

# Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps

RICHARD M. WARREN, KERI RIENER HAINSWORTH,  
BRADLEY S. BRUBAKER, JAMES A. BASHFORD, JR., and ERIC W. HEALY  
*University of Wisconsin, Milwaukee, Wisconsin*

In order to function effectively as a means of communication, speech must be intelligible under the noisy conditions encountered in everyday life. Two types of perceptual synthesis have been reported that can reduce or cancel the effects of masking by extraneous sounds: Phonemic restoration can enhance intelligibility when segments are replaced or masked by noise, and contralateral induction can prevent mislateralization by effectively restoring speech masked at one ear when it is heard in the other. The present study reports a third type of perceptual synthesis induced by noise: enhancement of intelligibility produced by adding noise to spectral gaps. In most of the experiments, the speech stimuli consisted of two widely separated narrow bands of speech (center frequencies of 370 and 6000 Hz, each band having high-pass and low-pass slopes of 115 dB/octave meeting at the center frequency). These very narrow bands effectively reduced the available information to frequency-limited patterns of amplitude fluctuation lacking information concerning formant structure and frequency transitions. When stochastic noise was introduced into the gap separating the two speech bands, intelligibility increased for "everyday" sentences, for sentences that varied in the transitional probability of keywords, and for monosyllabic word lists. Effects produced by systematically varying noise amplitude and noise bandwidth are reported, and the implications of some of the novel effects observed are discussed.

In everyday life, noise often reduces the intelligibility of speech. However, under some conditions studied in the laboratory, the presence of noise can increase intelligibility. Two types of noise-induced restorations have been reported: phonemic restoration and contralateral induction. The present study reports evidence that there is a third type of restoration that can be induced by noise: spectral restoration.

## Phonemic Restoration

When a single phoneme (or an entire syllable) in a sentence is removed and replaced by a broadband extraneous sound, such as white noise, listeners can neither identify the missing fragment nor locate the position of the noise, even when they are informed that a portion of the sentence has been deleted and replaced by noise and are allowed to listen to the stimulus several times (Warren, 1970; Warren & Obusek, 1971; Warren & Sherman, 1974). Also, when the intelligibility of sentences is reduced by deleting multiple segments, filling the gaps with broadband noise makes the sentence seem uninterrupted, and intelligibility is increased (Bashford, Myers, Brubaker, & Warren, 1988;

Bashford, Riener, & Warren, 1992; Bashford & Warren, 1979, 1987a, 1987b; Bashford, Warren, & Brown, 1996; Powers & Wilcox, 1977; Verschuure & Brocaar, 1983).

## Contralateral Induction

When speech is delivered to only one ear, it is normally lateralized to the side of that ear. However, if broadband noise is delivered to the opposite ear, contralateral induction can produce the effect of hearing speech in both ears, so that the monaural speech appears to be located in the medial plane (Warren & Bashford, 1976). Contralateral induction not only prevents mislateralization of a monaurally masked signal, but it also has been linked to an early stage in the binaural processing normally leading to a fused image of the source that is located accurately in space.

## Spectral Restoration

The sounds interfering with speech in everyday life are not always broadband; some spectral regions of speech may be spared, whereas others are obliterated. If comprehension is to be achieved under these conditions, it must be accomplished with limited spectral information. A recent study by Warren, Riener, Bashford, and Brubaker (1995) examined the intelligibility of sentences heard through narrow spectral slits. In one of their experiments employing sentences representing "everyday American speech" (Silverman & Hirsh, 1955), the spectra were restricted to single  $\frac{1}{3}$ -octave bands. Intelligibility scores for the standard keywords in the sentences were better than 95% when center frequencies of the speech bands ranged from 1100 to 2100 Hz. This high intelligibility could not

---

This work was supported by Grant DC00208 from the National Institutes of Health to the first author. A major portion of this paper is based on the second author's doctoral dissertation, an outline of which was presented at the 127th meeting of the Acoustical Society of America, 1994. The authors thank Christopher A. Brown for technical assistance related to this work. Correspondence should be addressed to R. M. Warren, Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201 (e-mail: rmwarren@csd.uwm.edu).

have been predicted from the earlier literature. Although the Articulation Index (American National Standards Institute, 1969) cannot be used to predict the intelligibility of stand-alone bands of speech, the Articulation Index's frequency importance function indicates that each of the individual  $\frac{1}{3}$ -octave bands of speech in the range of 1100–2100 Hz contributes only 9%–11% to the intelligibility of broadband speech (for tables giving the frequency importance values for  $\frac{1}{3}$ -octave bands of sentences, see Pavlovic, 1994, and Studebaker, Pavlovic, & Sherbecoe, 1987). As discussed by Warren et al. (1995), despite the small proportion of the total information carried by the individual  $\frac{1}{3}$ -octave bands, the top-down sentential context combined with the residual spectral information was responsible for the near-perfect intelligibility of the band-passed sentences. When Warren et al. restricted the spectral information still further, to an extremely narrow band consisting only of high-pass and low-pass slopes of 115 dB/octave that intersected at 1500 Hz, intelligibility scores as high as 77% were reported despite the impoverishment of spectral and formant information and the severe spectral tilt along the steep high-pass and low-pass slopes. This study also determined that when the sentence frequencies were restricted to two widely separated narrow bands—one with a center frequency well above the range of maximal intelligibility (6000 Hz) and the other well below this range (370 Hz)—the intelligibility score for the bands presented together was greater than twice the sum of the score of the two bands when presented separately. In these experiments, the regions between the two bands of speech remained silent: the present study was designed to determine whether adding noise to the spectral gap could increase intelligibility through a “spectral restoration,” much as the addition of noise to temporal gaps in speech can increase intelligibility through phonemic restoration.

There have been a few reports that noise can induce the restoration of missing spectral components of nonverbal sounds. Houtgast (1976) reported that the addition of broadband noise at a level slightly below the masked threshold of a single sinusoid resulted in perception of the lower pitch of a missing fundamental that had previously been heard along with that sinusoid. Plomp and Larkin (cited in Plomp, 1981) showed that adjacent harmonics could be deleted from a complex tone without affecting its timbre if listeners had been primed with the intact complex tone and if the narrowband noise replacing the missing region had a level sufficient to mask the harmonics had they been present in the signal.

There is also some evidence that missing spectral components in speech can be restored. Bashford and Warren (1987a) suggested that a “spectral completion effect” may have contributed to the increased intelligibility observed when complementary filtered noise was added to alternating high-pass and low-pass segments of sentences; however, it was pointed out that the experimental design, which was chosen for other purposes, permitted alternative interpretations of this enhancement of intelligibility.

Subsequently, support for spectral restoration was provided by Shriberg (1992). She found that listeners confused front and back vowels when isolated exemplars were low-pass filtered, but that identification became more accurate when noise replaced the missing spectral components of the vowels.

The present study was designed to examine directly the effects upon intelligibility produced by the addition of noise bands to spectral gaps in speech. Experiments 1–3 examined the effects of adding noise to the gap between widely separated, extremely narrow bands of speech—the center frequencies of the bands were 370 and 6000 Hz, and each band had slopes of 115 dB/octave that intersected at the center frequency (resulting in nominal bandwidths of  $\frac{1}{20}$  octave). The purpose of Experiment 1 was to compare the effects of graded levels of noise in the spectral gap upon the intelligibility of Central Institute of the Deaf (CID) “everyday American speech” sentences. The purpose of Experiment 2 was to determine the effect of bandwidth of the added noise upon the intelligibility of these sentences. The purpose of Experiment 3 was to determine the effects of adding noise to the gap upon the intelligibility of three types of sentences (everyday American speech, sentences with high-predictability keywords, and sentences with low-predictability keywords), as well as the intelligibility of monosyllabic word lists.

In Experiments 1–3, the high-frequency speech band was boosted to match the level of the low-frequency band. The stimulus used in Experiment 4 was prepared differently: The broadband sentences were subjected to band-reject filtering that removed frequencies between 300 and 4500 Hz. The remaining spectral components were maintained at their original relative amplitudes, and the effect upon intelligibility produced by replacing the missing speech frequencies with noise was determined.

## EXPERIMENT 1

### Method

**Subjects.** The 140 subjects (seven groups of 20 subjects) were native English speakers recruited from introductory psychology classes at the University of Wisconsin–Milwaukee. All participants reported having normal hearing and were given course credit for participation.

**Stimuli.** The CID sentences used for testing intelligibility were designed originally to represent “everyday American speech” (Silverman & Hirsh, 1955). These sentences are arranged in 10 lists of 10 sentences, with 50 keywords in each list for a total of 500 keywords. They were produced by a male native English speaker at a fundamental frequency of approximately 100 Hz and were recorded digitally on a Macintosh IIfx computer (22-kHz sampling rate with 16-bit precision).

The speech stimuli were presented in the form of two extremely narrow bands ( $\frac{1}{20}$  octave), each having high-pass and low-pass slopes of 115 dB/octave intersecting at 370 Hz (for the low-frequency band) and 6000 Hz (for the high-frequency band). The two bands were prepared by converting the digital recording to analog form and then splitting the signal. Each of the broadband signals was first sent through a filter (Rockland Model 852), with slopes of 48 dB/octave set at a bandwidth of  $\frac{2}{3}$  octave and a center frequency of either 370 or 6000 Hz, and then passed through a final filter (Wavetek/Rockland Model 751A), with the high-pass and

low-pass slopes of 115 dB/octave set to intersect at the center frequency of the first filter (the prefiltering served to increase the signal/noise ratio of the final filtering stage). The 370- and 6000-Hz bands of speech were then recorded simultaneously (in temporal alignment) at 15 ips on separate tracks of an Otari 8-track recorder (Model MX 5050). The amplitude of each speech band was rescaled prior to the final analog recording so that the peaks of each band (measured after playback through the matched Sennheiser HD 250 headphones used to deliver the stimuli diotically to listeners) were within  $\pm 2$  dB of the nominal overall 70-dB presentation level for each sentence (unless otherwise noted, all levels are given in dBA SPL). Amplitude measurements were made with a Brüel & Kjaer Model 2204 sound-level meter operating in its slow response mode. The spectra of all stimuli were examined after headphone transduction using a Hewlett-Packard Model 3561A spectrum analyzer to confirm the reliability of the band-pass filtering and to ensure that the signal contained no spectral artifacts. The noise floor of the tape recording, as measured by the spectrum analyzer, was approximately 30 dB below the peak levels of the speech signal.

The 370- and 6000-Hz speech bands were presented together, and the spectral gap separating them was filled with one of the seven levels of noise shown in Figure 1 (the noise bands did not stop and start with the individual sentences but were on continuously). The five highest levels were bandpass filtered slopes of 115 dB/octave with a Wavetek/Rockland Model 751A filter, and their amplitudes were adjusted before recording so that on playback they would be at 50, 55, 60, 70, or 80 dB when the flanking speech bands were played at average peaks of 67 dB (combined speech level approximating 70 dB). Since increases in the amplitude of noise are accompanied by an increase in the spread of masking, the interpolated bands were recorded with different cutoff frequencies so that masking of the flanking speech bands would be avoided. Kryter (1960) stated that information contributing to the intelligibility of filtered speech might be present for frequencies along the filter slopes that were attenuated by as much as 30 dB (which, coincidentally, corre-

sponded to the noise floor of the recorded signal in the present study). In order to ensure that none of the noise bands presented at various amplitudes masked spectral components within the speech bands, the high-pass and low-pass cutoff frequencies of each of the noise bands were adjusted so that the noise would just mask sinusoidal tones having frequencies and amplitudes corresponding to the 30-dB downpoints based on long-term averages for the low-pass skirt of the 370-Hz speech band (470 Hz) and the high-pass skirt of the 6000-Hz speech band (4200 Hz). This adjustment resulted in the following settings for the levels of interpolated noise: 50-dB level, 550–3800 Hz; 55-dB level, 600–3400 Hz; 60-dB level, 700–3000 Hz; 70-dB level, 880–2400 Hz; 80-dB level, 1100–1900 Hz. Separate groups of subjects were assigned to each of these five conditions, and a sixth group of subjects heard the speech played without the addition of noise. The background noise level within the spectral gap separating the two speech bands was approximately 40 dB. In order to present the listeners with a noise level in the spectral gap that was below the threshold for noise, a seventh group of subjects heard the recorded speech bands after on-line passage through two Wavetek Model 751A filters (slopes of 115 dB/octave): one filter having a low-pass cutoff at 600 Hz, and the other set at a high-pass cutoff of 4000 Hz (the resulting level of noise in the spectral gap had a level of 10 dB).

**Procedure.** Separate groups of 20 subjects were assigned to each of the seven experimental conditions. The 100 CID sentences presented to each listener were grouped in 10 sets of 10 sentences each. Since there was no assurance that these sets, which were designed to be matched in intelligibility when heard broadband, would maintain matched intelligibility when filtered, the order of these sets was blocked and pseudorandomized so that each sentence set occurred twice in each serial position in each of the experimental conditions for each group. Thus, the pooled intelligibility scores for each successive block of 10 sentences were based upon all 100 CID sentences.

The listeners were tested individually while seated in an audiometric chamber with the experimenter. Testing took about 30 min.

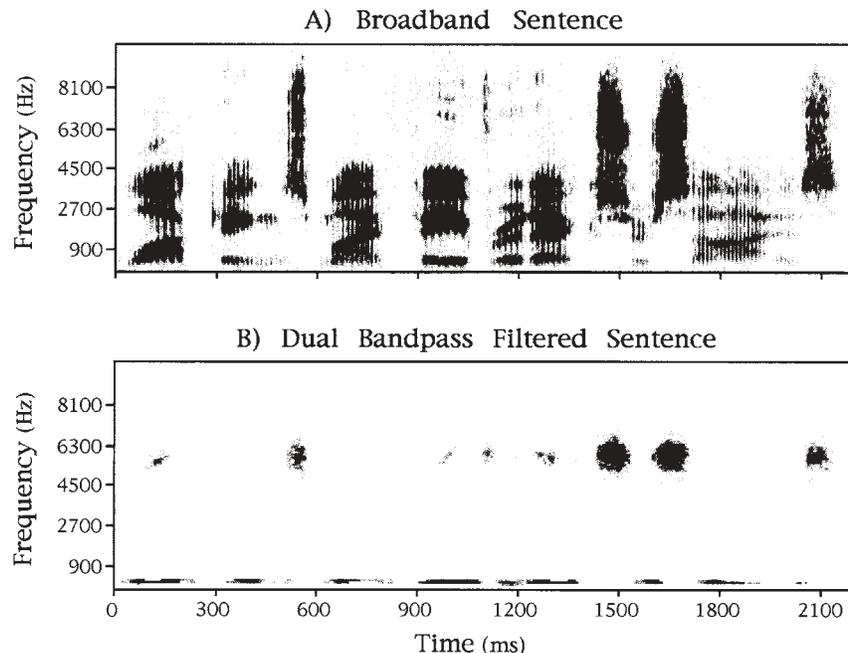


Figure 1. (A) Broadband spectrogram of one of the Central Institute of the Deaf “everyday” sentences, “Walking’s my favorite exercise.” (B) Spectrogram of the same sentence after filtering that produced two widely separated very narrow bands (center frequencies 370 and 6000 Hz, 115-dB/octave filter slopes that intersected at the center frequencies). These bands were presented at equal amplitudes.

Before starting the formal experiment, the subjects were familiarized with the nature of the experimental stimuli by listening to 10 sentences taken from the high-predictability set of the SPIN (Speech Perception in Noise) test of Kalikow, Stevens, and Elliot (1977) (e.g., "Throw out all this useless junk."). These SPIN sentences were presented first broadband and then heard filtered in the same manner along with the same level of interpolated noise as the experimental CID sentences that followed.

Following the initial practice phase of the experiment, each listener was presented with the 100 experimental sentences containing the 500 keywords. The listeners were instructed to repeat each sentence as accurately as possible and were encouraged to guess if uncertain. The experimenter scored the number of keywords reported correctly.

## Results

Figure 2 presents the mean intelligibility scores (and standard error confidence intervals) obtained for the seven experimental conditions, averaged across the 10 blocks of CID sentences. It can be seen that the insertion of noise into the spectral gap separating the two speech bands produced a substantial increase in intelligibility for all conditions. Interestingly, the maximum increase in intelligibility produced by noise occurred at 60 dB. Further increases in noise amplitude produced decreasing intelligibility scores, despite the adjustments that had been made in the noise bandwidths to ensure that peripheral masking of the speech bands did not occur at any of the noise levels; hence, the interference with speech processing at 70 and 80 dB occurred at a central level.

The listeners' repetition accuracy scores were subjected to a two-way (7 noise levels  $\times$  10 blocks of sentences) analysis of variance (ANOVA), which revealed a significant main effect of noise level [ $F(6,133) = 7.66, p < .0001$ ], a significant main effect of block [ $F(9,1197) =$

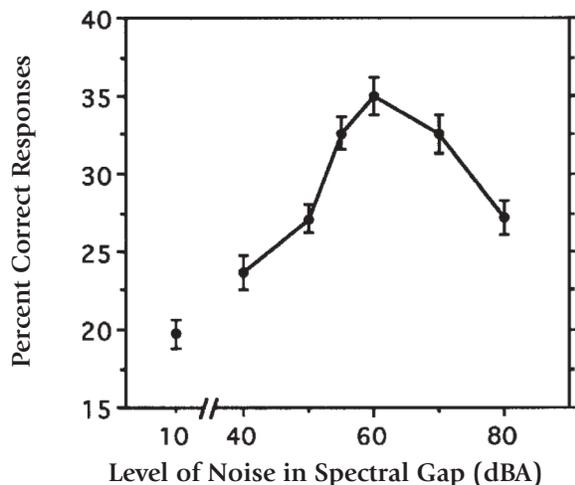


Figure 2. Effects of adding noise at various amplitudes to the spectral gap between two very narrow bands (center frequencies 370 and 6000 Hz) of filtered Central Institute of the Deaf "everyday" sentences. Mean percent accuracy scores and their standard errors are shown for identification of keywords for seven noise-level conditions. Separate groups of 20 listeners were used for each condition.

18.65,  $p < .0001$ ], and a nonsignificant noise  $\times$  block interaction [ $F(54,1197) = 0.55, p > .99$ ]. Analysis of the block effect, using Helmholtz single- $df$  contrasts (Hays, 1988), indicated that repetition accuracy improved consistently across blocks of sentences ( $F \geq 3.87, p < .05$  or better) until reaching an apparent asymptote at the 7th block. This pattern of improvement, which resulted in an intelligibility increase of 14% across blocks, held for all seven noise-level conditions, as indicated by the nonsignificant noise level  $\times$  block interaction. The effect of noise level was assessed using Tukey HSD tests, which indicated that intelligibility was equivalent at SPLs of 55, 60, and 70 dB, averaging 33%, and that those conditions produced significantly higher intelligibility ( $p < .05$ ) than did the 10- and 40-dB conditions, which averaged 22%. The 50- and 80-dB noise conditions, which produced intermediate intelligibility scores averaging 27%, did not differ significantly from the remaining conditions.

## EXPERIMENT 2

The results of Experiment 1 demonstrated that the intelligibility of sentences limited to two widely separated narrow bands can be enhanced by filling the spectral gaps with noise. The purpose of Experiment 2 was to examine the effect upon intelligibility of a systematic reduction of the bandwidth of this interpolated noise, resulting in effective gaps of different widths separating the noise from the flanking speech bands.

### Method

**Subjects.** Sixty additional listeners (three groups of 20 subjects) were recruited from introductory psychology classes at the University of Wisconsin-Milwaukee. The qualifications and compensation procedures employed in Experiment 1 were also employed in Experiment 2.

**Stimuli.** The 100 sentences consisting of the two widely separated narrow bands of speech employed in Experiment 1 were also employed in Experiment 2. As in Experiment 1, the speech bands were presented at a combined level of 70 dB, and the interpolated noise band producing the greatest enhancement of intelligibility in Experiment 1 (700–3000-Hz band of white noise presented at 60 dB, with slopes of 115 dB/octave) was employed again with a different group of 20 subjects. Two additional conditions employed narrower noise bands (880–2400 Hz and 1100–1900 Hz). The reduced bandwidths represented successive  $\frac{2}{3}$ -octave decreases ( $\frac{1}{3}$  octave removed from each end of the band). All noise bands were presented at an equal spectrum level of 26 dB per hertz, corresponding to 60 dBA SPL for the 700–3000-Hz band, 58 dBA SPL for the 880–2400-Hz band, and 55 dBA SPL for the 1100–1900-Hz band. The stimuli were prepared and presented to the listeners using the equipment employed in Experiment 1.

**Procedure.** Separate groups of 20 listeners were assigned to the three conditions. The order of the 10 sets of sentences was pseudorandomized as in Experiment 1 so that each set occurred twice in each serial position across listeners. The familiarization with the SPIN practice sentences, the presentation of the experimental stimuli, and the scoring method were the same as in Experiment 1.

### Results

Figure 3 presents the mean intelligibility scores for the sentences under three interpolated noise conditions. As

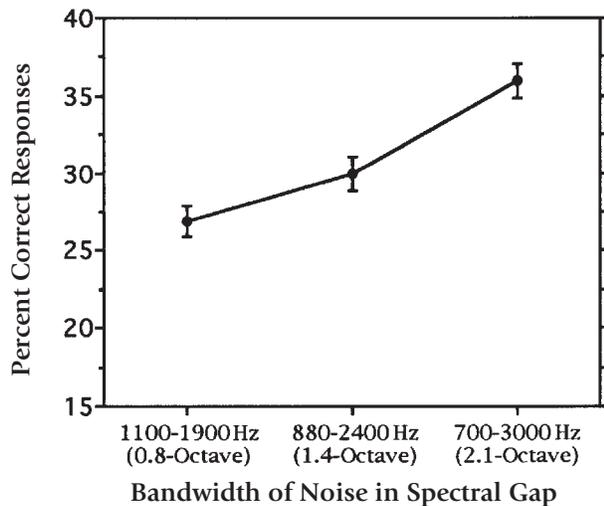


Figure 3. Effects of adding noise at fixed spectrum level and different bandwidths to the spectral gap between the two very narrow bands (center frequencies 370 and 6000 Hz) of filtered Central Institute of the Deaf “everyday” sentences. Mean percent accuracy scores and their standard errors are shown for identification of keywords for three noise-bandwidth conditions. Separate groups of 20 listeners were used for each condition.

shown, intelligibility was greatest when the spectral gap was filled with the widest band of noise.

The sentence data were subjected to a two-way (3 noise bandwidths  $\times$  10 blocks of sentences) ANOVA that yielded a significant effect of bandwidth [ $F(2,57) = 5.46, p < .01$ ] and block [ $F(9,513) = 9.69, p < .0001$ ] and a nonsignificant bandwidth  $\times$  block interaction [ $F(18,513) = 0.81, p > .6$ ]. A subsequent post hoc analysis (Tukey HSD) showed that the bandwidth effect was due to a significant ( $p < .01$ ) enhancement of intelligibility for the 700–3000-Hz condition (36%) over the 1100–1900-Hz condition (27%). The 880–2400-Hz condition produced an intermediate level of restoration (30%), and it did not differ from either the 2.1-octave or the 0.8-octave band of noise ( $p > .05$ ). The 700–3000-Hz condition represents an exact replication of conditions employed with a separate group in Experiment 1, and the 35% intelligibility score in Experiment 1 agrees quite closely with the score of 36% obtained in Experiment 2.

Analysis of the block effect using Helmholtz single-*df* contrasts indicated that repetition accuracy improved consistently across blocks of sentences ( $F \geq 4.04, p < .05$  or better) until reaching an apparent asymptote at the 5th block. This pattern of improvement, which resulted in an intelligibility increase of about 15% across blocks, held for all noise bandwidth conditions, as indicated by the nonsignificant bandwidth  $\times$  block interaction.

### EXPERIMENT 3

Experiment 3 was designed to examine the extent to which intelligibility is restored for different types of

speech stimuli when noise is added to spectral gaps in speech. The CID sentences used in Experiments 1 and 2 were used again. These “everyday” sentences have a high degree of redundant top-down information concerning the identity of the keywords. Two additional sets of sentential stimuli were employed that had been originally designed to test for the influence of top-down semantic cues upon the intelligibility of degraded speech: For one set, the keywords had a high predictability; for the other, the keywords could not be predicted on the basis of the preceding linguistic context. Finally, to examine the extent of spectral restoration for words lacking supralexical semantic, syntactic, and prosodic cues, a set of 500 monosyllabic words were used as stimuli. Four separate groups heard these four types of verbal stimuli without interpolated noise, and an additional four groups of subjects heard the stimuli with interpolated noise.

### Method

**Subjects.** An additional 176 subjects (four groups of 20 subjects and four groups of 24 subjects) were employed. All were native English speakers with no known hearing problems and, as in Experiments 1 and 2, were recruited from the introductory psychology courses at the University of Wisconsin–Milwaukee. They were given course credit for participation.

**Stimuli.** In addition to the 100 CID sentences representing everyday American speech that had been employed in Experiments 1 and 2, there were two additional sets of sentences that provided different levels of transitional probability of keywords and that required the listener to repeat only the monosyllabic keywords rather than an entire sentence. One of these sets consisted of high-predictability sentences and the other of low-predictability sentences originally designed for the SPIN test of Kalikow et al. (1977). These SPIN sentences consist of eight sets of 25 sentences, across which 200 monosyllabic keywords appear as the final words in carrier sentences. A sample sentence from the high-predictability set is: “Throw out all this useless junk.” A sample sentence from the low-predictability set is: “Ruth hopes she called about the junk.” The other set of stimuli consisted of ten 50-item lists of phonemically balanced (PB) monosyllabic words (originally published by Egan, 1948). The isolated words provide no supralexical cues to identification.

The preparation of the filtered CID sentences was described in the Method section of Experiment 1. The other stimuli were produced by the same speaker and prepared in a manner similar to the CID sentences. For all four types of stimuli (“everyday speech” CID sentences, high-predictability SPIN sentences, low-predictability SPIN sentences, and monosyllabic word lists), the two narrow bands of speech ( $1/20$ -octave bands centered at 370 and 6000 Hz) either were mixed with a band of continuous white noise or had no noise added. The noise added to the spectral gaps in the speech was at the optimal level for intelligibility enhancement (60 dB) found in Experiment 1. As in Experiment 1, the noise was band-passed from 700 to 3000 Hz; the speech stimuli were presented at a level of 67 dB for each band, yielding a combined level of approximately 70 dB. The procedure for recording and delivering the stimuli was identical with that employed in Experiment 1.

**Procedure.** Separate groups of 20 listeners were assigned to four of the eight conditions (the CID and PB word conditions, with noise and without noise), and four additional groups, each consisting of 24 subjects, were assigned to the high-predictability and low-predictability SPIN sentence conditions (two conditions with and two conditions without spectrally interpolated noise). The 10 sets of CID sentences and the 10 sets of PB words were presented twice in each serial position across listeners; the eight sets of the high-

predictability SPIN sentences and the eight sets of the low-predictability SPIN sentences were presented three times in each serial position.

Before beginning the formal experiment, the listeners were presented a list of practice stimuli in order to familiarize them with the effects of filtering. The practice stimuli were first presented in broadband form and then filtered in the same manner as the experimental stimuli, with interpolated noise either present or absent, in keeping with the conditions employed for the experimental stimuli of the group. As in Experiments 1 and 2, the listeners were told that these initial stimuli were intended to provide practice for the experiment that would follow. For the groups tested with CID sentences, the practice stimuli consisted of the 10 high-predictability SPIN sentences (Kalikow et al., 1977) used for practice in Experiments 1 and 2. For the groups tested with the high-predictability SPIN sentences and the low-predictability SPIN sentences, the practice stimuli consisted of 10 CID sentences. Finally, the practice stimuli for groups tested with the PB words consisted of 10 words taken from the NU-6 monosyllabic word test (Tillman & Carhart, 1966).

## Results

Figure 4 presents the mean intelligibility scores for the four types of speech with interpolated noise between the dual speech bands either present or absent. As can be seen, intelligibility was enhanced in each case when stochastic noise was added to the spectral gaps. The data are averaged across the 8 blocks of SPIN sentences (for both the high-predictability and the low-predictability sets), and the first 8 blocks of CID sentences and PB word lists. The listeners' scores for these blocks of stimuli were subjected to a three-way (4 stimulus types  $\times$  2 noise conditions  $\times$  8 blocks of stimuli) ANOVA that yielded significant main effects of verbal stimulus type [ $F(3,168) = 115.96, p < .0001$ ], noise condition [ $F(1,168) = 55.43, p < .0001$ ], and block [ $F(7,1176) = 18.98, p < .0001$ ]. The analysis also revealed significant two-way interactions

of stimulus type  $\times$  noise [ $F(3,168) = 3.74, p < .02$ ], stimulus type  $\times$  block [ $F(21,1176) = 3.31, p < .0001$ ], and noise  $\times$  block [ $F(7,1176) = 2.98, p < .01$ ]. The three-way interaction was nonsignificant [ $F(21,1176) = 0.93, p > .6$ ].

Simple effects analysis of the stimulus type  $\times$  noise interaction indicated that the interpolation of noise in the spectral gap between bands improved intelligibility for all speech signals, including PB words (3.5%) [ $F(1,38) = 45.08, p < .0001$ ], low-predictability SPIN sentences (3.7%) [ $F(1,46) = 13.76, p < .001$ ], high-predictability SPIN sentences (7.2%) [ $F(1,46) = 24.99, p < .0001$ ], and the CID sentences (10.5%) [ $F(1,38) = 12.52, p < .002$ ]. The spectral restoration of isolated monosyllabic words contrasts with results obtained from experiments examining phonemic restoration: While addition of stochastic noise to the *spectral* gap increased intelligibility of word lists, previous studies have shown that the addition of noise to *temporal* gaps in word lists does not increase intelligibility (Bashford et al., 1992; Bashford et al., 1996; Dirks & Bower, 1970; Miller & Licklider, 1950).

As in Experiments 1 and 2, and as would be expected, the listeners' performance increased across presentation blocks. The significant stimulus type  $\times$  block interaction was analyzed further using Helmholtz single-*df* contrasts, which indicated that performance reached an apparent asymptote by the 3rd block of stimuli for both the PB words and the low-predictability SPIN sentences but did not asymptote until the 4th block of stimuli for the high-predictability SPIN sentences and the 6th block for the CID sentences. Simple effects analysis of the noise  $\times$  block interaction indicated that interpolation of noise between the speech bands improved intelligibility of the speech stimuli in each of the 8 presentation blocks [ $F(1,174) \geq 4.59, p < .05$  or better]. Helmholtz tests indicated that performance reached an apparent asymptote

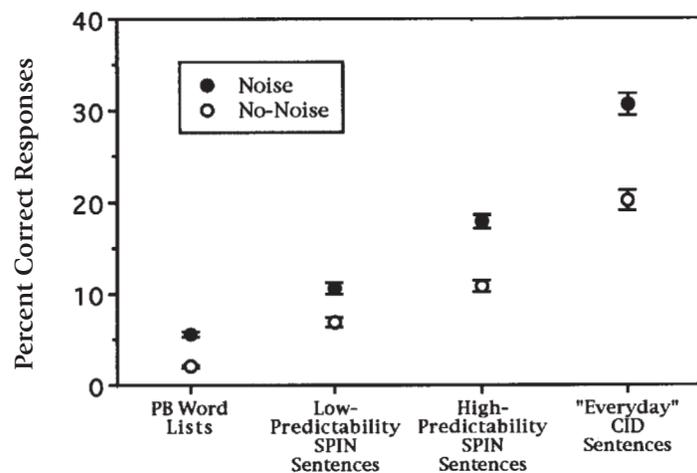


Figure 4. Effects of adding noise at fixed level and bandwidth to the spectral gap between two very narrow bands (center frequencies 370 and 6000 Hz) for four types of speech having different levels of linguistic content. Mean percent accuracy scores and their standard errors are shown for items in word lists and for three types of sentences. Separate groups of either 20 or 24 listeners were used for each condition.

later in the noise condition (Block 5) than in the no-noise condition (Block 4).

#### EXPERIMENT 4

Experiment 4 examined the restoration of intelligibility for CID sentences subjected to band-reject filtering that removed frequencies between 300 and 4500 Hz. The remaining spectral components were presented at their original relative levels, rather than with matched levels as used for the dual narrow bands in Experiments 1–3.

#### Method

**Subjects.** An additional group of 20 subjects was recruited from introductory psychology classes at the University of Wisconsin–Milwaukee. The qualifications and compensation procedures remained the same as in Experiments 1–3.

**Stimuli.** The broadband recording of the CID sentences was subjected to on-line band-reject filtering during this experiment, with filtering cutoffs set to remove frequencies between 300 and 4500 Hz (cascaded Wavetek/Rockland Model 852 filters producing low-pass and high-pass filter slopes of 96 dB/octave). Since the averaged spectrum level of intact speech decreases with frequency, the noise that was added to the spectral gap in the sentences was pink noise (stochastic noise having a spectrum level decreasing 3 dB/octave) that was band-pass filtered from 400 to 3000 Hz using a Wavetek/Rockland Model 751A filter producing slopes of 115 dB/octave. The filtered sentences were presented at a slow-peak level of 75 dB, and the band-passed pink noise (when present) was added at a slow-peak level of 60 dB.

**Procedure.** As in Experiments 1 and 2, the 10 sets of CID sentences presented to each listener were blocked and pseudorandomized in their order of delivery so that each sentence set occurred twice in each serial position across listeners. Noise was added to the spectral gap during odd-numbered blocks of sentences (e.g., Blocks 1, 3, 5, 7, and 9) for half of the listeners and during even-numbered blocks of sentences for the remaining listeners. Thus, unlike the earlier experiments in which separate groups of listeners received the noise-filled and silence-filled gap conditions, the noise variable in Experiment 4 was a within-subjects effect, with noise and silence conditions presented alternately across blocks. Aside from this difference in procedure, the testing of the listeners and scoring were the same as in the earlier experiments.

#### Results

The silent gap condition employed in this experiment produced a much higher intelligibility score (70%) for the CID sentences than did the corresponding dual narrow band condition employed in Experiments 1 and 3 (20%). However, the effects of interpolating noise in the spectral notch were comparable to those observed in the earlier experiments. A preliminary ANOVA examining the effects of noise, block, and group yielded nonsignificant main and interaction effects for the group variable ( $F \leq 2.05$ ,  $p > .10$ ), indicating that the listeners' performance was not substantially affected by the order in which the noise-filled and silence-filled spectral gap conditions alternated across blocks of sentences. Hence, the data were collapsed across the group variable and a noise  $\times$  blocks ANOVA was conducted, yielding significant main effects of noise [ $F(1,19) = 28.3$ ,  $p < .0001$ ] and block [ $F(4,76) = 25.3$ ,  $p < .0001$ ] and a nonsignificant interaction [ $F(4,76) = 1.2$ ,  $p > .30$ ]. Analysis of the block effect, using Helmholtz

single- $df$  contrasts indicated that repetition accuracy improved up to the 3rd block of sentences ( $F > 9.8$ ,  $p < .005$ ). The nonsignificant noise  $\times$  block interaction indicates that this pattern held with both unfilled and noise-filled spectral gaps in the sentences. As indicated by the significant main effect of noise, filling the spectral gap in the band-reject filtered sentences produced a reliable increase in intelligibility, with performance increasing from the average level of 70% correct responses in the silent gap condition to an average level of 84% in the noise-filled gap condition. This 14% enhancement of intelligibility by noise for the band-reject CID sentences having a normal spectral tilt was comparable to the average 10% increase obtained using narrow bands of speech matched in level in Experiments 1–3.

#### SUMMARY AND DISCUSSION

The present study is one of a series reporting that noise enables listeners to effectively restore missing portions of verbal and nonverbal acoustic signals through a process called *auditory induction*. Earlier experiments have found that the addition of noise to temporal gaps in speech produces a linguistic type of temporal induction that was named *phonemic restoration* (Warren, 1970; Warren & Obusek, 1971). Another variety of noise-induced restoration occurs when the noise is presented to one ear while speech is presented to the other: This *contralateral induction* can cause the image of the speech to be located in the medial plane, as if the monaural speech was also present at the ear receiving the noise (Warren & Bashford, 1976). (For a discussion of the literature dealing with restoration of missing sounds, see Warren, 1984, and Warren, Bashford, Healy, & Brubaker, 1994.)

When multiple temporal gaps in sentences reduce intelligibility, filling the gaps with noise produces contextually appropriate phonemic restorations that can make the sentences comprehensible (see Bashford et al., 1992; Bashford & Warren, 1979). The present study was designed to determine whether a spectral restoration of intelligibility occurs when wide frequency gaps in otherwise intact speech are filled with noise. The first three experiments in this study employed filtered speech consisting of two extremely narrow bands (nominal bandwidths of  $1/20$  octave) having steep filter slopes (115 dB/octave). These bands were located at opposite regions of the speech spectrum (370 and 6000 Hz), in ranges that contribute little to the total information content of broadband speech (see American National Standards Institute, 1969). The acoustic cues available in these speech bands consisted essentially of fixed-frequency patterns of amplitude fluctuation lacking information concerning formant structure and frequency transitions.

Experiment 1 employed filtered CID "everyday American speech" sentences, and it was found that placing stochastic noise in the spectral gaps resulted in an appreciable enhancement of the intelligibility of keywords. Various signal-to-noise ratios were employed, and the greatest increase in intelligibility (from 19% to 35%)

was found when the added noise was 10 dB below the slow-peak level of the speech. Intelligibility scores decreased at higher noise amplitudes, dropping to 27% when the noise level was 10 dB above the sentence peaks. This reduction occurred despite the adjustments made in noise bandwidth to ensure that changes in peripheral masking potential could not be responsible for changes in intelligibility. Apparently, there is an optimal level of noise for spectral restoration of intelligibility, with noise above that level producing an "informational masking" (involving higher level processing centers) that is distinct from masking at the peripheral level (see Pollack, 1975, and Watson, Kelly, & Wroton, 1976, for other types of informational masking).

Experiment 2 was designed to determine whether or not optimal spectral restoration of sentence intelligibility requires the concurrent noise to effectively span the frequencies from one speech band to the other. It was found that this was indeed the case: When the bandwidth of interpolated noise was decreased while maintaining the same spectrum level (i.e., decibels per hertz) within the noise band, the intelligibility of sentences was reduced, with the extent of reduction depending upon the extent of the silent region remaining in the gap.

Experiment 3 employed the four types of speech stimuli that had been used previously by Bashford et al. (1992) in a phonemic restoration study examining the enhancement of intelligibility when noise was added to multiple temporal gaps rather than the spectral gaps employed in the present study. These stimuli were (1) phonemically balanced word lists, (2) SPIN sentences with low-probability keywords, (3) SPIN sentences with high-probability keywords, and (4) CID "everyday speech" sentences. In Experiment 3, all four stimuli were reduced to two extremely narrow and widely separated bands prepared in the manner described for the CID sentences employed in Experiments 1 and 2. As shown in Figure 4, intelligibility was increased for all four types of stimuli by the addition of noise to the spectral gap. Similar results were obtained in the phonemic restoration study with one notable exception. The addition of noise to the spectral gaps in word lists increased intelligibility, whereas no increase in intelligibility occurs when noise is added to temporal gaps in monosyllable word lists (Bashford et al., 1992; Dirks & Bower, 1970; Miller & Licklider, 1950). Indeed, one of the experiments in a recent study has found that the addition of broadband stochastic noise at various levels to temporal gaps in broadband monosyllabic word lists consistently reduced intelligibility (Bashford et al., 1996). However, Samuel (1981, 1991) has reported that phonemic restoration does occur with isolated words. This apparent conflict may be resolved if we consider that Samuel's method for studying phonemic restoration did not measure intelligibility, but rather it measured whether listeners could distinguish between conditions in which noise either was added to a phoneme or replaced that phoneme. The inability of listeners to make this distinction (which Samuel attributed to phonemic restoration) appears to be related to what Miller and Licklider (1950)

called the *picket fence effect*. They observed that when noise was added to temporal gaps in word lists, speech appeared to continue through the noise-filled gaps; however, appropriate phonemes were not restored by the addition of noise, and Miller and Licklider were the first to report that intelligibility was no better than when the gaps in the isolated words remained silent.

The stimuli in Experiments 1–3 were reduced to extremely narrow bands centered at 370 and 6000 Hz, with the amplitude of the 6000-Hz band boosted above its normal relative level to match that of the lower frequency band. Experiment 4 was undertaken to examine spectral restoration using a speech stimulus more closely resembling that produced by a speaker. The same CID sentences employed in Experiments 1–3 were subjected to band-reject filtering that produced a spectral gap extending from 300 to 4500 Hz. Fewer spectral components were removed by this procedure than by the filtering procedure employed in Experiments 1–3, and, in addition, the components not subject to filtering maintained their original relative amplitudes. This closer resemblance to natural speech resulted in a considerably higher intelligibility (70%) than that of the dual narrow-band sentences used in Experiments 1–3, and the insertion of noise into the spectral notch increased the intelligibility to 84%, despite the absence or severe distortion of formants and other spectral features having frequencies between 300 and 4500 Hz.

The present study examined only the effects of added noise upon the intelligibility of speech, and it did not determine whether concomitant changes occurred in the noise producing spectral restoration. There have been conflicting suggestions concerning the mechanism by which noise induces phonemic restorations: Warren (1984) suggested that a portion of the noise was reallocated to provide the substrate for perceptual synthesis of the missing speech fragment, and Bregman (1990) suggested that when phonemic restoration occurred, the noise triggered an appropriate gestalt-like closure of the speech gap without any reallocation of the noise's auditory representation. Repp (1992) attempted to resolve this controversy but found evidence for reallocation (changes in the ratings by listeners for the timbre or brightness of the inducing noise) in only two of his five experiments. Subsequently, Warren and colleagues (Warren, Bashford, & Healy, 1992; Warren et al., 1994) used a direct loudness matching procedure to measure changes in the apparent amplitude of the inducing noise. It was found that the apparent amplitude did drop during phonemic restoration, and the extent of this reduction (in decibels) was reported. It remains to be determined whether a portion of the inducing noise is reallocated for perceptual synthesis in spectral restoration as in phonemic restoration. It also remains to be established whether a spectral gap in broadband speech is required for the spectral restoration of intelligibility: can the addition of complementary noise to high-pass speech or to low-pass speech enhance comprehension as suggested by the experiments of Bashford and Warren (1987a) and Shriberg (1992)?

The effectiveness of speech as a method of communication depends upon the great amount of redundant information that is present in the signal produced by speakers. This redundancy permits listeners with mastery of a language to employ a variety of mechanisms and strategies that make comprehension possible when the information present in everyday speech is reduced or degraded. Processing mechanisms leading to comprehension are overlearned and automatic, so that listeners are generally unaware of the masking of portions of speech by noise as long as there is sufficient information remaining to establish what the speaker is saying. Spectral restoration represents only one of several mechanisms available to listeners for minimizing interference when extraneous sounds replace portions of speech.

#### REFERENCES

- AMERICAN NATIONAL STANDARDS INSTITUTE (1969). *American national standard methods for the calculation of the articulation index* [ANSI-S3.5, 1969 (R1986)]. New York: Author.
- BASHFORD, J. A., JR., MEYERS, M. D., BRUBAKER, B. S., & WARREN, R. M. (1988). Illusory continuity of interrupted speech: Speech rate determines durational limits. *Journal of the Acoustical Society of America*, **84**, 1635-1638.
- BASHFORD, J. A., JR., RIENER, K. R., & WARREN, R. M. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Perception & Psychophysics*, **51**, 211-217.
- BASHFORD, J. A., JR., & WARREN, R. M. (1979). Perceptual synthesis of deleted phonemes. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th meeting of the Acoustical Society of America* (pp. 423-426). New York: Acoustical Society of America.
- BASHFORD, J. A., JR., & WARREN, R. M. (1987a). Effects of spectral alternation on the intelligibility of words and sentences. *Perception & Psychophysics*, **42**, 431-438.
- BASHFORD, J. A., JR., & WARREN, R. M. (1987b). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics*, **42**, 114-121.
- BASHFORD, J. A., JR., WARREN, R. M., & BROWN, C. A. (1996). Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. *Perception & Psychophysics*, **58**, 342-350.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- DIRKS, D. D., & BOWER, D. (1970). The role of knowledge-based expectations in music perception: Evidence from musical restoration. *Journal of Experimental Psychology: General*, **119**, 123-144.
- EGAN, J. P. (1948). Articulation testing methods. *Laryngoscope*, **58**, 955-991.
- HAYS, W. L. (1988). *Statistics*. Chicago: Rinehart & Winston.
- HOUTGAST, T. (1976). Subharmonic pitches of a pure tone at low S/N ratio. *Journal of the Acoustical Society of America*, **60**, 405-409.
- KALIKOW, D. N., STEVENS, K. N., & ELLIOT, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, **61**, 1337-1351.
- KRYTER, K. D. (1960). Speech bandwidth compression through spectrum selection. *Journal of the Acoustical Society of America*, **26**, 530-538.
- MILLER, G. A., & LICKLIDER, J. C. R. (1950). The intelligibility of interrupted speech with and without noise. *Journal of the Acoustical Society of America*, **22**, 167-173.
- PAVLOVIC, C. V. (1994). Band importance functions for audiological applications. *Ear & Hearing*, **15**, 100-104.
- PLOMP, R. (1981). Perception of sound signals at low signal-to-noise ratio. In D. J. Getty & J. H. Howard (Eds.), *Auditory and visual pattern recognition* (pp. 27-35). Hillsdale, NJ: Erlbaum.
- POLLACK, I. (1975). Auditory informational masking [Abstract]. *Journal of the Acoustical Society of America*, **57**, S5.
- POWERS, G. L., & WILCOX, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *Journal of the Acoustical Society of America*, **61**, 195-199.
- REPP, B. H. (1992). Perceptual restoration of a "missing" speech sound: Auditory induction or illusion? *Perception & Psychophysics*, **51**, 14-32.
- SAMUEL, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110**, 474-494.
- SAMUEL, A. G. (1991). A further examination of attentional effects in the phonemic restoration illusion. *Quarterly Journal of Experimental Psychology*, **43A**, 679-699.
- SHRIBERG, E. E. (1992). Perceptual restoration of filtered vowels with added noise. *Language & Speech*, **35**, 127-136.
- SILVERMAN, S. R., & HIRSH, I. J. (1955). Problems related to the use of speech in clinical audiometry. *Annals of Otolaryngology & Laryngology*, **64**, 1234-1245.
- STUDEBAKER, G. A., PAVLOVIC, C. V., & SHERBECOE, R. L. (1987). A frequency importance function for continuous discourse. *Journal of the Acoustical Society of America*, **81**, 1130-1138.
- TILLMAN, T. W., & CARHART, R. (1966). *An expanded test for speech discrimination utilizing CVC monosyllabic words* (Northwestern University Auditory Test No. 6). Brooks Air Force Base, TX: USAF School of Aerospace Medicine.
- VERSCHUURE, J., & BROCAAR, M. P. (1983). Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. *Perception & Psychophysics*, **33**, 232-240.
- WARREN, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, **167**, 392-393.
- WARREN, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, **96**, 371-383.
- WARREN, R. M., & BASHFORD, J. A. (1976). Auditory contralateral induction: An early stage in binaural processing. *Perception & Psychophysics*, **20**, 380-386.
- WARREN, R. M., BASHFORD, J. A., JR., & HEALY, E. W. (1992). The subtractive nature of illusory continuity: Reciprocal changes in alternating sounds [Abstract]. *Journal of the Acoustical Society of America*, **91**, 2334.
- WARREN, R. M., BASHFORD, J. A., JR., HEALY, E. W., & BRUBAKER, B. S. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Perception & Psychophysics*, **55**, 313-322.
- WARREN, R. M., & OBUSEK, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, **9**, 358-362.
- WARREN, R. M., RIENER, K. R., BASHFORD, J. A., JR., & BRUBAKER, B. S. (1995). Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Perception & Psychophysics*, **57**, 175-182.
- WARREN, R. M., & SHERMAN, G. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, **16**, 150-156.
- WATSON, C. S., KELLY, W. J., & WROTON, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of uncertainty. *Journal of the Acoustical Society of America*, **60**, 1176-1186.

(Manuscript received June 16, 1995;  
revision accepted for publication April 15, 1996.)