

# On the minimum audible difference in direct-to-reverberant energy ratio<sup>a)</sup>

Erik Larsen,<sup>b)</sup> Nandini Iyer,<sup>c)</sup> Charissa R. Lansing, and Albert S. Feng  
*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign,  
Urbana, Illinois 61801*

(Received 30 March 2006; revised 17 April 2008; accepted 5 May 2008)

The goals of this study were to measure sensitivity to the direct-to-reverberant energy ratio (D/R) across a wide range of D/R values and to gain insight into which cues are used in the discrimination process. The main finding is that changes in D/R are discriminated primarily based on spectral cues. Temporal cues may be used but only when spectral cues are diminished or not available, while sensitivity to interaural cross-correlation is too low to be useful in any of the conditions tested. These findings are based on an acoustic analysis of these variables and the results of two psychophysical experiments. The first experiment employs wideband noise with two values for onset and offset times to determine the D/R just-noticeable difference at  $-10$ ,  $0$ ,  $10$ , and  $20$  dB D/R. This yielded substantially higher sensitivity to D/R at  $0$  and  $10$  dB D/R ( $2-3$  dB) than has been reported previously, while sensitivity is much lower at  $-10$  and  $20$  dB D/R. The second experiment consists of three parts where specific cues to D/R are reduced or removed, which enabled the specified rank ordering of the cues. The acoustic analysis and psychophysical experiments also provide an explanation for the “auditory horizon effect.”

© 2008 Acoustical Society of America. [DOI: 10.1121/1.2936368]

PACS number(s): 43.66.Ba, 43.66.Qp [AK]

Pages: 450–461

## I. INTRODUCTION

Reverberation has a significant influence on speech communication, sound localization, and auditory perception in general. Although speech understanding is degraded in highly reverberant environments (Nábělek and Dagenais, 1986; Nábělek, 1988), musical listening can be enhanced if the amount of reverberation is appropriate for the type of music, which is often reflected in the design of listening spaces (Blessner, 2001).

Reverberation facilitates distance judgments because in the absence of sound reflections, distance is confounded with intensity at the ear, and thus it is nearly impossible to assess how far away a sound source is unless the listener has *a priori* knowledge of sound power. For unfamiliar sounds in anechoic environments, distance judgments typically converge to a default value (Coleman, 1962), which has been termed the *specific distance tendency* (see also Gogel, 1961; Mershon and King, 1975), irrespective of the actual sound source distance, although learning does occur and judgments tend to improve over time. For familiar sounds such as speech, anechoic distance localization is possible due to in-

tensity cues, but these can be unreliable or ambiguous (Philbeck and Mershon, 2002).

In reflective environments, acoustic cues having a one-to-one relationship with distance and which are not confounded with source characteristics are available, in particular, the “direct-to-reverberant energy ratio” or D/R (von Békésy, 1938; Mershon and King, 1975; Mershon *et al.*, 1989; Nielsen, 1993; Bronkhorst and Houtgast, 1999).<sup>1</sup> In a typical listening room, the direct sound field energy decays proportionally to (logarithmic) distance, while the reverberant sound field has approximately equal energy irrespective of distance. Thus, D/R can, in principle, be used to estimate the distance of a sound source. In anechoic environments, cues to distance are available for nearby sources, within approximately 1 m from the head (Brungart and Rabinowitz, 1999; Brungart *et al.*, 1999; Brungart, 1999; Shinn-Cunningham, 2000), but we do not further consider this special case.

In this study, we investigate the proficiency with which listeners discriminate signals with different D/R at various reference D/R levels. We also present an acoustic analysis of the properties of the reverberant sound field as a function of D/R, which can be used as a starting point for a more complete modeling effort. The goal of the psychophysical and acoustical analyses is to improve our understanding of perception of D/R, and thus distance perception in enclosed spaces.

Although listeners are sensitive to changes in D/R, it is not clear that this is based on an actual perception of D/R or on some other parameter that covaries with D/R. If listeners in fact do assign some D/R-equivalent measure to sounds, most prior work has assumed that this is achieved by a tem-

<sup>a)</sup>Portions of this work were presented in “On temporal vs. frequency-based discrimination of direct-to-reverberant energy ratio,” 146th Conference of the Acoustical Society of America, Austin, TX, November 2003, and “An evaluation of two frequency-based mechanisms for direct-to-reverberant energy ratio discrimination,” 147th Conference of the Acoustical Society of America, New York, NY, May 2004.

<sup>b)</sup>Present address: Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology, Cambridge, MA. Electronic mail: elarsen@alum.mit.edu

<sup>c)</sup>Currently with General Dynamics Advanced Information Systems. Electronic mail: nandini.iyer@wpaf.af.mil

poral integration method, and a model exists (Bronkhorst and Houtgast, 1999). If this hypothesis is correct, then variation in the onset time of sounds should affect the ability to discriminate changes in D/R, but most studies (Santarelli *et al.*, 2000; Bronkhorst, 2001; Zahorik, 2002a, 2002c) have found that the effect of temporal modulation (as measured by signal onset/offset time) on identification or discrimination of D/R is at most minor. Other variables, such as interaural cross-correlation (IACC), spectral variance (frequency-by-frequency variation in power spectrum), and spectral envelope, may be more important. In Sec. II, we will analyze how these variables covary with D/R to investigate their potential roles in D/R perception. This analysis reveals that spectral and binaural cues offer alternative explanations for listeners' sensitivity to D/R. In Sec. IV, we present further evidence for this by showing degradation in sensitivity to D/R when removing certain acoustic features of signals that otherwise leave D/R, as calculated from its acoustic definition, unchanged.

Two previous studies aimed to find the just-noticeable difference (JND) for D/R. Reichardt and Schmidt (1966) used classical music (presented in an anechoic chamber with four loudspeakers and a yes/no procedure) with adjustable D/R to establish the JND at various reference D/R values, holding the overall sound level constant, to develop a scale of "spatial impression" ("Räumlichkeit"). This scale ranged from fully anechoic to highly reverberant. Based on the obtained JNDs, they found 14 discernible steps in the range (-23, 23) dB D/R. The JND versus D/R value had a U shape, with a minimum of 2 dB at 0 dB D/R and rising to about 20 dB at  $\pm 20$  dB D/R. Zahorik (2002c) used virtual acoustics to assess the JND at 0, 10, and 20 dB D/R (roving the overall intensity level of signals) in a 2AFC procedure. He found roughly constant JNDs of 5–6 dB at these D/R for four different signal types (two noise signals, a speech syllable, and an impulse) and for both medial and lateral sources. Zahorik attributed the large discrepancy between the results of his and Reichardt and Schmidt's study mainly to the fact that Reichardt and Schmidt held the overall level of the stimuli constant, thus allowing listeners to focus on changes in the reverberant (or direct) energy level to discriminate the signals as D/R was manipulated.

In this paper, we start by analyzing ear-canal signals as a function of D/R, uncovering how room acoustics affects these signals. Applying this analysis to D/R discrimination, we find that the general properties of reverberant sound fields in enclosed spaces create well-defined physical/acoustical constraints. We hypothesize that these constraints will be reflected in behaviorally obtained discrimination thresholds; if applied to distance perception, they provide a natural explanation for the "auditory horizon effect" (see, e.g., Bronkhorst and Houtgast, 1999). We complement the acoustic analysis with behavioral discrimination data over a broad range of D/R values in a manner similar to Zahorik (2002c) but extended to negative D/R and also for "impoverished" signals. The first experiment, described in Sec. III, was conducted to establish base line D/R JNDs at -10, 0, 10, and 20 dB D/R for signals with two different onset times. In the second experiment (Sec. IV), the signals presented to listeners were

modified to selectively remove potential D/R cues. Experiment 2a employs monaural listening (removing binaural cues); experiment 2b removes most spectral cues by band limiting the signals, while experiment 2c further reduces spectral cues by roving the frequency region of the narrow band signals. The JNDs for these impoverished signals, compared to the JNDs of experiment 1, can indicate the relative contribution of specific acoustic cues to discrimination of D/R.

## II. ACOUSTIC ANALYSIS

External changes in a sound field correlate to changes in internal (psychological) variables. Thus, the internal processes can be probed by manipulating the external signals and linking the two with a quantitative model, as originally outlined by Fechner (1912). To properly discuss how the external environment influences perceptual processes, e.g., in the case of D/R perception, we need to briefly consider such a quantitative model. Following the approach and notation from Allen and Neely (1997), the JND for the external variable (we use the D/R JND  $\Delta\nu$ ) is proportional to the JND for the internal variable ( $\Delta\psi$ ). The proportionality factor depends on the functional relationship between external and internal variables ( $d\psi/d\varphi$ ) and the physical relationship between the external variable and the quantity that is being manipulated, which is D/R ( $\nu$ ) in our case. The exact relationship is

$$\Delta\nu = \Delta\psi \left( \frac{d\psi d\varphi}{d\varphi d\nu} \right)^{-1}. \quad (1)$$

We will propose various physical variables  $\varphi_i$  (e.g., interaural correlation or spectrum) that have the potential for being useful in D/R discrimination, and we compute their dependence on D/R [ $\varphi_i(\nu)$ ]. This will allow us to compute  $d\varphi_i/d\nu$  in Eq. (1) but not the other components,  $\Delta\psi$  or  $d\psi/d\varphi$ . In the remainder, we will refer to this as modeling the effect of the *physical relationship* between D/R and the acoustic variables. This stands in contrast to analyzing the effect of *perceptual sensitivities* on the JND, which, though equally important, is beyond the scope of this work.

A complication arises in a listening task (e.g., D/R perception) where redundant cues are available. When multiple cues can be used at once, these could be combined to improve discrimination performance, leading to smaller JNDs than would be obtained if any cue was used in isolation (Lutfi and Wang, 1999; Ernst *et al.*, 2000; Ernst and Banks, 2002; Hillis *et al.*, 2004). Such an approach requires quantitative knowledge about the contribution of the various cues; in Sec. IV, we selectively remove potential cues to qualitatively assess their individual contribution for D/R discrimination without considering how they interrelate.

### A. Interaural cross-correlation

Reverberation introduces binaural cues by altering the sound attributes at the two ears differentially. IACC ( $\varrho$ ) is a powerful binaural cue (Blauert, 1984), and sensitivity to changes in IACC can be high: the JND  $\Delta\varrho$  at  $\varrho=1$  is about 0.02–0.04 but increases strongly as  $\varrho$  decreases (Pollack and

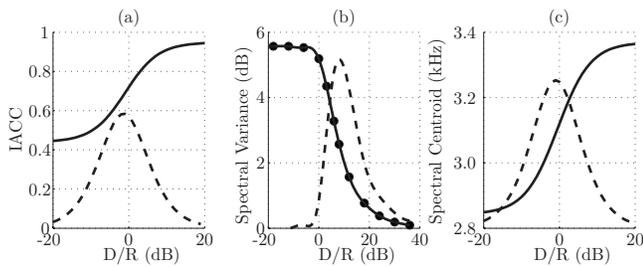


FIG. 1. Dependence of acoustic variables on D/R. Panel (a): Relation between D/R and IACC, calculated from experimental signals. Panel (b): Relation between D/R and variance of the power spectrum, adapted from [Jetzt \(1979\)](#). Panel (c): Relation between D/R and spectral centroid (or spectral CoG), calculated from experimental signals. In all panels, the solid line indicates the dependence of the acoustic variable on D/R, while the dashed line indicates the (arbitrarily) scaled derivative. In all cases, the acoustic variables change monotonically as a function of D/R but asymptote at large positive and negative D/R.

[Trittipoe, 1959](#); [Gabriel and Colburn, 1981](#); [Boehnke et al., 2002](#)). Because reverberation decorrelates signals at the ears ([Blauert, 1984](#)), IACC decreases as D/R decreases, making it a potential cue for D/R.

IACC is calculated as

$$\rho(\tau) = \frac{\overline{x_1(t)x_2(t-\tau)}}{\sqrt{\overline{x_1^2(t)} \overline{x_2^2(t)}}, \quad (2)$$

where  $x_1(t)$  and  $x_2(t)$  are signals at the left and right ears, respectively, and  $\tau$  is a time delay ( $\overline{x_1}$  indicates the expected value of  $x_1$ ). Applying Eq. (2) to ear signals measured in the room that was used to obtain experimental data for experiments 1 and 2 (described later) yields the relationship between IACC and D/R as shown in panel (a) of Fig. 1 (solid line); the scaled derivative of the IACC is also shown (dashed line). The maximum rate of change in IACC appears to occur at about 0 dB D/R, and otherwise, the rate of change is approximately symmetrical around this value. Equation (1) shows that the contribution of the physical relationship of D/R and IACC minimizes the JND around D/R values of 0 dB and to enlarge it at increasingly negative or positive D/R.

## B. Spectral variance

Changes in sound spectrum produce salient perceptual cues (for a review, see [Green, 1988](#)). There are at least two different kinds of spectral changes that occur when a sound source moves away from a receiver or, equivalently, when D/R decreases. In this and the following section, we will analyze these two effects. Here we discuss changes in the fine structure of the spectrum, by which we mean the frequency-to-frequency variations in the magnitude spectrum that occur as a result of the interference of reflected sound waves. Prior studies have indicated that people are sensitive to such fine-structure changes ([Green, 1988](#); [Berkley and Allen, 1993](#)). In a multitonal background, sensitivity to changes in the amplitude of a single component seems to be highest when the overall amplitude variability is small ([Kidd et al., 1986](#)).

[Jetzt \(1979\)](#) showed that the variance ( $\sigma^2$ ) of the spectral response between two locations in a room is exclusively de-

termined by D/R, as reproduced in panel (b) of Fig. 1.<sup>2</sup> The dots are data as given by [Jetzt](#) in his Table I, and the solid line is a cubic spline interpolation of those data; the dashed trace is the scaled derivative of  $\sigma^2$  as a function of D/R. The derivative has a maximum at about 8 dB D/R and is large only at positive D/R values. Thus, according to Eq. (1), the contribution of the physical relationship between D/R and spectral variance is to provide cues at moderately large positive D/R values, peaking at about 8 dB D/R.

## C. Spectral envelope

A second kind of spectral change that occurs when D/R changes is the spectral envelope (sometimes also referred to as the “spectral shape”). Spectral envelope changes are thought to influence distance perception because empirical findings show that low-pass filtered signals are judged to be further away as the cut-off frequency decreases; the effect has been attributed to the relatively larger absorption of high sound frequencies in air, which creates progressive low-pass filtering as the source distance increases ([Coleman, 1968](#); [Butler et al., 1980](#); [Little et al., 1992](#); [Zahorik, 2002a, 2002c](#)). However, the reduction in high-frequency content necessary to yield such increases in source distance is always much greater than that produced by air absorption alone. A different explanation is based on the fact that most materials commonly used in rooms absorb more high- than low-frequency energy, such that each reflected sound wave will be low-pass filtered. As the reverberant sound field consists of waves that have been reflected numerous times, the spectral envelope of the reverberant sound field will be shifted toward lower frequencies relative to the direct sound (see also [Nielsen, 1993](#)).

To remain consistent with the other acoustic parameters under consideration, we analyze spectral envelope changes with a one-parameter model.<sup>3</sup> To capture perceived changes in sound timbre as a function of D/R, we use the spectral centroid, or spectral center of gravity (CoG),  $\chi$ , according to

$$\chi = \frac{\sum_{i=2}^N f_i X_i}{\sum_{i=2}^N X_i}, \quad (3)$$

where  $X_i$  is the  $i$ th component of the discrete power spectrum of the signal  $x$ , and the summation includes all  $N$  frequency components  $f_i$  up to the Nyquist frequency, except dc (hence the summation starts at index value 2). As explained above, the reverberant sound field has less high-frequency energy than the direct sound field, which means that CoG decreases as the relative content of reverberation is larger, i.e., as D/R decreases.

Aside from its dependence on D/R, the spectral envelope depends on the characteristics of the source signal also. For the signals that were used in experiment 1 (Sec. III), we obtain the CoG  $\chi(\nu)$  as shown in panel (c) of Fig. 1 (solid line), as well as its derivative (dashed line). The maximum rate of change in CoG occurs at about 0 dB D/R and is otherwise approximately symmetrical around this value, becoming near zero at large positive and negative values. Thus, the physical relationship between D/R and the spectral envelope acts to minimize the JND around 0 dB D/R [cf. Eq. (1)].

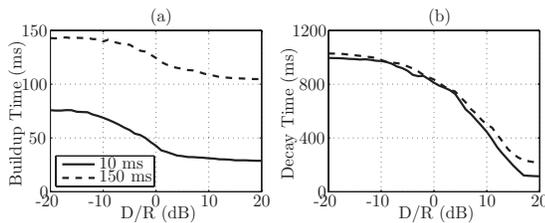


FIG. 2. Dependence of temporal cues on D/R. Panel (a): Relation between D/R and sound buildup time, calculated from experimental signals, using broadband noise with onset times of 10 and 150 ms. Panel (b): Relation between D/R and sound decay time, calculated from experimental signals, using the same two source signal types. Both sound buildup and decay times vary monotonically with D/R but asymptote at large positive and negative D/R. Buildup time varies mainly at negative D/R, while decay time varies mainly at positive D/R.

#### D. Temporal integration

At large positive D/R values, buildup and decay of sounds at the ear canal will closely match the onset and offset times of the sound source,<sup>4</sup> while at large negative D/R values, the room's reflections may alter these patterns. In most practical cases, the response time of the room, as parametrized by reverberation time (Sabine, 1962; Kuttruff, 1991), is longer than the onset/offset time of typical source signals, e.g., speech. Thus, as D/R decreases, sound buildup and decay are expected to become more sluggish, possibly providing cues to D/R.

We compute the buildup and decay of the signals used in experiment 1 by integrating the squared signal amplitude over time based on the Schroeder (1965) method for calculating reverberation time (for decay time analysis, we integrate backward in time starting at the end of the signal). We express buildup (decay) time as the period required for the signal to increase (decrease) by 60 dB at the start (end) of the signal by using straight portions of the energy buildup (decay) curves. The results of this analysis for broadband noise signals with fast (10 ms) and slow (150 ms) onset/offset times, as used in experiment 1, are shown in Fig. 2. There is a strong dependence on D/R (the derivatives of these curves are quite noisy and thus omitted).

Panel (a) of Fig. 2 shows buildup time versus D/R for the 10 and 150 ms onset time signals. These curves appear similar except for a vertical offset, i.e., changes in buildup time as a function of D/R are similar for both signals. However, for purposes of discriminating differences in buildup time, relative changes may be more relevant (Kewley-Port and Pisoni, 1984), in which case, the fast-onset signal should provide the strongest temporal discrimination cues. Panel (b) shows decay time versus D/R for both signals, and the results appear very similar, at least for D/R values up to about 5 dB, after which changes in the slow-offset signal become more gradual. Thus, potential mechanisms employing decay time to discriminate D/R should work equally well for fast- or slow-offset signals, except at relatively large D/R values, where fast-offset signals may provide stronger cues.

Another feature of these curves is that most of the change in buildup time occurs at moderately negative D/R values, between about -10 and 0 dB, especially for the fast-onset signal. In contrast, most of the change in decay time occurs at positive D/R values, between about 0 and 15 dB,

for both types of signals. This may indicate that temporally based D/R discrimination mechanisms may rely primarily on differences in either buildup or decay time depending on the D/R regime.

#### E. Acoustic analysis: Discussion

We have analyzed the dependence of four acoustic variables on D/R: IACC, spectral variance, spectral envelope, and buildup/decay time. We have shown that these variables have a monotonic relationship with D/R, and prior studies have shown that listeners are sensitive to variations in these variables. We therefore hypothesize that discrimination of D/R can make use of any or all of these cues.

Our quantitative models of these cues are without doubt significant simplifications of how the auditory system could assess differences between signals at various D/R values. For example, spectral envelope changes are more complicated than can be captured by a single parameter such as the spectral CoG. However, these simple single-parameter approximations do capture the important fact that physical changes in ear-canal signals asymptote at large positive and negative D/R values. More sophisticated analyses would refine these results, but we do not expect that they would alter this basic fact. Also note that the particular details of the equations we used to analyze acoustic changes are somewhat arbitrary; therefore numerical values in Figs. 1 and 2 should be taken with a grain of salt and are most useful to make *relative* comparisons between D/R values or between signal types.

Equation (1) shows that the D/R JND depends on (i) the physical relationship between D/R and the acoustic variable, (ii) the transformation of the acoustic (physical) to psychological variable, and (iii) the internal sensitivity to that psychological variable. We have only considered aspect (i), such that our conclusions are limited in scope. Specifically, our analysis suggests the following:

- (1) Acoustic variables change as a function of D/R only within a limited range, loosely defined as moderately valued positive and negative D/R.
- (2) There are D/R regions where none of the three variables we studied changes (except in an asymptotic sense). Equation (1) predicts that as these regions are approached, the D/R JND increases and eventually becomes very large.

A corollary is that D/R discrimination is possible within a limited range of D/R values only. Our analysis does *not* suggest any of the following:

- (1) At what value(s) of D/R the D/R JND is smallest (discrimination is best). This requires knowledge about how sensitivity to the underlying variables changes with D/R.
- (2) The exact extent of the region in which D/R discrimination is possible. The acoustic analysis merely implies that this region is limited in extent.<sup>5</sup>
- (3) A rank ordering of the acoustic variables in terms of contribution for D/R discrimination. This requires knowledge regarding the internal sensitivity to each of the underlying variables.

We can test these suggestions derived from the acoustic analysis by measuring D/R JNDs over a large range of D/R

values. The prediction is that at the edges of the measured region, D/R JNDs will increase. Although we do not know *a priori* how large of a D/R region is sufficient, we choose  $-10$  to  $20$  dB D/R in experiment 1. From Figs. 1 and 2, we can see that at least the values of the physical variables start to asymptote at the edges of this region, such that we might expect to see some variation in D/R JND.

Many experiments on auditory distance perception yield a compressive function of perceived versus actual distance; the restricted range of perceived distances is commonly known as the “auditory horizon” effect (Bronkhorst and Houtgast, 1999; Zahorik, 2002a). This effect is consistent with our acoustic analysis. The fact that relevant acoustic properties of the perceived sound signal (IACC, spectral and temporal cues) remain essentially constant once the sound source moves to distances well beyond critical distance (large negative D/R) is an unavoidable property of room acoustics and is responsible for the auditory horizon effect. Sources well beyond the critical distance should be judged to be closer than they actually are because the signal reaching the listener’s ears is very similar to the signal of a closer sound source, which explains the compression of perceived distances.<sup>6</sup>

### III. EXPERIMENT 1: DISCRIMINATION ABILITY AS A FUNCTION OF DIRECT-TO-REVERBERANT ENERGY RATIO

The goal of experiment 1 was to assess the JND for D/R at various D/R values ( $-10$ ,  $0$ ,  $10$ , and  $20$  dB) and to mimic the experiments by Zahorik (2002c), although some details were different (described in Sec. III B). One major difference with Zahorik’s study is that we also determine the JND at a negative D/R value. According to the analysis of Sec. II, JNDs should increase at sufficiently large positive and negative D/R. Thus this experiment also investigated the extent to which this prediction could be observed in behavioral performance.

#### A. Methods

##### 1. Subjects

Eight listeners (four female, four male; age 19–36 years) participated in the experiment. All listeners had audiometric thresholds below 20 dB hearing level between 250 and 8000 Hz, prior experience in psychoacoustic experiments, and participated in daily 1 h sessions for a period of 2 weeks. Two of the authors participated as subjects. The experiment was performed both at the University of Illinois at Urbana-Champaign and at Wright-Patterson Air Force Base; four subjects were tested at each location. Since no systematic differences in mean thresholds were found between the two groups, the data from the two laboratories will be combined for subsequent analysis.

##### 2. Stimuli

*Virtual sound source technique.* Two types of anechoic source signals (sample rate of 20 kHz) were convolved with Binaural room impulse response (BRIRs) measured in an auditorium to create the virtual sources that were used in the experiment. The signals were as follows:

- (1) Wideband (white) noise (WBN) of 300 ms duration, including 150 ms raised cosine onset and offset (i.e., no steady-state portion). The bandwidth of this signal was limited only by the bandwidth of the acoustic response of the auditorium and the measurement system (100–10 000 Hz).
- (2) WBN of 300 ms duration, including 10 ms raised cosine onset and offset.

The auditorium was a rectangular-shaped room of approximately  $1000\text{ m}^3$  with a shallow sloping seating area. Reverberation time  $T_{60}$  measured in the auditorium was 0.78 s (average of 0.5 and 1 kHz octave bands). To analyze the variation in D/R in the auditorium, BRIRs at several source-to-receiver distances were measured, and the corresponding D/R values were computed. The D/R decayed at a rate of approximately 4.7 dB per doubling of distance.

A power amplifier (ADCOM GFA-535II) and loudspeaker (Analog and Digital Systems L200e) were used to play maximum length sequences (Rife and Vanderkooy, 1989) of order 14 at 20 kHz (duration: 0.82 s). The loudspeaker was positioned at a distance of 4.0 m from the recording location, in the seating area of the auditorium, at  $0^\circ$  azimuth (straight ahead). Recording was achieved with a KEMAR and ER-1 microphones (coupled to Zwislocki ear simulators) and preamplifiers (Knowles), placed on the center of a raised stage in the front of the auditorium. The KEMAR was positioned such that there were no obstructions between it and the loudspeaker. Data were acquired by a DAQPad-6052E (National Instruments), interfaced with a laptop computer, used for signal generation and storage through custom MATLAB (The Mathworks) software and MATLAB’s data acquisition toolbox. Note that the use of nonindividualized BRIRs is not expected to affect perception of D/R, as shown previously by Zahorik (2002b, 2002c). We know of no specific study that has investigated potential issues of using KEMAR for distance perception, but it appears unlikely that discrimination of D/R would be problematic with room signals recorded through KEMAR.

The measured BRIRs were manipulated in postprocessing such that any desired D/R could be obtained. The two anechoic source signals were then convolved with all manipulated BRIRs to obtain the test stimuli. Signals were stored on file in 16 bit wav format.

*Direct-to-reverberant energy ratio manipulation.* The measured BRIR at 4.0 m source distance was used in all subsequent signal generation procedures. It was used to construct BRIRs with any desired D/R value by scaling the direct sound portion (defined by a window of 3 ms length after direct sound onset) with an appropriate amount; these modified BRIRs were convolved with the anechoic source signals to synthesize signals used in the experiment. Although D/R manipulation can also be done by scaling the reverberant portion of the BRIR, we chose to scale the direct sound as this corresponds more closely to the physical situation in a room, where direct sound decays with increasing source distance and the reverberant level remains more or less constant. Thus, all manipulated BRIRs had the same reverberant energy level (before the level rose, see the next section).

### 3. Procedure

A 2I 2AFC procedure was used to determine the D/R JND. In order to assess differences in JND as a function of D/R, each signal type was used at four D/R values of 20, 10, 0, and -10 dB (order counterbalanced between subjects); these were the reference signals. The 20 dB D/R condition was only tested at the Wright-Patterson Air Force site, i.e., only with four instead of eight subjects. In each presentation trial, the reference and target signals were presented in random order, with an interstimulus interval of 500 ms. At each reference, the two signal types (10 and 150 ms) were tested in random order.

At the beginning of each block of trials, the target signal had a D/R that was 10 dB higher than the reference signal. Listeners had to identify the most reverberant sounding signal (lowest D/R), i.e., the reference signal. The adaptive variable was the D/R of the target signal (the reference signal D/R never changed within a block of trials). An adaptive step size was used, initially set to 4 dB, decreased to 2 dB after two reversals, and set to the final value of 1 dB after another two reversals. Thresholds estimated the 79.4% point on the psychometric function (probability level  $p_1=0.794$ ) using a three-down, one-up adaptive procedure (Levitt, 1971). Signals were presented in blocks of 60 trials. Thresholds per block were obtained by averaging the D/R difference between the target and reference signals for the final ten reversals or at all reversals if fewer than ten occurred in a particular block (as was the case for approximately 10% of all blocks). All listeners practiced for at least four blocks per condition before data were collected. Following the practice sessions, data collection continued until thresholds were stable across six consecutive blocks of trials. The JND for each listener per condition was then obtained as the average threshold value obtained for the last six blocks.

Because our method of creating D/R changes leads to changes in signal level, listeners could potentially discriminate D/R using level cues. To control for this confound, the overall level of the signals was roved by  $R=20$  dB, and listeners were specifically instructed to ignore the loudness of the signals. Thus, for each interval, the signal level was chosen as the nominal level plus an offset randomly chosen from a uniform distribution in the range -10 to +10 dB. The nominal presentation level was adjusted by each individual to a comfortable listening level while also ensuring sufficient audibility at 10 dB below the nominal level. The level rove ensures that detection of changes based only on signal level will lead to a threshold  $C$  that is at least

$$C = R[1 - \sqrt{2(1 - p_1)}], \quad (4)$$

as shown by Green (1988). If thresholds are lower, discrimination must have been based on other cues, in our case D/R. In this equation,  $p_1$  is the probability level tracked, which is 0.79 in our case. With  $R=20$  dB, this yields  $C=7.15$  dB.

Theoretically available level cues were mediated by direct and/or overall (direct and reverberant) sound because D/R was manipulated by changing direct sound while holding reverberation constant. This method of signal manipulation means that a specific change in the D/R value is created

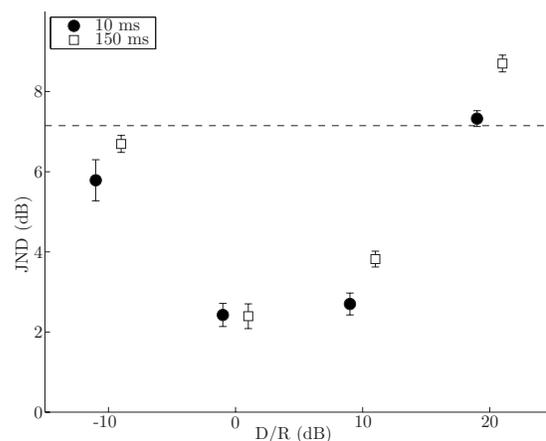


FIG. 3. Mean D/R discrimination thresholds from experiment 1 at -10, 0, 10, and 20 dB D/R; error bars indicate standard error. Filled symbols indicate fast-onset noise; open symbols indicate slow-onset noise. The dashed line at 7.15 dB indicates the lowest threshold that could be obtained if subjects used level cues only instead of D/R.

by the exact same change in the direct sound level. Changes in overall level are always smaller, as can be shown theoretically,<sup>7</sup> as well as empirically by computing the overall level as a function of D/R by using the experimental signals (results not shown). It is undisputed that listeners could use the overall level of signals within trials for discrimination, but whether direct level can be extracted from the total signal and used as a discrimination cue is not certain. Until it is known whether the direct level can be a useful cue, it appears prudent to assume that it is because this leads to more stringent rejection criteria. In conclusion, as long as obtained thresholds are well below 7.15 dB at any D/R reference, use of level cues can be excluded, and thresholds are reliable indicators of D/R discrimination performance. In practice, we consider thresholds reliable if the mean is at least one standard deviation below the 7.15 dB ceiling imposed by the level rove.

Listeners were seated in a sound-attenuating room and listened binaurally through headphones (Sennheiser HDA 200). Signal presentation was accompanied by visual displays (box outlines) on a computer monitor and listeners were required to respond to each trial by pressing the appropriate response box using a mouse (corresponding to the listening intervals) on the computer monitor. After the response was given, the box outline corresponding to the correct listening interval was highlighted to provide visual feedback regarding the correct response.

### B. Results and discussion

Mean thresholds and standard errors are shown in Fig. 3 for both the fast- and slow-onset noise signals. The dashed line at 7.15 dB indicates the threshold that would be obtained if listeners had employed level cues alone instead of D/R [see Eq. (4)]. Averaged thresholds (with standard error in parentheses) for the noise signal with 150 ms onset were 6.7 (0.2), 2.4 (0.3), 3.8 (0.2), and 8.7 (0.2) dB for reference D/R of -10, 0, 10, and 20 dB, respectively. Similarly, the average thresholds for the noise signal with 10 ms onset were 5.8 (0.5), 2.4 (0.3), 2.7 (0.3), and 7.3 (0.2) dB for the same ref-

erence D/R values. These JNDs correspond to changes in the overall level of about 1, 1.5, 3, and 8 dB at -10, 0, 10, and 20 dB D/R references. The threshold at 20 dB D/R is thus likely contaminated by the overall level cues, as overall level changes exceed the 7.15 dB ceiling. If direct level cues, which vary by exactly the same amount as D/R does, are also available to the listener, the threshold at -10 dB D/R reference is also unreliable because it is less than one standard deviation below ceiling (in both conditions). Thus, for both -10 and +20 dB D/R, the thresholds we found cannot be taken as reliable indicators of the JND at those D/R. Nonetheless, it is clear that whatever the true JNDs at -10 and 20 dB D/R are, they are significantly greater than the JNDs at 0 and 10 dB D/R by at least 4 dB.

Considering for the moment that only the data at 0 and 10 dB D/R are reliable, a two-way repeated-measure analysis of variance (ANOVA) with D/R (0 and 10 dB D/R) and signal type as factors was performed. This yielded a significant main effect of signal type ( $F_{(1,7)}=10.56$ ,  $p=0.01$ ) and D/R ( $F_{(1,7)}=22.24$ ,  $p=0.002$ ). The interaction effect of D/R and signal type was also significant ( $F_{(1,7)}=26.76$ ,  $p=0.001$ ) due to the tendency for thresholds to increase when D/R increased from 0 to 10 dB in the slow-onset noise condition (by 1.4 dB) but not in the fast-onset noise condition (by 0.3 dB, which was not significant). Prior studies generally indicate that sound onset time has no, or at most a minor effect, on D/R discrimination and distance perception (Santarelli *et al.*, 2000; Bronkhorst, 2001; Zahorik, 2002a, 2002c). These views may need to be re-examined in view of the present results, as we did find a modest yet statistically significant effect of onset time on D/R discrimination thresholds.

At positive D/R values, sound decay time provides more powerful acoustic cues than sound buildup time (which is nearly constant in that regime) according to the analysis in Sec. II D. This may imply that at the reference of 10 dB D/R, where the JND for the slow-offset signal is larger than that for the fast-offset signal, it is more difficult for listeners to use variations in the decay time of slower- versus faster-offset sounds to discriminate D/R. At 0 dB D/R, no such difference in JND exists between the signal types. It is possible that at this particular D/R value, other cues are more powerful, rendering temporal aspects of sounds less important.

*Comparison to prior studies.* The obtained JNDs at 0 and 10 dB D/R are in good agreement with the data from Reichardt and Schmidt (1966) but are lower than those reported by Zahorik (2002c) (JNDs at -10 and 20 dB are quite high by Reichardt and Schmidt's results as well, but it is difficult to make quantitative comparisons as our JNDs are at or near ceiling at those D/R values). This is somewhat surprising, given that our signals and methods were more similar to those used by Zahorik (2002c). Zahorik previously pointed out that low JND estimates in the study by Reichardt and Schmidt (1966) might be the result of the methods and procedures used in that study, which did not effectively control for level confounds. Nonetheless, the results of Reichardt and Schmidt (1966) are partially supported by our data: our methods were similar to Zahorik's and involved

controlling for level confounds (confirmed by our JNDs being well below the 7.15 dB threshold level at 0 and 10 dB D/R), but we still found similarly low JNDs as Reichardt and Schmidt.

Comparing our data with those of Zahorik (2002c), who measured JNDs at 0, 10, and 20 dB D/R (using methods and signals that were broadly similar to ours), we note that at 0 and 10 dB D/R, our JNDs are lower: 2.5–3.5 dB versus 5–6 dB. Differences in experimental procedure or signals used (similar but different rooms, KEMAR versus individualized BRIRs, 20 versus 40 kHz sample frequency, manipulating direct versus reverberant level to vary D/R) make a direct comparison difficult, although differences of nearly 4 dB are considerable and cannot be easily ignored. Zahorik manipulated the reverberant level of his signals to vary D/R, so it is possible that cues from the reverberant sound level had some effect on the obtained JNDs (similar to possible confounds from the direct sound level in our study at -10 and 20 dB D/R). Zahorik showed that changes in overall level were too small to account for his results, but he did not discuss the issue of the reverberant level cue.

A more qualitative and perhaps more important difference is that our JNDs steadily increase at reference D/R values of 0 to 10 to 20 dB D/R (JNDs: 2.5 to 3.3 to 8 dB), while Zahorik's JNDs remain constant at 5–6 dB over the entire range of D/R values. It is remarkable that discrimination performance could stay constant over a 20 dB range of D/R values, considering our acoustic analysis, which indicates that acoustic variables change rapidly around 0 dB D/R but are nearly constant at 20 dB D/R. Although D/R discrimination likely makes use of redundant cues, all cues that we have been able to consider become weak at large D/R values, so that increases in JND appear unavoidable. As our results differ both quantitatively and qualitatively from Zahorik's, it would appear necessary in future studies to further replicate this style of experiment to gain more insight into the importance of experimental parameters on D/R JNDs.

*Comparison to acoustic analysis.* In Sec. II E, we predicted on the basis of our acoustic analysis that if a sufficiently large range of D/R was sampled, one would find an increase in JND at the edges (due to the fact that acoustic variables are nearly constant in those regions). Our results presented here appear to mirror these predictions, in that we find a large increase in JNDs at -10 and 20 dB relative to JNDs at 0 and 10 dB. The increase is at least 4–5 dB but possibly more because JNDs at -10 and 20 dB were near or at the ceiling level. Some of this increase could also be explained by a decrease in sensitivity to the variables that are being discriminated, but this would be an independent additive effect. Our theoretical explanation for the auditory horizon effect (Sec. III E) then also appears to find support in the experimentally obtained D/R JNDs presented here.

*Implications for distance perception.* D/R JNDs of about 2.4 dB at 0 dB D/R means that JNDs for distance perception are about 25% for sources near the critical distance.<sup>8</sup> At about 1/3 of the critical distance (10 dB D/R, JNDs of 2.7–3.8 dB), distance JNDs are 25%–35%, while at 1/10 of the critical distance (20 dB D/R, JNDs at least 7 dB), they

are at least 55% (unless this is within 1 m of the head, in which case binaural cues can be used to maintain errors to within 30%–40%, [Brungart et al., 1999](#)). For sources at three times the critical distance (–10 dB D/R, JNDs of 5.8 dB versus at least 7 dB for the fast- versus slow-onset signal), distance JNDs are 35% versus at least 55% for the fast- versus slow-onset signal. All these distance JNDs are with respect to changes that bring the source *closer* to the listener, as the D/R JNDs were measured for *positive* changes in D/R. JNDs in the opposite direction may be different.

The transformed distance JNDs of 25%–35% (valid for reference distances of about 30%–100% of the critical distance) are somewhat better than the estimated distance JNDs reported by [Zahorik \(2002c\)](#), which used data from [Zahorik \(2002a\)](#): these were around 50% of the reference distance (changes in source distance also *toward* the listener). The difference can be directly attributed to the lower D/R JNDs we obtained.

## IV. EXPERIMENT 2: DISCRIMINATION ABILITY FOR IMPOVERISHED SIGNALS

The aim of experiment 2 is to reduce or remove each of the three previously discussed cues for D/R from the test signals (IACC, spectral variance, and spectral envelope). It is hoped that by observing the effect of these signal manipulations on the JND, we will gain insight into the relative importance of each cue for D/R discrimination. Binaural cues will be removed in experiment 2a, while experiment 2b will remove a large portion of both spectral variance and spectral envelope cues. Finally, experiment 2c will remove the remaining spectral envelope cues.

### A. Methods

#### 1. Subjects

The same listeners participated as in experiment 1; sessions were 1 h daily for a period of 2 weeks. One-half of the subject group ran experiment 2 before experiment 1 to control for practice effects.

#### 2. Stimuli

*Experiment 2a.* Experiment 2a presented signals to the listeners monaurally (right ear only). This removes all binaural cues, including IACC (Sec. II A). The level of the signal at the right ear was raised by 3 dB relative to the level of the signals used in experiment 1 to preserve the overall intensity of the sound. The signals were otherwise identical to those used in experiment 1.

*Experiment 2b.* Experiment 2b presented signals dichotically but with reduced spectral cues. This was achieved by filtering the signals into narrow frequency bands, three Equivalent rectangular bandwidth (ERB) wide, around center frequencies of 500 Hz and 3 kHz. Here and in experiment 2c, only the fast-onset (10 ms) noise signal was used, convolved with appropriately manipulated BRIRs.

The reduction in independent spectral samples in the signals used in experiment 2b relative to those of experiment 1 significantly increases the uncertainty with which the variance of the power spectrum may be estimated. As this uncer-

tainty increases, the usefulness of the spectral variance cue for D/R discrimination is reduced. Spectral envelope cues are also reduced by bandwidth reduction because variations in spectral centroid/CoG are diminished. We analyze the reduction in both spectral cues more fully in Sec. IV B. Temporal cues (sound buildup and decay) were not greatly altered for the signals and D/R regime we used (data not shown) and were similar for the 500 Hz and 3 kHz center frequency signals.

*Experiment 2c.* For experiment 2c, similar signals as in experiment 2b were used (dichotic presentation) but with the spectral envelope cues completely removed. This was achieved by roving the center frequency of the three ERB-wide narrow band signals by one ERB unit at both center frequencies (500 Hz and 3 kHz). By ensuring that the frequency rove was larger than the variation in the spectral CoG as a function of D/R, this effectively removed the spectral envelope cues. Note that the spectral envelope cues are already significantly reduced for narrow band versus wideband signals; the frequency rove removes the remaining spectral envelope cues. Spectral variance cues should not be affected by the frequency rove.

The low-frequency signal had a center frequency roved in the range of 500–582 Hz, while the high-frequency signal had a center frequency roved in the range of 3000–3367 Hz. The center frequency that was used for each stimulus was determined randomly out of a sample of ten center frequencies distributed uniformly between the lower and upper limits for each frequency range.

### 3. Procedure

Thresholds were determined using a 4I 2AFC procedure; the first and last interval always contained the target signal, which was also present in either the second or third interval, and varied randomly. The random 20 dB level rove was applied to each signal in the four intervals. In experiment 2c, all four signals were randomly roved in the center frequency by 1 ERBu. Listeners were instructed to indicate whether the reference signal (most reverberant signal) was present in either the second or third interval.

Thresholds were not collected at –10 or 20 dB D/R as the JNDs found at those D/R in experiment 1 were already at or above the ceiling level of 7.15 dB. In experiment 2a, thresholds were collected at 0 and 10 dB D/R, while in experiments 2b and 2c, thresholds were only collected at 0 dB D/R. This value was chosen because experiment 1 found lowest thresholds at that value, and it thus allows a potentially large increase in JND of the impoverished signals used in experiment 2 (before the threshold level of 7.15 dB is approached). All other procedures were identical to those reported in experiment 1.

## B. Results and discussion

### 1. Experiment 2a: Monaural thresholds

Results from Experiment 2a are shown in Fig. 4. The subject-averaged monaural thresholds (standard error in parentheses) for the slow-onset noise were 2.9 (0.3) and 3.8 (0.2) dB for reference D/R of 0 and 10 dB, respectively. Av-

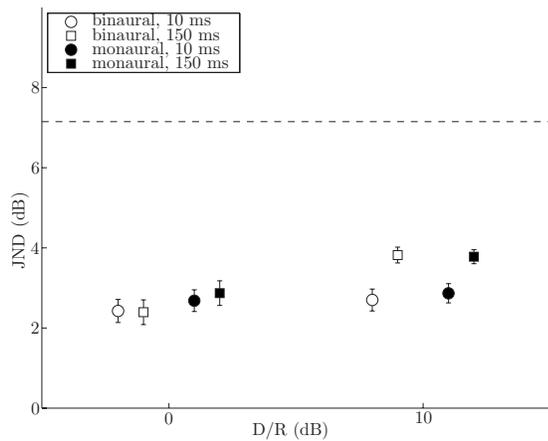


FIG. 4. Mean D/R discrimination thresholds for experiment 2a (monaural, right ear only) at 0 and 10 dB D/R, indicated by filled symbols. Open symbols are thresholds from experiment 1 (binaural) in the same conditions and are included for comparison; error bars indicate standard error. Circles indicate fast-onset noise signal; squares indicate slow-onset noise signal. The dashed line at 7.15 dB indicates the lowest threshold that could be obtained if subjects used level cues only instead of D/R.

average JNDs for the fast-onset noise were 2.7 (0.3) and 2.9 (0.2) dB for the 0 and 10 dB reference D/R, respectively. All thresholds are more than one standard deviation below the ceiling of 7.15 dB and can be assumed reliable indicators of the D/R JND. The data from experiment 1 (0 and 10 dB D/R only) were combined with those from experiment 2a and submitted to a within-subject three-factor ANOVA, with listening mode (binaural versus monaural), reference D/R (0 and 10 dB), and signal type (noise with slow or fast on-off time) as factors. There was a significant main effect of signal type ( $F_{(1,7)}=16.45$ ,  $p=0.01$ ) and D/R ( $F_{(1,7)}=15.31$ ,  $p=0.01$ ). The two-way interaction of signal type and D/R was also significant ( $F_{(1,7)}=7.09$ ,  $p=0.04$ ) due to the tendency for thresholds in the slow-onset noise condition to increase as D/R increased from 0 to 10 dB (by 1.2 dB) but not for the fast-onset noise (0.3 dB, not significant). None of the other two-way (D/R and listening mode, signal type and listening mode) or the three-way interactions were significant ( $p \geq 0.05$ ). As in experiment 1, the effect of onset time on JNDs was not expected based on prior studies (Santarelli *et al.*, 2000; Bronkhorst, 2001; Zahorik, 2002a, 2002c), although the effect is quite small (average over all conditions is 0.8 dB). These results show that monaural listening does not lead to statistically different discrimination thresholds as compared to binaural listening at reference D/R of 0 and 10 dB and for both signal types. If a difference does exist, the 95% confidence interval is  $-0.2$  to  $0.7$  dB (subtracting monaural mean from binaural mean thresholds, pooled over D/R and signal types).

To interpret these findings with respect to our acoustic analysis in Sec. II, we make use of the study of Pollack and Trittipoe (1959), who reported psychometric functions for IACC discrimination at various reference values by using 1 s broadband noise stimuli (we used 300 ms broadband noise convolved with the auditorium impulse response). In the current study, IACC values are about 0.7 at 0 dB D/R and 0.9 at 10 dB D/R (see Fig. 1). In the study by Pollack and Trittipoe (1959), reference IACC values of 0.7 and 0.9 yielded JNDs

of 0.2 and 0.07, respectively. If these IACC JNDs are extrapolated to the current study, it would appear that if listeners utilized IACC as a cue to D/R discrimination, it should have yielded D/R JNDs of approximately 10 dB at a reference D/R of 0 dB (see Fig. 1). Instead, in experiment 1, average JNDs were in the 2–3 dB range at 0 dB D/R. Near the reference D/R of 10 dB, the IACC function saturates at 0.95, and thus the requisite value (i.e., 0.97) cannot be obtained. Seemingly, listeners did not use IACC as a discrimination cue in experiment 1, and hence, thresholds remained unchanged in experiment 2a, when IACC cues were removed. At D/R values outside of the range 0–10 dB, it is not likely that IACC is an important cue; above 10 dB, IACC changes are too small to be discriminable, while below 0 dB, IACC is in the range 0.5–0.7, where discrimination is very poor and would lead to very large ( $>10$  dB) discrimination thresholds for D/R.

Even though binaural cues are not powerful enough to be useful in D/R discrimination in ordinary circumstances, this does not imply that distance perception (for which D/R is thought to be an important cue) is equally effective via monaural or binaural listening. One line of evidence was provided by Bronkhorst (2001), who performed distance judgment experiments with manipulated BRIRs that had a range of IACC values. He showed that distance perception was progressively impaired as IACC approached unity, as distance judgments converged to small values. Based on panel (a) of Fig. 1 we might anticipate such a result, as  $IACC \approx 1$  corresponds to large D/R values (implying small source-receiver distance). Although changes in D/R may be discriminated equally well for monaural as binaural listening, monaural stimuli do not evoke natural distance percepts, and the results we report here cannot be linked directly to tasks involving identification of auditory distance.

## 2. Experiments 2b and 2c: Narrow band thresholds with fixed and roving centers frequency

Mean thresholds from experiments 2b and 2c are shown in Fig. 5, together with those of experiment 1 for the same signal at 0 dB D/R. Note that all data were collected at 0 dB D/R by using only the fast-onset noise. Average thresholds and standard errors (in parentheses) obtained by listeners for the fixed center frequency conditions were 3.6 (0.2) and 4.2 (0.2) dB for center frequencies of 500 and 3000 Hz, respectively. The thresholds for the roved center frequency condition at 500 and 3000 Hz were 4.9 (0.4) and 6.2 (0.4) dB, respectively. The frequency-roved threshold at 3 kHz may be contaminated by direct sound level cues, as it is less than one standard deviations below the ceiling level of 7.15 dB; the other thresholds are all reliable D/R JNDs by our criterion. Similar to experiment 2a, the data obtained in the fixed center frequency condition were compared to those obtained in experiment 1 (0 dB D/R, fast-onset noise, wideband signal). To investigate the effect of spectral content (wideband, low-frequency narrow band, or high-frequency narrow band), the data were submitted to three paired-sample *t* tests (wideband versus low-frequency narrow band, wideband versus high-frequency narrow band, and low-frequency versus high-frequency narrow band) with the Bonferroni correction for

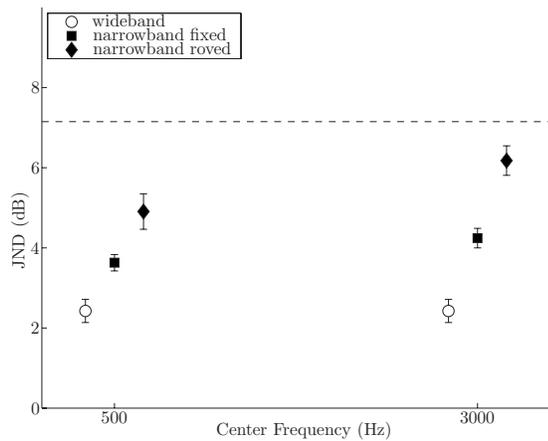


FIG. 5. D/R discrimination thresholds for the narrow band noise signals (10 ms onset/offset) used in experiments 2b and 2c at 0 dB D/R at center frequencies of 500 Hz and 3 kHz (bandwidth: 3 ERB), indicated by filled symbols. Open symbols are thresholds from experiment 1 (WBN) in the same condition, included for comparison. Squares indicate narrow band signals with a fixed center frequency; lozenges indicate narrow band signals with roved center frequency; error bars indicate the standard error. The dashed line at 7.15 dB indicates the lowest threshold that could be obtained if subjects used level cues only instead of D/R.

multiple comparisons. The difference in the mean JNDs between the wideband and both narrow band signals was statistically significant: with the low-frequency narrow band signal, the mean difference was 1.2 dB ( $t_{14}=3.4$ ,  $p=0.0043$ ), and with the high-frequency narrow band signal, the mean difference was 1.8 dB ( $t_{14}=4.8$ ,  $p<10^{-3}$ ). The mean JNDs for the two narrow band signals were not significantly different. As this experiment was designed to investigate the contribution of spectral cues for D/R discrimination, we may conclude that at 0 dB D/R, for fast-onset (10 ms) noise signals, these are important for D/R discrimination.

In order to compare the effect of roving the center frequency, a within-subject ANOVA was conducted with center frequency (500 or 3000 Hz) and spectral envelope (fixed versus roved) as main factors. The analysis revealed a significant main effect of center frequency ( $F_{(1,7)}=55.81$ ,  $p<0.01$ ) and a main effect of spectral envelope ( $F_{(1,7)}=17.45$ ,  $p<0.01$ ). The JNDs for the high-frequency signals were on average 0.9 dB higher than those for the low-frequency signals and the JNDs in the roved conditions were on average 1.6 dB higher than in the fixed conditions. The interaction between the frequency and roving condition was not significant.

To assess the effect of bandwidth reduction on spectral variance cues, it is relevant to consider the number of independent samples of the power spectrum that are available to the brain. This number is limited by the resolving power (frequency selectivity) of the cochlea, such that the number of independent samples in a given frequency band is given by the number of “nonoverlapping” auditory filters in this band, which can be computed by counting how many ERB wide the band is. Experiment 1 used signals with energy in the frequency range of 100–10000 Hz, corresponding to a bandwidth of 32 ERB, whereas experiments 2b and 2c used signals of 3 ERB wide. Thus, spectral variance estimates had available 32 versus 3 independent samples for the wideband

versus the narrow band signals. This reduces the effectiveness of the spectral variance cue by an estimated factor<sup>9</sup> of 7.

To assess the effect of bandwidth reduction on spectral envelope cues, we calculated the spectral CoG as a function of D/R and expressed the slope as a percentage change in CoG per decibel increase in D/R. For the WBN signals from experiment 1, the peak slope value occurs around 0 dB D/R [cf. panel (c), Fig. 1] and equals 1% / dB. For the narrow band signals, the peak value also occurs around 0 dB D/R, but its value is reduced by a factor of about 6–9. For the 500 Hz signal, it is 0.11% / dB, while for the 3 kHz signal, it is about 0.17% / dB. Comparing the two narrow band signals, the low-frequency signal has slightly smaller CoG changes with D/R relative to the high-frequency signal (0.11% / dB versus 0.17% / dB), but according to [Wier et al. \(1977\)](#), discrimination of the relative changes in frequency is slightly better at 500 Hz versus 3 kHz ( $\Delta F/F=0.17$  versus 0.23, respectively). Data pooled from experiments 2b and 2c indicated that JNDs for the high-frequency narrow band signals were significantly higher than the low-frequency JNDs, although the effect is small (mean difference: 0.9 dB). This seems to indicate that although physical changes in spectral envelope (as measured by CoG) are greater for the high-frequency signal, better sensitivity to spectral changes [as measured by pure tone Frequency difference limen (FDLs) relative to the reference frequency] at low frequencies leads to better discrimination performance for the low-frequency narrow band signal. However, given the limited modeling and psychophysical data used, this latter conclusion should be regarded as tentative.

Experiment 2c used the same narrow band signals, but with roved center frequencies, which prevented listeners from relying on changes in spectral envelope, thereby completely removing this cue. This was accompanied by a statistically significant increase in JNDs with respect to the fixed center frequency condition of experiment 2b (mean increase: 1.6 dB). Remaining cues to D/R are IACC, spectral variance, or temporal cues. As discussed with respect to experiment 2a, the obtained JNDs in experiment 2c are too low to be mediated by IACC. Spectral variance and temporal cues were not altered with respect to experiment 2b, so the increase in JND seems to be explained by the complete removal of spectral envelope cues.

The mean difference in JNDs from experiment 2c (narrow band, roved center frequency) and experiment 1 (broadband, 0 dB D/R, fast-onset noise) is 3.1 dB. Because binaural cues were found to be too weak to mediate discrimination in any of our experimental conditions and the signals in experiment 2c had most if not all spectral cues removed, we hypothesize that discrimination in experiment 2c relied primarily on temporal cues (differences in buildup or decay of sound as a function of D/R), which then appear to be relatively weak cues to D/R (mediating JNDs of 5–6 dB). Discrimination in experiment 2b (narrow band, fixed center frequency) probably did *not* primarily use temporal cues because these cues should have been equally effective in experiments 2b and 2c, but thresholds increased by 1.6 dB in experiment 2c versus 2b. Thus, the reason that prior studies have failed in finding strong effects of onset/offset time on

D/R discrimination and distance perception (Santarelli *et al.*, 2000; Bronkhorst, 2001; Zahorik, 2002a, 2002c) may be that in ordinary circumstances, spectral cues are more powerful (at least around 0 dB D/R), so that temporal characteristics are less important. Temporal cues would be important in such tasks when the spectral cues are less powerful or not available.

## V. SUMMARY

We have proposed that D/R discrimination is based on discrimination of several underlying acoustic cues: specifically, we considered IACC, spectral variance, spectral envelope, and buildup/decay time. This proposition is based on the fact that these variables have a monotonic relationship with D/R and have a rapid rate in change around specific D/R values (Figs. 1 and 2). We have considered only the acoustic aspects of these cues as they occur in a typical auditorium, without modeling perceptual sensitivity. Our main finding was that each of these acoustic cues varied over a limited range of D/R values only, and that at large negative and positive D/R, the acoustic variables were constant. This implies that D/R discrimination, as based on these variables, becomes poor at large negative and positive D/Rs.

We measured D/R JNDs at D/R values of  $-10$ ,  $0$ ,  $10$ , and  $20$  dB for noise signals in “ordinary” (full-cue) and at  $0$  and  $+10$  dB in reduced-cue conditions. We found the following:

- D/R JNDs for WBN signals are  $2$ – $3$  dB at  $0$  and  $+10$  dB D/R and at least  $6$ – $8$  dB at  $-10$  and  $20$  dB (experiment 1). The increase in JND at these last two D/R values is consistent with the prediction based on our acoustic analysis.
- Contrary to most prior studies, we found an effect of onset/offset time on D/R JNDs in that fast-onset/offset ( $10$  ms) signals maintain the same JND at  $0$  and  $10$  dB D/R, while slow-onset/offset ( $150$  ms) signals show a JND increase of about  $1$  dB in this range. This may indicate that even in “full-cue” conditions, temporal mechanisms are used to some extent in discriminating D/R.
- D/R discrimination does *not* rely on binaural cues such as IACC because changes in D/R corresponding to a JND produce subthreshold changes in IACC (IACC data from Pollack and Trittipoe, 1959). This was confirmed by experiment 2a, which showed that monaurally obtained JNDs are not statistically different from binaurally obtained JNDs at  $0$  and  $10$  dB D/R.
- Large reductions in frequency bandwidth lead to statistically significant increases (mean increase:  $1.5$  dB) in D/R JND at a reference D/R of  $0$  dB. The effect of bandwidth reduction is to strongly reduce both spectral variance and spectral envelope cues without affecting temporal cues. Therefore, spectral cues are important for obtaining low D/R JNDs in ordinary conditions.
- Complete elimination of spectral envelope cues by roving the center frequency of narrow band signals leads to a further significant increase in JND at a reference D/R of  $0$  dB (mean increase:  $1.6$  dB). For frequency-roved, narrow band signals, temporal cues appear to be the main mechanism listeners use to discriminate D/R. The obtained

thresholds are  $5$ – $6$  dB, i.e.,  $3$  dB higher than those that are obtained when spectral cues are available.

- Narrow band noise signals ( $3$  ERB) elicit larger JNDs at high frequencies versus low frequencies ( $3$  kHz versus  $500$  Hz, mean difference:  $0.9$  dB), which may be due to greater sensitivity to changes in frequency at  $500$  Hz versus  $3$  kHz.

The smallest JNDs we report in full cue conditions (about  $2$ – $3$  dB) are considerably smaller than previously established JNDs (about  $5$ – $6$  dB) in what appear to be broadly similar conditions (Zahorik, 2002c).

Our acoustic analysis reveals the auditory horizon effect, i.e., the tendency to underestimate source distance for far sources (beyond critical distance). Acoustic variables thought to provide cues to D/R do not change appreciably beyond the critical distance, also consistent with the fact that we obtained large D/R JNDs at large negative D/R values.

## ACKNOWLEDGMENTS

Research was supported by grants from the NIDCD of the National Institutes of Health (R21DC-04840), the Beckman Institute, the Charles M. Goodenberger Fund, and Phonak AG. Jont Allen and Bill Hartmann pointed out the spectral variance and interaural cross-correlation as potential cues to D/R. The authors thank the members of the Intelligent Hearing Aid group at the University of Illinois at Urbana-Champaign for comments on the original version of this manuscript. Associate Editor Armin Kohlrausch and three reviewers contributed many insightful comments that improved the quality of the final manuscript.

<sup>1</sup>By using  $h(t)$  to indicate the impulse response between two locations in an enclosure, D/R can be computed as

$$D/R \equiv 10 \log \frac{\int_0^T h^2(\tau) d\tau}{\int_T^\infty h^2(\tau) d\tau},$$

where  $T$  is chosen such that it separates the direct sound from all reflections in the impulse response (typically  $T \approx 2$ – $3$  ms).

<sup>2</sup>The source signal itself may also have spectral variations, but in general, these will be uncorrelated to the room response and therefore have no effect on the change in spectral variance as a function of D/R.

<sup>3</sup>A more comprehensive model for spectral envelope that is also perceptually relevant is the Intensity Weighted Average of Instantaneous Frequency (IWAIF) model, described by Anantharaman *et al.* (1993).

<sup>4</sup>To avoid confusion, we use the terms “onset” and “offset” times to describe the temporal characteristics of the sound source; “buildup” and “decay” times will be used for the ear-canal signals.

<sup>5</sup>In this regard, it is “unfortunate” that the greatest physical changes in all variables we analyzed occur in roughly the same D/R range. This might limit the benefit that could be obtained from the availability of redundant cues. A counterexample where redundant cues operate in different physical regimes is in sound localization, where interaural time versus level differences are most useful in different frequency regions.

<sup>6</sup>A similar argument for large positive D/R values implies that sources well within the critical distance should also be judged closer to it than they in fact are; thus the perceived distance would *overestimate* the actual distance. It is true that near sources are indeed usually judged to be further away than they actually are (e.g., Zahorik, 2002c).

<sup>7</sup>Overall level  $L_O$  equals the logarithm of the sum of the direct and reverberant intensities (not level), i.e.,

$$L_O = 10 \log_{10}(10^{L_D/10} + 10^{L_R/10}),$$

using  $L_D$  and  $L_R$  for direct and reverberant levels in decibel sound pressure level. It is easily verified that by holding  $L_R$  constant, a change  $\Delta L_D$  will lead to  $\Delta L_O < \Delta L_D$ .

<sup>8</sup>The dependence of D/R on source distance is caused by the fact that the energy in the direct sound decays with distance, while the energy of reverberation is approximately constant throughout the entire room (Kuttruff, 1991). Conservation of energy implies that direct sound level decreases by 6 dB for every doubling of the distance  $d_{rs}$  between the source and receiver. Therefore, D/R decreases by 6 dB for every distance doubling, and we can write

$$D/R = -6 \log_2 \left( \frac{d_{rs}}{r_c} \right) = 20 \log \left( \frac{r_c}{d_{rs}} \right),$$

where  $r_c$  is the “critical distance” of the room, defined as the distance where D/R equals 0 dB (equal energy in direct and reverberant sounds).

<sup>9</sup>The variance of a sound power spectrum can be estimated as a stochastic variable, the *sample spectral variance*. With  $N$  independent spectral samples  $x_i$ , we define the sample spectral variance  $s^2$  as

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - m)^2,$$

using  $m$  for the sample mean (which has to be estimated first). Assuming for simplicity that the  $x_i$  are normally distributed, then  $\text{var}(s^2)$ , the variance of the sample variance, is (Kenney and Keeping, 1951)

$$\text{var}(s^2) = \frac{N-1}{N^2} 2\sigma^4,$$

using  $\sigma^2$  for the population spectral variance, which is estimated by  $s^2$ . For wideband signals (experiments 1 and 2a)  $N \approx 32$ ,  $\text{var}(s^2) = 0.061 \sigma^4$ ; the narrow band signals (experiments 2b and 2c) were designed to have  $N = 3$ , and thus  $\text{var}(s^2) = 0.44 \sigma^4$ , an increase by a factor of about 7. The effectiveness of a cue is inversely proportional to its variance, so this bandwidth reduction should diminish the effectiveness of spectral variance cues by approximately a factor of 7.

Allen, J., and Neely, S. (1997). “Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise,” *J. Acoust. Soc. Am.* **102**, 3628–3646.

Anantharaman, J., Krishnamurthy, A., and Feth, L. (1993). “Intensity-weighted average of instantaneous frequency as a model for frequency discrimination,” *J. Acoust. Soc. Am.* **94**, 723–729.

Berkley, D., and Allen, J. (1993). in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by Studebaker and Hochberg (Allyn and Bacon, Boston, MA), Chap. 1.

Blauert, J. (1984). *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd ed. (MIT Press, Cambridge, MA).

Blessner, B. (2001). “An interdisciplinary synthesis of reverberation viewpoints,” *J. Audio Eng. Soc.* **49**, 867–903.

Boehnke, S., Hall, S., and Marquardt, T. (2002). “Detection of static and dynamic changes in interaural correlation,” *J. Acoust. Soc. Am.* **112**, 1617–1626.

Bronkhorst, A. (2001). “Effects of stimulus properties on auditory distance perception in rooms,” in *Proceedings of the 12th ISH: Physiological and Psychological Bases of Auditory Function* (Shaker, Maastricht, The Netherlands).

Bronkhorst, A., and Houtgast, T. (1999). “Auditory distance perception in rooms,” *Nature (London)* **397**, 517–520.

Brungart, D. (1999). “Auditory localization of nearby sources. III. Stimulus effects,” *J. Acoust. Soc. Am.* **106**, 3598–3602.

Brungart, D., Durlach, N., and Rabinowitz, W. (1999). “Auditory localization of nearby sources. II. Localization of a broadband source,” *J. Acoust. Soc. Am.* **106**, 1956–1968.

Brungart, D., and Rabinowitz, W. (1999). “Auditory localization of nearby sources: Head-related transfer functions,” *J. Acoust. Soc. Am.* **106**, 1465–1479.

Butler, R., Levy, E., and Neff, W. (1980). “Apparent distance of sounds recorded in echoic and anechoic chambers,” *J. Exp. Psychol. Hum. Percept. Perform.* **6**, 745–750.

Coleman, P. (1962). “Failure to localize the source distance of an unfamiliar sound,” *J. Acoust. Soc. Am.* **34**, 345–346.

Coleman, P. (1968). “Dual role of frequency spectrum in determination of auditory distance,” *J. Acoust. Soc. Am.* **44**, 631–632.

Ernst, M., and Banks, M. (2002). “Humans integrate visual and haptic in-

formation in a statistically optimal fashion,” *Nature (London)* **415**, 429–433.

Ernst, M., Banks, M., and Bühlhoff, H. (2000). “Touch can change visual slant perception,” *Nat. Neurosci.* **3**, 69–73.

Fechner, G. (1912). *Elements of Psychophysics* (Houghton Mifflin, Boston, MA).

Gabriel, K., and Colburn, H. (1981). “Interaural correlation discrimination. I. Bandwidth and level dependence,” *J. Acoust. Soc. Am.* **69**, 1394–1401.

Gogel, W. (1961). “Convergence as a cue to absolute distance,” *J. Psychol.* **52**, 287–301.

Green, D. (1988). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York, NY).

Hillis, J., Watt, S., Landy, M., and Banks, M. (2004). “Slant from texture and disparity cues: Optimal cue combination,” *J. Vision* **4**, 967–992.

Jetzt, J. (1979). “Critical distance measurements of rooms from the sound energy spectral response,” *J. Acoust. Soc. Am.* **65**, 1204–1211.

Kenney, J., and Keeping, E. (1951). *Mathematics of Statistics*, 2nd ed. (Van Nostrand, New York, NY), pt. 2, pp. 164.

Kewley-Port, D., and Pisoni, D. (1984). “Identification and discrimination of rise time: Is it categorical or noncategorical?,” *J. Acoust. Soc. Am.* **75**, 1168–1176.

Kidd, G., Mason, C., and Green, D. (1986). “Auditory profile analysis of irregular sound spectra,” *J. Acoust. Soc. Am.* **79**, 1045–1053.

Kuttruff, H. (1991). *Room Acoustics*, 3rd ed. (Elsevier, London).

Levitt, H. (1971). “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Am.* **49**, 467–477.

Little, A., Mershon, D., and Cox, P. (1992). “Spectral content as a cue to perceived auditory distance,” *Perception* **21**, 405–416.

Mershon, D., Ballenger, W., Little, A., McMurtry, P., and Buchanan, J. (1989). “Effects of room reflectance and background noise on perceived auditory distance,” *Perception* **18**, 403–416.

Mershon, D., and King, E. (1975). “Intensity and reverberation as factors in the auditory perception of egocentric distance,” *Percept. Psychophys.* **18**, 409–415.

Nábělek, A. (1988). “Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss,” *J. Acoust. Soc. Am.* **84**, 476–484.

Nábělek, A., and Dagenais, P. (1986). “Vowel errors in noise and in reverberation by hearing-impaired listeners,” *J. Acoust. Soc. Am.* **80**, 741–748.

Nielsen, S. (1993). “Auditory distance perception in different rooms,” *J. Audio Eng. Soc.* **41**, 755–770.

Philbeck, J., and Mershon, D. (2002). “Knowledge about typical source output influences perceived auditory distance (L),” *J. Acoust. Soc. Am.* **111**, 1980–1983.

Pollack, I., and Trittipoe, W. (1959). “Binaural listening and interaural cross correlation,” *J. Acoust. Soc. Am.* **31**, 1250–1252.

Reichardt, W., and Schmidt, W. (1966). “Die hörbaren Stufen des Raumeindrucks bei Musik (The audible steps of spatial impression in music performances),” *Acustica* **17**, 175–179.

Rife, D., and Vanderkooy, J. (1989). “Transfer-function measurement with maximum-length sequences,” *J. Audio Eng. Soc.* **37**, 419–443.

Sabine, W. (1962). *Collected Papers on Acoustics*, Dover ed. (Peninsula, Los Altos Hills, CA).

Santarelli, S., Kopčo, N., and Shinn-Cunningham, B. (2000). “Distance judgements of nearby sources in a reverberant room: Effect of stimulus envelope,” *J. Acoust. Soc. Am.* **107**, 2822.

Schroeder, M. (1965). “New method for measuring reverberation time,” *J. Acoust. Soc. Am.* **37**, 409–412.

Shinn-Cunningham, B. (2000). “Distance cues for virtual auditory space,” *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (IEEE, New York)*.

von Békésy, G. (1938). “Über die Entstehung der Entfernungsempfindung beim Hören (On the origin of distance perception in hearing),” *Akust. Z.* **3**, 21–31.

Wier, C., Jesteadt, W., and Green, D. (1977). “Frequency discrimination as a function of frequency and sensation level,” *J. Acoust. Soc. Am.* **61**, 178–184.

Zahorik, P. (2002a). “Assessing auditory distance perception using virtual acoustics,” *J. Acoust. Soc. Am.* **111**, 1832–1846.

Zahorik, P. (2002b). “Auditory display of sound source distance,” in *Proceedings of the International Conference Auditory Display*, Kyoto, Japan.

Zahorik, P. (2002c). “Direct-to-reverberant energy ratio sensitivity,” *J. Acoust. Soc. Am.* **112**, 2110–2117.