AFC and 4AFC counterparts.

Tables 4 and 5 list Δ/R values corresponding to $P_{c_{\text{unwanted}}}$ values between .33 or .25 and 1 (in steps of .01) for the three- and four-interval oddity paradigms, respectively.

On the Use of Discrete Uniform Roving Distributions With Few Bins

The results in Figures 1 and 2 and Tables 1–5 apply to continuous uniform roving distributions. However, in experimental studies, researchers sometimes use discrete uniform distributions with a relatively small number of levels (bins) on the roving continuum. For instance, Henning (1966) used a 10-dB roving range with levels spaced 0.5 dB apart, yielding 21 possible stimulus intensities. In Jesteadt and Bilger (1974), the uniform discrete distributions used for frequency and level roving contained five bins. In pitch-discrimination experiments with complex tones in which the lowest harmonic number had been roved, the roving distribution typically contained only two or three

bins (see Houtsma & Smurzynski, 1990; Moore, Glasberg, Flanagan, & Adams, 2006; Oxenham et al., 2009). Therefore, it is of interest to determine how the number of bins in a discrete uniform roving distribution influences the relationship between $P_{c_{\text{unwanted}}}$ and Δ/R . Provided that the bins of the roving distributions for the standard and signal stimuli coincide with each other within the region in which the two distributions overlap, the results for discrete uniform roving distributions with n bins can be derived from the results obtained using continuous uniform distributions by replacing Δ/R with $[(n - 1)/n]\Delta/R$ (Dai & Kidd, 2009). As an example, suppose that an experimenter desires to predict $P_{c_{\text{unwanted}}}$ in a *same–different* task for a uniform discrete distribution having a range of 2Δ and containing three bins. The result can be obtained first by calculating $[(n - 1)/n]\Delta/R$, then looking for the closest Δ/R value in Table 1, and finding the corresponding $P_{c_{\text{unwanted}}}$. In our example, $[(n - 1)/n]\Delta/R = .3333$, and the $P_{c_{\text{unwanted}}}$ corresponding to the closest Δ/R value (.3463) in Table 1 is .56. With a continuous roving distribution or, approximately, a

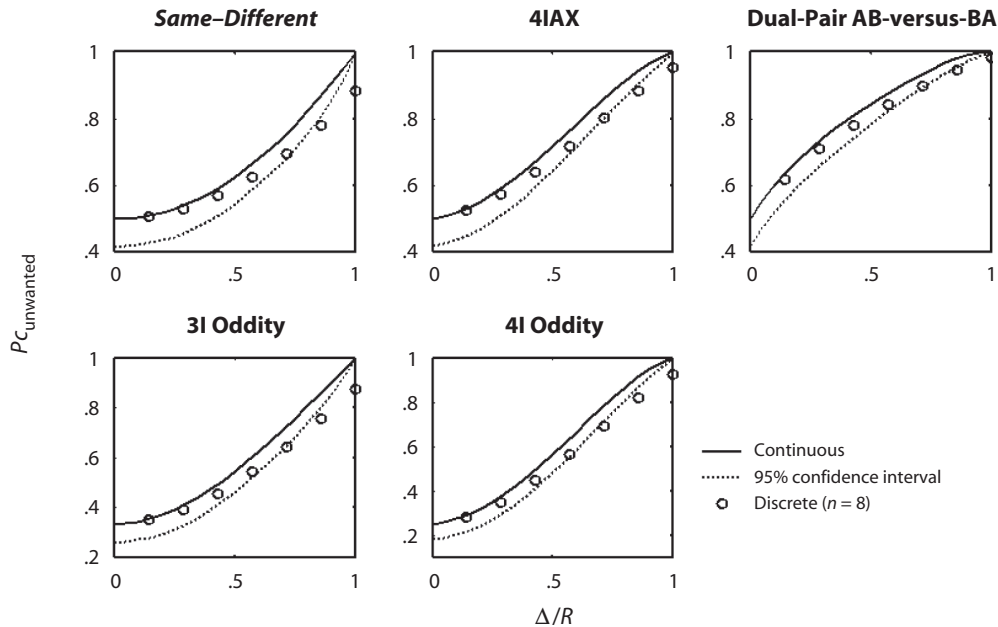


Figure 3. $P_{c_{unwanted}}$ produced by a continuous roving distribution (solid dark line) and a discrete distribution with eight bins (circles) as a function of Δ/R for various psychophysical paradigms (represented by the different panels). The lower bound of the 95% confidence interval around $P_{c_{unwanted}}$ values obtained using a continuous roving distribution is shown as a dotted line. The 95% confidence interval was determined on the basis of the assumption of binomial variability, with the parameter N (number of trials) of the binomial distribution set to 100.

discrete distribution containing a large number of bins, the same shift of $\Delta/R = .5$ would yield a $P_{c_{unwanted}}$ of about .63. Thus, for the same roving range, performance based on unwanted cues can be limited to a lower level by using a uniform discrete roving distribution with a relatively small number of bins, compared with a discrete distribution with a larger number of bins.

Finally, another question of practical interest to experimenters concerns the smallest number of bins that a discrete roving distribution should have in order for its effect on $P_{c_{unwanted}}$ to be essentially indistinguishable from that achieved with a continuous roving distribution. To answer this question, we computed the lower bound of the 95% confidence interval around the $P_{c_{unwanted}}$ values shown in Figures 1 and 2, which were derived using a continuous distribution. The confidence intervals were determined under the assumption of binomial variability (i.e., no overdispersion) and measures based on 100 trials. $P_{c_{unwanted}}$ values corresponding to discrete distributions with different numbers of bins were then computed, using the approach described in the previous paragraph, and these values were compared with the lower bound of the 95% confidence interval. Figure 3 shows for each paradigm (in the different panels) the lower bound of the 95% confidence interval (dotted line) for the $P_{c_{unwanted}}$ function derived from a continuous distribution (solid line, replotted from either Figure 1 or Figure 2) and $P_{c_{unwanted}}$ values derived from a discrete distribution with eight bins (open circles). The results are similar across all paradigms, showing that for Δ/R values of less than .5 (i.e., a roving range at least twice as large as the assumed size of the unwanted

cue, which is typically the case in experimental studies), the $P_{c_{unwanted}}$ values from the discrete distribution (open circles) fall within the 95% confidence interval and are thus statistically indistinguishable from that achieved using a continuous roving distribution. Therefore, in limiting the effectiveness of an unwanted cue via random roving, a discrete uniform distribution is practically identical to a continuous uniform distribution, provided that the discrete distribution consists of eight or more bins. This result provides a guideline for experimenters.

APPLICATION EXAMPLES

In this section, we provide two examples to illustrate how the information provided in this article can be used. The first example illustrates whether the roving range in an experiment using the dual-pair paradigm was large enough to warrant ruling out the possibility that the measured discrimination thresholds were based on an unwanted cue. The second example illustrates how using Green's (1988) formula (which was explicitly derived for the 2I-2AFC paradigm) when designing a *same-different* experiment can result in substantial overestimation of the roving range needed.

Retrospective Analysis of Published Data: Was the Roving Range Large Enough?

The first example comes from a recent study of pitch perception by Semal and Demany (2006). In this study, Semal and Demany used both the 4IAX and the AB-versus-BA versions of the dual-pair paradigm in order to

measure thresholds for the detection of frequency changes (4IAX) and thresholds for the identification of the direction of frequency changes (AB-versus-BA) between pure tones in the same listeners. In one of the experiments in this study, they sought to test the hypothesis that listeners' performance in these two tasks was based on level changes at the output of a single auditory channel (for the details of this explanation, see pp. 3910–3911 in Semal & Demany, 2006; see also Emmerich et al., 1989; Henning, 1966; Moore & Glasberg, 1989). The level of each tone was roved over a 10-dB range (± 5 dB) around the nominal level (65 dB SPL). Semal and Demany reasoned that such roving would lead to an increase in thresholds if the listeners' performance was based on changes in level at the output of a single auditory channel.

The question that we ask here is whether the 10-dB roving range used by Semal and Demany (2006) was sufficient to ensure that listeners could not reliably achieve 75% correct responses—the percentage-correct level targeted by the adaptive threshold-tracking procedure—in the pitch-change detection and pitch-change direction identification tasks on the basis of level changes at the output of an auditory channel. Taking into account both the nominal level of the tones (65 dB SPL) and the mean thresholds that were measured without roving the level in this experiment (about 29 cents, slightly less than 2%), the average size of excitation-level differences at the output of auditory filters in Semal and Demany's experiment can be estimated between 2 and 3 dB on average.⁴ To be on the safe side, we set $\Delta = 3$ dB. First, we consider the change-detection task, which corresponds to the 4IAX dual-pair paradigm. Table 2 indicates that, for this paradigm, the value of $P_{c_{\text{unwanted}}}$ corresponding to $\Delta/R = .3$ (i.e., 3 dB/10 dB) is .60. This is well below the targeted level of .75. Therefore, for the dual-pair change-detection (4IAX) task, we can confidently rule out the possibility that the listeners' performance was based on loudness cues alone.

Next, we consider the direction-identification task, which corresponds to the dual-pair AB-versus-BA paradigm. Table 3 reveals that, for this paradigm, the same value of Δ/R (.3) yields a substantially higher $P_{c_{\text{unwanted}}}$ (.74), which is practically indistinguishable from the targeted proportion correct level of .75. Since thresholds were measured using an adaptive procedure that visited different points (both below and above the targeted proportion correct of .75) on the psychometric function, one cannot rule out the possibility that level cues had some influence on the threshold measurements, even with roving.

This outcome illustrates the important point that when the same roving range is used in different experiments that involve superficially similar stimulus designs but different underlying psychophysical paradigms, the predicted influence of roving on the proportion of correct responses can be substantially different across experiments.

On the Importance of Using the Correct Paradigm-Specific Formulas or Tables When Selecting a Roving Range

Green's (1988) formula, which is reproduced in the introduction, is frequently used by auditory psychophysicists

to select appropriate roving ranges in their experiments. However, as was mentioned above, this formula was derived specifically with the 2I–2AFC paradigm in mind. If the formula is applied in the context of experiments in which a different paradigm is used, it can lead to the selection of unnecessarily large roving ranges. For example, consider an experimenter who is designing a spectral-shape discrimination experiment with a *same–different* task. To prevent listeners from performing the task reliably on the basis of simple loudness cues, the experimenter will rove the overall level of each complex. Suppose that the experimenter determines using Green's formula that a roving range of 30 dB is required to limit $P_{c_{\text{unwanted}}}$ to .55 at most. The information in Figure 1 and in Table 1 reveals that for the *same–different* paradigm, in fact, a roving range of merely 5 dB is sufficient, in principle, to limit $P_{c_{\text{unwanted}}}$ to .55. Therefore, in this example, using Green's formula would lead the experimenter to use a roving range about six times as large as that needed to achieve the objective. Outcomes such as this one should matter to experimenters. Unless the sensory dimensions involved are completely independent perceptually, random variations along the unwanted dimension may have detrimental effects on the processing of the primary cue, even if the unwanted cue does not provide any useful information for task performance (Ashby & Townsend, 1986; Garner, 1974). Therefore, experimenters should avoid using unnecessarily large ranges of rove, while ensuring that performance is safely below the level targeted in an experiment. The figures and tables in this article should help them achieve this objective.

SUMMARY

Stimulus randomization, or roving, is a technique commonly used to limit the use of unwanted cues by participants in psychophysical experiments. A practical question for experimenters who use the technique is how large the roving range should be. Previous publications have provided equations for selecting adequate roving ranges in the *m*AFC and *yes–no* paradigms (Dai & Kidd, 2009; Green, 1988). In the present article, these analyses were extended to several other psychophysical paradigms, including the *same–different* paradigm, two versions of the dual-pair paradigm (4IAX and AB-versus-BA), and the three- and four-interval oddity paradigms. Use of the information given in this article is subject to the same limitations as applications based on Green's or Dai and Kidd's formulas. In particular, it requires a valid estimate or measure of the unwanted-cue size. If the size of the unwanted cue is underestimated, the minimum roving range required to limit proportion correct on the basis of the unwanted cue to a predefined level will also be underestimated. However, to the extent that the size of the unwanted cue can be measured or correctly estimated, the predictions described in this article provide an upper bound on the performance that can be achieved on the basis of the unwanted cue.

AUTHOR NOTE

This work was supported by National Institutes of Health National Institute on Deafness and Other Communication Disorders Grant R01 DC

05216, as well as by the University of Arizona. The authors are grateful to D. Creelman, S. Grondin, and two anonymous reviewers for constructive comments, which greatly helped improve the manuscript. Correspondence concerning this article should be addressed to H. Dai, Department of Speech, Language, and Hearing Sciences, University of Arizona, 1131 E. 2nd Street, Tucson, AZ 85721 (e-mail: hdai@email.arizona.edu).

REFERENCES

- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- BERLINER, J. E., & DURLACH, N. I. (1973). Intensity perception. IV. Resolution in roving-level discrimination. *Journal of the Acoustical Society of America*, **53**, 1270-1287.
- BERLINER, J. E., DURLACH, N. I., & BRAIDA, L. D. (1977). Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects. *Journal of the Acoustical Society of America*, **61**, 1577-1585.
- BERNSTEIN, L. R., & TRAHOTIS, C. (1994). The effect of nonsimultaneous on-frequency and off-frequency cues on the detection of a tonal signal masked by narrow-band noise. *Journal of the Acoustical Society of America*, **95**, 920-930.
- BERNSTEIN, L. R., & TRAHOTIS, C. (1997). The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation. *Journal of the Acoustical Society of America*, **102**, 1113-1120.
- CREELMAN, C. D., & MACMILLAN, N. A. (1979). Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 146-156.
- DAI, H. (2008). On suppressing unwanted cues via randomization. *Perception & Psychophysics*, **70**, 1379-1382.
- DAI, H., & GREEN, D. M. (1992). Auditory intensity perception: Successive versus simultaneous, across-channel discriminations. *Journal of the Acoustical Society of America*, **91**, 2845-2854.
- DAI, H., & KIDD, G., JR. (2009). Limiting unwanted cues via random rove applied to the yes-no and multiple-alternative forced choice paradigms. *Journal of the Acoustical Society of America*, **126**, 62-67.
- DAI, H., NGUYEN, Q. T., & GREEN, D. M. (1995). A two-filter model for frequency discrimination. *Hearing Research*, **85**, 109-114.
- DAI, H., VERSFELD, N. J., & GREEN, D. M. (1996). The optimum decision rules in the same-different paradigm. *Perception & Psychophysics*, **58**, 1-9.
- DRENNAN, W. R., & WATSON, C. S. (2001). Sources of variation in profile analysis. II. Component spacing, dynamic changes, and roving level. *Journal of the Acoustical Society of America*, **110**, 2498-2504.
- DURLACH, N. I., BRAIDA, L. D., & ITO, Y. (1986). Towards a model for discrimination of broadband signals. *Journal of the Acoustical Society of America*, **80**, 63-72.
- EMMERICH, D. S., ELLERMEIER, W., & BUTENSKY, B. (1989). A re-examination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model. *Journal of the Acoustical Society of America*, **85**, 1653-1659.
- FARRAR, C. L., REED, C. M., ITO, Y., DURLACH, N. I., DELHORNE, L. A., ZUREK, P. M., & BRAIDA, L. D. (1987). Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises. *Journal of the Acoustical Society of America*, **81**, 1085-1092.
- FORMBY, C., & MUIR, K. (1989). Effects of randomizing signal level and duration on temporal gap detection. *Audiology*, **28**, 250-257.
- FORREST, T. G., & GREEN, D. M. (1987). Detection of partially filled gaps in noise and the temporal modulation transfer function. *Journal of the Acoustical Society of America*, **82**, 1933-1943.
- FRIJTERS, J. E. (1979a). The paradox of discriminatory nondiscriminators resolved. *Chemical Senses*, **4**, 355-358.
- FRIJTERS, J. E. (1979b). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced-choice signal detection theory models. *British Journal of Mathematical & Statistical Psychology*, **32**, 229-241.
- FURUKAWA, S., & MOORE, B. C. (1997). Dependence of frequency modulation detection on frequency modulation coherence across carriers: Effects of modulation rate, harmonicity, and roving of the carrier frequencies. *Journal of the Acoustical Society of America*, **101**, 1632-1643.
- GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- GELHOED, E. N., MACRAE, A. W., & ENNIS, D. M. (1994). Preference gives more consistent judgments than oddity only if the task can be modeled as forced choice. *Perception & Psychophysics*, **55**, 473-477.
- GLASBERG, B. R., & MOORE, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, **47**, 103-138.
- GREEN, D. M. (1988). *Profile analysis: Auditory intensity discrimination*. New York: Oxford University Press.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Krieger.
- HALL, J. W., & FERNANDES, M. A. (1983). The effect of random intensity fluctuation on monaural and binaural detection. *Journal of the Acoustical Society of America*, **74**, 1200-1203.
- HENNING, G. B. (1966). Frequency discrimination of random amplitude tones. *Journal of the Acoustical Society of America*, **39**, 336-339.
- HENNING, G. B., RICHARDS, V. M., & LENTZ, J. J. (2005). The effect of diotic and dichotic level-randomization on the binaural masking-level difference. *Journal of the Acoustical Society of America*, **118**, 3229-3240.
- HOUTSMA, A. J. M., & SMURZYNSKI, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America*, **87**, 304-310.
- IRWIN, R. J., & HAUTUS, M. J. (1997). Likelihood-ratio decision strategy for independent observations in the same-different task: An approximation to the detection-theoretic model. *Perception & Psychophysics*, **59**, 313-316.
- IRWIN, R. J., HAUTUS, M. J., & BUTCHER, J. C. (1999). An area theorem for the same-different experiment. *Perception & Psychophysics*, **61**, 766-769.
- JESTEADT, W., & BILGER, R. C. (1974). Intensity and frequency discrimination in one- and two-interval paradigms. *Journal of the Acoustical Society of America*, **55**, 1266-1276.
- KIDD, G., JR., & DAI, H. (1993). A composite randomization procedure for measuring spectral shape discrimination. *Journal of the Acoustical Society of America*, **94**, 1275-1280.
- KIDD, G., JR., MASON, C. R., BRANTLEY, M. A., & OWEN, G. A. (1989). Roving-level tone-in-noise detection. *Journal of the Acoustical Society of America*, **86**, 1310-1317.
- KIDD, G., JR., MASON, C. R., UCHANSKI, R. M., BRANTLEY, M. A., & SHAH, P. (1991). Evaluation of simple models of auditory profile analysis using random reference spectra. *Journal of the Acoustical Society of America*, **90**, 1340-1354.
- LYZENGA, J., & HORST, J. W. (1995). Frequency discrimination of band-limited harmonic complexes related to vowel formants. *Journal of the Acoustical Society of America*, **98**, 1943-1955.
- LYZENGA, J., & HORST, J. W. (1997). Frequency discrimination of stylized synthetic vowels with a single formant. *Journal of the Acoustical Society of America*, **102**, 1755-1767.
- LYZENGA, J., & HORST, J. W. (1998). Frequency discrimination of stylized synthetic vowels with two formants. *Journal of the Acoustical Society of America*, **104**, 2956-2966.
- MACMILLAN, N. A., & CREELMAN, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- MACMILLAN, N. A., GOLDBERG, R. F., & BRAIDA, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*, **84**, 1262-1280.
- MACMILLAN, N. A., KAPLAN, H. L., & CREELMAN, D. C. (1977). The psychophysics of categorical perception. *Psychological Review*, **84**, 452-471.
- MASON, C. R., KIDD, G., JR., HANNA, T. E., & GREEN, D. M. (1984). Profile analysis and level variation. *Hearing Research*, **13**, 269-275.
- MAUERMAN, M., LONG, G. R., & KOLLMEIER, B. (2004). Fine structure of hearing threshold and loudness perception. *Journal of the Acoustical Society of America*, **116**, 1066-1080.
- MCDERMOTT, H. J., LECH, M., KORNBLUM, M. S., & IRVINE, D. R. F. (1998). Loudness perception and frequency discrimination in subjects with steeply sloping hearing loss: Possible correlates of neural plasticity. *Journal of the Acoustical Society of America*, **104**, 2314-2325.

- MICHEYL, C., & DAI, H. (2008). A general area theorem for the *same–different* paradigm. *Perception & Psychophysics*, **70**, 761–764.
- MICHEYL, C., & DAI, H. (2009). Likelihood ratio, optimal decision rules, and relationship between proportion correct and d' in the dual-pair AB-versus-BA identification paradigm. *Perception & Psychophysics*, **71**, 1426–1433.
- MICHEYL, C., KAERNBACH, C., & DEMANY, L. (2008). An evaluation of psychophysical models of auditory change perception. *Psychological Review*, **115**, 1069–1083.
- MICHEYL, C., & MESSING, D. P. (2006). Likelihood ratio, optimal decision rules, and correct response probabilities in a signal detection theoretic, equal-variance Gaussian model of the observer in the 4IAX paradigm. *Perception & Psychophysics*, **68**, 725–735.
- MOORE, B. C. J., & GLASBERG, B. R. (1989). Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. *Journal of the Acoustical Society of America*, **86**, 1722–1732.
- MOORE, B. C. J., GLASBERG, B. R., FLANAGAN, H. J., & ADAMS, J. (2006). Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure. *Journal of the Acoustical Society of America*, **119**, 480–490.
- MOORE, B. C. J., GLASBERG, B. R., LOW, K. E., COPE, T., & COPE, W. (2006). Effects of level and frequency on the audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America*, **120**, 934–944.
- MOORE, B. C. J., & SEK, A. (1998). Discrimination of frequency glides with superimposed random glides in level. *Journal of the Acoustical Society of America*, **104**, 411–421.
- NOREEN, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology* (Vol. 13, pp. 237–279). Providence, RI: American Mathematical Society.
- OXENHAM, A. J., & BUUS, S. (2000). Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions. *Journal of the Acoustical Society of America*, **107**, 1605–1614.
- OXENHAM, A. J., MICHEYL, C., & KEEBLER, M. V. (2009). Can temporal fine structure represent the fundamental frequency of unresolved harmonics? *Journal of the Acoustical Society of America*, **125**, 2189–2199.
- ROUSSEAU, B., & ENNIS, D. M. (2001). A Thurstonian model for the dual pair (4IAX) discrimination method. *Perception & Psychophysics*, **63**, 1083–1090.
- ROUSSEAU, B., & ENNIS, D. M. (2002). The multiple dual-pair method. *Perception & Psychophysics*, **64**, 1008–1014.
- SEMAL, C., & DEMANY, L. (2006). Individual differences in the sensitivity to pitch direction. *Journal of the Acoustical Society of America*, **120**, 3907–3915.
- SPIEGEL, M. F., PICARDI, M. C., & GREEN, D. M. (1981). Signal and masker uncertainty in intensity discrimination. *Journal of the Acoustical Society of America*, **70**, 1015–1019.
- STELLMACK, M. A., VIEMEISTER, N. F., & BYRNE, A. J. (2006). Discrimination of depth of sinusoidal amplitude modulation with and without roved carrier levels. *Journal of the Acoustical Society of America*, **119**, 37–40.
- THAI-VAN, H., MICHEYL, C., MOORE, B. C. J., & COLLET, L. (2003). Enhanced frequency discrimination near the hearing loss cut-off: A consequence of central auditory plasticity induced by cochlear damage? *Brain*, **126**, 2235–2245.
- VERSFELD, N. J., DAI, H., & GREEN, D. M. (1996). The optimum decision rules for the oddity task. *Perception & Psychophysics*, **58**, 10–21.
- VERSFELD, N. J., & HOUTSMA, A. J. (1991). Perception of spectral changes in multi-tone complexes. *Quarterly Journal of Experimental Psychology*, **43A**, 459–479.

NOTES

1. For instance, in a frequency-discrimination task, the experimenter can try to equalize loudness. Precise loudness equalization of tones that differ in frequency by a variable amount can be very difficult to achieve in practice because of irregularities and individual differences in equal-loudness contours (Mauermann, Long, & Kollmeier, 2004), especially at low sound levels or in hearing-impaired listeners (see McDermott, Lech, Kornblum, & Irvine, 1998; Thai-Van, Micheyl, Moore, & Collet, 2003).

2. Numerous studies have demonstrated that random variation along an irrelevant stimulus dimension can adversely affect performance in various perceptual tasks if the irrelevant and relevant dimensions are not independent, or separable (e.g., Ashby & Townsend, 1986; Garner, 1974). In addition, a few studies have demonstrated detrimental effects of increasing roving range on performance or thresholds in auditory intensity- and frequency-discrimination tasks (Jesteadt & Bilger, 1974) and spectral-shape discrimination (Mason et al., 1984). It is not entirely clear whether, and to what extent, these detrimental effects were due to roving actually limiting listeners' ability to use unwanted cues or to the random and irrelevant variations having a distracting influence.

3. Although Lyzenga and Horst (1995, 1997, 1998) mentioned using a 3AFC design, they instructed their listeners to select the odd stimulus. This suggests that, from the point of view of the listener, the task was essentially a form of three-interval oddity task.

4. This estimate is based on the formulas provided in Glasberg and Moore (1990) for calculating the shapes of auditory filters, as is defined by the rounded exponential function with a p value of 25, which corresponds to normal auditory filters. For such filters, a 2% frequency change on the steepest-slope side yields a change in output level of about 2–3 dB.

(Manuscript received July 1, 2009;
revision accepted for publication October 7, 2009.)