Stream segregation with high spatial acuity

John C. Middlebrooks^{a)} and Zekiye A. Onsan

Department of Otolaryngology, University of California at Irvine, Irvine, California 92697-5310

(Received 26 March 2012; revised 25 September 2012; accepted 12 October 2012)

Spatial hearing is widely regarded as helpful in recognizing a sound amid other competing sounds. It is a matter of debate, however, whether spatial cues contribute to “stream segregation,” which refers to the specific task of assigning multiple interleaved sequences of sounds to their respective sources. The present study employed “rhythmic masking release” as a measure of the spatial acuity of stream segregation. Listeners discriminated between rhythms of noise-burst sequences presented from free-field targets in the presence of interleaved maskers that varied in location. For broadband sounds in the horizontal plane, target-masker separations of $\geq 8^\circ$ permitted rhythm discrimination with $d' \geq 1$; in some cases, such thresholds approached listeners’ minimum audible angles. Thresholds were the same for low-frequency sounds but were substantially wider for high-frequency sounds, suggesting that interaural delays provided higher spatial acuity in this task than did interaural level differences. In the vertical midline, performance varied dramatically as a function of noise-burst duration with median thresholds ranging from $>30^\circ$ for 10-ms bursts to 7.1° for 40-ms bursts. A marked dissociation between minimum audible angles and masking release thresholds across the various pass-band and burst-duration conditions suggests that location discrimination and spatial stream segregation are mediated by distinct auditory mechanisms.

© 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4764879]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.66.Ts [RYL]

Pages: 3896–3911

I. INTRODUCTION

In a complex auditory scene, a listener is confronted with multiple sequences of sounds that must somehow be assigned to their corresponding sound sources (reviewed by Bregman, 1990). This process has been referred to as “stream segregation,” where each “stream” is the perceptual correlate of a sound sequence from a particular source. Successful stream segregation facilitates recognition of signals of interest amid competing sounds. A variety of features of sounds contribute to stream segregation, including fundamental frequency, spectral or temporal envelope, and “ear of entry” (reviewed by Moore and Gockel, 2002).

In his classic reference to the “cocktail party problem,” Cherry (1953) listed “the voices come from different directions” as a major factor, and spatial separation of signal and masker(s) is widely assumed to enhance signal recognition in the presence of competing sounds. Nevertheless, the importance of spatial separation compared to other factors has been questioned as has the spatial acuity, and the roles of various acoustical spatial cues in stream segregation are a matter of some controversy. Some authors have argued that spatial cues make little contribution to stream segregation compared to, for example, fundamental frequency or spectral envelope (e.g., Stainsby *et al.*, 2011). Others have suggested that space is important for segregation of sequences of sounds interleaved in time but less important for separation of temporally overlapping (i.e., concurrent) sounds (e.g.,

Shinn-Cunningham, 2005). Studies of competing speech streams (which include both interleaved and concurrent mixtures of sounds) have demonstrated improved performance with signal/masker separations as narrow as 15° (Marrone *et al.*, 2008), whereas many studies have tested physical separations (or the corresponding interaural differences) no smaller than 60° (e.g., Hartmann and Johnson, 1991; Edmonds and Culling, 2005). Among the various acoustic cues that contribute to spatial hearing, some studies stress the importance of “better-ear” listening or interaural level differences (ILDs) for spatial stream segregation (e.g., Culling *et al.*, 2004; Phillips, 2007), whereas others emphasize interaural time differences (ITDs; e.g., Hartmann and Johnson, 1991; Darwin and Hukin, 1999; Culling *et al.*, 2004).

The present study attempted to isolate the spatial component of stream segregation. We adapted the “rhythmic masking release” procedure that has been described by Turgeon and colleagues (2002) and by Sach and Bailey (2004). In our version of the procedure, the listener was required to discriminate between two temporal patterns (“rhythms”) carried by sequences of brief noise bursts presented from a free-field sound source (the “target”). A similar sequence of noise bursts from a second sound source, the “masker,” was interleaved with the target sequence. Noise bursts from the target and masker never overlapped in time, thereby eliminating concurrent masking. This was an objective measure of stream segregation in which stream segregation enhanced a listener’s performance in a two-alternative task. It contrasts with other objective measures that test whether spatial cues can defeat efforts to fuse two sound streams (e.g., Stainsby *et al.*, 2011). It also contrasts with subjective measures in which the listener is asked whether a sequence of sounds is heard as one or two

^{a)}Author to whom correspondence should be addressed. Also at: Dept. of Neurobiology & Behavior, Dept. of Cognitive Sciences, and Dept. of Biomedical Engineering, University of California at Irvine, Irvine, CA 92697–5310. Electronic mail: j.middle@uci.edu

streams (e.g., contrasted by [Michey and Oxenham, 2010](#)). We employed an objective task using non-verbal stimuli because we wanted to minimize top-down linguistic components of segregation and because we hoped eventually to relate the results to ongoing animal studies.

The results demonstrated robust spatial stream segregation in the horizontal dimension with acuity that approached the limits of listeners' horizontal location discrimination, i.e., their "minimum audible angles." Investigation of specific spatial cues that might permit high-acuity stream segregation demonstrated the importance of interaural difference cues, particularly ITD. Nevertheless, most listeners also could perform the task when target and masker were both located in the vertical midline where interaural difference cues are negligible. The spatial acuity of masking release varied substantially across various pass-band and noise-burst-duration conditions, whereas minimum audible angles were relatively constant across those conditions. That observation raises the hypothesis that spatial stream segregation is accomplished by brain pathways distinct from those responsible for localization discrimination *per se*.

II. GENERAL METHODS

A. Listeners

Paid listeners were recruited from the student body of the University of California at Irvine and from the laboratory staff. They were screened for audiometric thresholds of 15 dB Hearing Level or better at one-octave intervals from 0.25 to 8 kHz. None had experience in formal psychoacoustic experiments. Eighteen of the students and three of the laboratory staff passed the audiometric screening and completed one or more of the three experiments. Listeners each received ≥ 8 h (≥ 1300 trials) of training and practice in their respective tasks with trial-by-trial feedback.

B. Experimental apparatus and stimulus generation

Experiments were conducted in a double-walled sound-attenuating booth (Industrial Acoustics, Inc.) that was lined with 60-mm-thick SONEXone absorbent foam. Room dimensions inside the foam were $2.6\text{ m} \times 2.6\text{ m} \times 2.4\text{ m}$. The booth contained a circular hoop, 1.25 m in radius, that was oriented in the horizontal plane at a height corresponding to that of the seated listener's interaural axis. The horizontal hoop supported 3.5 in. two-way loudspeakers located at azimuths of $\pm 0^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ, 30^\circ$, and 40° ; 0° was straight in front of the listener and negative values were to the listener's left. A similar hoop was oriented in the vertical midline in front of the listener, holding speakers at elevations of $-30^\circ, -20^\circ, -10^\circ, -5^\circ, 0^\circ, 5^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ$, and 50° relative to the horizontal plane containing the listener's interaural axis. In Experiment 3, the listener sat in the center of the hoop, 1.2 m from the speakers. In Experiment 1, the listener sat further back, 2.4 m from the speakers, so that the effective angular spacing of the speakers was 2.5° or 5° . Listeners were instructed to maintain orientation of their heads fixed toward the target source; compliance was checked with

occasional inspection of a video monitor. Experiment 2 was conducted in the same booth, but stimuli were presented through headphones.

Sounds were generated with 50-kHz and 24-bit precision using System III equipment from Tucker-Davis Technologies (TDT; Alachua, FL). The TDT system was interfaced to a personal computer that ran custom MATLAB scripts (The Mathworks, Natick, MA). The responses of the loudspeakers were flattened and equalized such that for each loudspeaker, the standard deviation across the calibrated pass band of the magnitude spectrum was < 1 dB. A precision 1/2 in. microphone (ACO Pacific) was positioned at the usual location of the center of the listener's head, 1.2 or 2.4 m from the speakers, depending on the experiment. Golay codes ([Zhou et al., 1992](#)) were used as calibration probe sounds. The calibration procedure yielded a 1029-tap finite-impulse-response correction filter for each speaker.

Stimuli consisted of 3.2-s sequences of brief noise bursts that were generated in real time by gating a continuous Gaussian noise source generated by the TDT RX6 digital signal processor. The noise presented from each speaker was filtered with the corresponding correction filter and then was band-pass filtered with fourth-order Butterworth filters to the following bands: 400-16 000 Hz for the broad-band condition; 400-1600 Hz for the low-band condition; and 4000-16 000 Hz for the high-band condition. The spectra of the filtered noise bursts differed in detail from burst to burst because of their origin as a continuous Gaussian noise source, but within any particular pass-band condition, the spectral envelopes of all bursts were identical within the limits of accuracy of the calibration and equalization procedure. The noise bursts were gated with raised cosine functions with 1-ms rise/fall times. Burst durations were 20 ms in Experiments 1 and 2 and were 10, 20, and 40 ms in Experiment 3. The base rate of presentation of noise bursts was 10/s, meaning that the shortest interval between bursts was 100 ms, onset to onset. Sounds were presented at 50 dB SPL in the free-field conditions and at a comfortable listening level, around 40 dB sensation level, in the dichotic-stimulation conditions.

C. Rhythmic masking release (RMR) and minimum audible angle (MAA) procedures

We used an objective measure of stream segregation adapted from the RMR procedure that was described by [Turgeon and colleagues \(2002\)](#) for sound sequences that were distinguished by center frequency and by [Sach and Bailey \(2004\)](#) for sequences of 500-Hz tones that were distinguished by interaural difference. In our study, sound sequences were distinguished by source location or by interaural difference. Listeners discriminated between two target rhythms (Rhythm 1 or 2), shown schematically in Fig. 1. An interfering sound sequence (the masker) consisted of a complementary sequence of pulses, such that either a target or masker pulse was presented at each 100-ms interval. When target and masker were co-located, the aggregate signal was a uniform 10/s sequence of noise bursts. The illustrated rhythms were repeated four times without interruption, so

FIG. 1. Schematic representations of rhythms. Rhythms consisted of sequences of 10-, 20-, or 40-ms noise bursts. (A) and (B) show target Rhythms 1 and 2. For each rhythm, the masker pattern was complementary to the target such that the combination of target and masker produced a uniform 10/s pattern. Each panel shows only one instance of each rhythm. In the actual psychophysical task, each rhythm was presented four times without interruption.

that the listener heard 16 target noise bursts and 16 masker bursts on each trial at the uniform 10/s base rate. The RMR procedure used in the previous studies differed from ours in that the timing of masker pulses in those studies was pseudo-random. In pilot tests using pseudo-random maskers, we found that some listeners could distinguish the rhythms even when target and masker were co-located. That problem was eliminated in our experiments by the use of the deterministic masker patterns. Listeners were instructed to attend to the target location although the rhythms could be distinguished by attending to either target or masker.

Target sound sequences all were presented from the loudspeaker positioned at 0° azimuth, 0° elevation. In the 0° target condition of Experiments 1 and 3, the listeners sat facing the target speaker and were instructed to orient their heads to that speaker. In the 40° condition of Experiment 1, listeners sat and oriented their heads so that the target speaker was 40° to their right. The orientation of the listener relative to the target was fixed throughout each ~10-15-min block of trials. The location of the masker source was fixed throughout the sequence of noise bursts constituting each trial, but it varied from trial to trial.

The experimental design was single-interval two-alternative forced choice, requiring the listener to discriminate between the two rhythms. On each trial, Rhythm 1 or 2 was presented (with equal probability), and listeners were instructed to press one of two response buttons to indicate the perceived identity of the rhythm. In the first training trials, target rhythms were presented without maskers, and rhythm identification was learned rapidly. Later in training, and on all data-collection trials, the target sound sequence was interleaved with the complementary masker sequence. After each trial, a light was illuminated next to the correct response button to provide trial-by-trial feedback. Masker locations varied in random order among trials. Each 10- to 15-min block consisted of 135-160 trials (depending on the experiment), and listeners typically completed five to seven blocks in each visit to the laboratory. Each listener completed 51-60 trials (depending on the experiment) for each condition of target and masker location, pass band, etc.

Performance in rhythm discrimination was expressed as a discrimination index (d' ; Green and Swets, 1988). For each condition, trials in which Rhythm 1 was presented and the listener replied “1” were scored as hits, and trials in which Rhythm 2 was presented and the listener replied “1” were scored as false alarms. The proportions of hits and false alarms were converted to standard deviates (i.e., z -scores), and the difference in z -scores gave the discrimination index (d'):

$$d' = z(\text{Hits}) - z(\text{False Alarms}).$$

Values of d' were plotted versus masker location, and the interpolated location at which the plot crossed the criterion of $d' = 1$ was taken as the RMR threshold.

Our rationale for employing this version of RMR as a measure of stream segregation is that when target and masker were separated sufficiently in location, the listener would hear two segregated streams of sound bursts corresponding to target and masker stimuli, thus enabling identification of the rhythm within one or the other stream. An alternative interpretation is that the listener might have heard just one, connected, stream that moved back and forth between target and masker location and could identify the rhythm from the pattern of back-and-forth movements. In pilot studies, we did observe a single connected stream in conditions in which the base rate of sound-burst presentation was 3/s (i.e., >200 ms onset to onset). At a slightly faster base rate, 5/s, it was difficult for listeners to decide whether they were hearing one connected or two segregated streams. At base rates of 7/s and at the 10/s rate that was used for this study, however, listeners were unable to track the back-and-forth movement of source locations and, instead, reported hearing two segregated streams, each at half the aggregate base rate. We also conducted a small subjective experiment with five of the listeners who had completed Experiment 3 (reported in Sec. III A).

The MAA procedure was similar to that of the RMR procedure. For MAA measurements in the horizontal dimension, the listener heard on each trial a single 20-ms noise burst from the fixed target location (referred to here as “A”) followed 280 ms later (i.e., 300 ms onset to onset) by a noise burst from a second location (“B”) to the left or right of A; on half the trials, the order was reversed and B was followed by A. The listener was instructed to press response button 1 or 2 to indicate whether the second burst was to the left or right of the first. Sound A was fixed at 0° or 40° (differing between blocks of trials), and sound B locations varied among trials in random order. The d' and the threshold for correct identification of the change in location (the MAA) were computed as for the RMR threshold. A similar MAA procedure was used for stimuli in the vertical midline (i.e., at azimuth 0°) except that Sound A was fixed at elevation 0°, Sound B varied in elevation, and listeners used response buttons 1 or 2 to indicate whether the second burst was heard below (button 1) or above (button 2) the first. Also, vertical MAAs were tested for 10-, 20-, and 40-ms burst durations in contrast to horizontal MAAs, which were tested only with 20-ms durations. Trial-by-trial feedback was provided.

Distributions of thresholds generally were not normally distributed. Also, in some conditions, one or more of the

listeners could not attain criterion performance within the range of target/masker separations that were tested. For that reason, non-parametric statistics were used for comparisons of median thresholds between conditions.

III. EXPERIMENT 1: HORIZONTAL ACUITY AND SPATIAL CUES

A. Experiment 1A: Azimuth acuity of stream segregation

The goal of Experiment 1A was to estimate the minimum horizontal separation of target and masker at which listeners could segregate sequences of sounds from target and masker locations and thereby discriminate between Rhythms 1 and 2. The interpolated separation that yielded $d' = 1$ was taken as the RMR threshold. Four female and three male listeners completed the experiment; ages ranged from 19 to 22 yr. Stimuli were sequences of 20-ms broadband (400–16 000 Hz) noise bursts that were gated and filtered from ongoing Gaussian noise. In separate blocks, the target was located in the horizontal plane in front of the listener on the listener's midline plane (0° azimuth) or 40° to the right of the listener's midline. Masker locations ranged in horizontal location from -20° to $+20^\circ$ relative to the target in 2.5° or 5° increments. The same seven listeners completed measurements of MAAs in the horizontal plane. The A and B sound locations for the MAA measurements were the same locations in the horizontal plane that were used as target and masker locations, respectively, for the RMR testing.

Listeners demonstrated remarkably high spatial acuity in the horizontal plane in this rhythm discrimination task. An example of the performance of one listener is shown in Fig. 2 for conditions in which the target was located at 0° [Fig. 2(A)] or right 40° [Fig. 2(B)]. In the 0° target condition, performance was around chance levels (i.e., d' near zero) when the masker was also located at 0° . This was the expected result because the aggregate of target and masker stimuli

was simply a train of 10/s sound bursts with nothing to distinguish bursts belonging to the target from bursts belonging to the masker. A shift of the masker by as little as 5° – 10° to the left or right of the target, however, produced a noticeable improvement in rhythm discrimination. Performance at a criterion of $d' = 1$ was obtained in this example with the masker located at interpolated positions of 6° to the left or 9° to the right of the 0° target. Performance improved rapidly with increasing target-masker separations. When the target was at right 40° , threshold target-masker separations were somewhat broader: 9° and 11° in this example.

Task performance by the seven listeners is summarized in Fig. 3. The median values of discrimination thresholds for maskers located to the left or right of the target ($\times$ and $\circ$, respectively) were not significantly different when the target was at 0° ($P = 0.16$; paired Wilcoxon signed rank test), whereas median thresholds were slightly larger for maskers located to the right of the right 40° target ($P < 0.05$). In the broadband stimulus condition represented in Fig. 3, median thresholds (data for left and right maskers combined) were 8.1° and 11.2° for targets at 0° and 40° , respectively. The difference in performance between target locations was significant ($P < 0.0005$, paired signed rank test).

We performed a *post hoc* subjective experiment to test our assumption that listeners' ability to perform rhythm discrimination in the masked condition reflected their perception of target and masker sound sequences as segregated streams. This employed five listeners who had completed Experiment 3. We presented only Rhythm 1 with its complementary masker and on each trial asked the listener to press one of two buttons indicating whether he or she heard one connected stream or heard two segregated half-rate streams. Listeners consistently reported "one stream" for 0° target-masker separation and "two streams" for wide separations. The target-masker separation at which those listeners reported two streams on 50% of trials ranged from 3.5° to 6.1° (median, 4.6°). For each listener, that separation was narrower than his or her corresponding RMR thresholds in azimuth (median RMR thresholds in that cohort: 8.2° , range: 5.9° to 15.2°).

FIG. 2. Rhythm discrimination by one listener. (A) and (B) show conditions with target located at 0° and 40° , respectively. Performance, represented by the discrimination index (d'), improved with increasing separation of target and masker. The dashed line at $d' = 1$ indicates the criterion for threshold rhythm discrimination.

FIG. 3. Distributions of rhythmic masking release (RMR) thresholds for the seven listeners of Experiment 1. Each listener is represented by two symbols, representing thresholds for maskers to the left ($\times$) and right ($\circ$) of the target. A random horizontal offset is added to each symbol to minimize overlap. In each box, horizontal lines represent 25th, 50th, and 75th percentiles.

That indicates that above-threshold rhythm discrimination occurred at target-masker separations at which two segregated streams were heard on more than half of trials.

The spatial acuity observed for the rhythm discrimination task, in some instances, appeared to approach the limits of auditory spatial discrimination. For that reason, we measured minimum audible angles (MAAs) for discrimination of two successive sound sources for the same seven listeners. The measured MAAs are compared with RMR thresholds in Fig. 4; we focus here on the broadband (i.e., 400 to 16 000 Hz) condition [Figs. 4(A) and 4(B)] and will address the other pass-band conditions in Sec. III B. For both target locations, RMR thresholds for each listener were significantly larger than their MAAs ($P < 0.0005$ for both target locations; paired signed rank test). Although RMR thresholds generally were wider than MAAs, one can see instances in Fig. 4 in which the symbols touch the diagonal line that corresponds to RMR thresholds equal to MAAs even among the subjects that exhibited some of the narrowest MAAs. That indicates that some listeners were able to perform spatial stream segregation at the limits of the resolution for auditory spatial cues.

FIG. 4. Comparison of RMR and minimum audible angle (MAA) for seven listeners. Three rows of panels indicate pass-band conditions, as indicated, and left and right columns of panels indicate conditions of target at 0° and 40° , respectively. Triangles indicate cases in which d' for RMR was < 1 at a 30° target/masker separation or MAA was < 1 at a 20° separation. Each listener is represented in each panel by two symbols, for maskers located to the left and right of the target.

B. Experiment 1B: Acoustic cues for spatial stream segregation

We attempted to identify the acoustic cues that listeners utilized for spatial stream segregation. We exploited the property that different types of spatial cues are available in low- versus high-frequency ranges (reviewed by Middlebrooks and Green, 1991). At low frequencies, the predominant cues would be interaural time differences (ITDs), primarily ITDs in temporal fine structure (ITD_{fs} ; e.g., Buell *et al.*, 1991). At higher frequencies, listeners are not sensitive to temporal fine structure (e.g., Licklider *et al.*, 1950), and the wavelengths of sound are equal or shorter than the dimensions of the head and external ears. At those higher frequencies, the dominant binaural cues presumably would be ILDs. Other high-frequency cues might include differences in sound level between target and masker that would be present at each ear (a “head-shadow level effect” considered in Sec. III C), spectral shape cues produced by the direction-dependent filter properties of the head and external ears, and ITDs in sound envelopes (ITD_{env}).

Listeners and conditions were identical to those in Experiment 1A except that in individual blocks, we tested a “low-band” stimulus having a pass band of 400–1600 Hz and a “high-band” stimulus having a pass band of 4–16 kHz. Comparisons among pass-band conditions used a Kruskal–Wallis test, and comparisons between pairs of pass-band conditions used a Bonferroni adjustment for multiple comparisons.

Listeners’ performance with the band-limited stimuli is compared with that for the broadband (400–16 000 Hz) stimulus in Fig. 5. As noted in the preceding text, median thresholds in the broadband condition were 8.1° and 11.2° , respectively, for 0° and 40° target locations. Performance varied significantly across the three pass-band conditions ($P < 0.0005$ for the 0° target and $P < 0.02$ for the 40° target). The effect of pass-band, however, almost entirely reflected

FIG. 5. Distributions of RMR thresholds for various conditions of pass band (3 conditions in each panel) and target location. Symbols represent two data from each of seven listeners. Triangles indicate cases in which d' was < 1 for a target/masker separation of 30° . Other details of plot format are as described for Fig. 3.

the impaired performance in the high-band condition. Elimination of high frequencies (i.e., the low-band condition) resulted in performance that was not significantly different from that for the broadband condition (median thresholds, 5.9° and 13.2° for 0° and 40° target locations, respectively; $P > 0.05$ for pair-wise comparisons of broadband versus low-band, both target locations). In contrast, thresholds in the high-band condition were significantly broader (i.e., worse) than for the broadband condition (high-band median thresholds 15.4° and 20.6° for 0° and 40° target locations; $P < 0.005$ for the 0° target and $P < 0.05$ for the 40° target, pair-wise comparisons).

As was the case for broadband stimuli, a comparison of RMR thresholds with MAAs for low- and high-band stimuli [Figs. 4(C)–4(F)] showed that rhythm thresholds were significantly larger than corresponding MAAs ($P < 0.0001$ for both the 0° and 40° targets, signed rank test across all pass bands). Also similar to the broadband condition, there were several instances in the low-band condition in which rhythm thresholds fell very close to the corresponding MAAs. In contrast, all rhythm thresholds for the high-band condition were appreciably larger than the corresponding MAAs. The MAAs did not differ consistently among pass bands for the 0° target ($P > 0.05$) and, for the 40° target, were only slightly broader in the high-band condition (median 7.8° for high-band compared to 4.8° for broad band; $P < 0.05$). That is despite the RMR thresholds all being significantly larger for the high-band condition than for broadband and low-band conditions, as stated in the preceding text.

C. Experiment 1C: “Head-shadow” level cues

In many situations in which a target must be heard in the presence of a spatially separated masker, acoustic shadowing of the masker by the head can result in a target-to-masker ratio that is higher at one ear than at the other. In such a condition, assignment of greater weight to input from the “better ear” can aid the listener in hearing the target. In the present study, there was no better ear in the usual sense. By design of the experiment, the listener could identify the rhythm by listening to either the target or the masker. Nevertheless, differences in the proximal sound levels of target and masker at one and/or the other ear introduced by the head shadow might have distinguished the target from the masker, thereby permitting successful rhythm discrimination. Note that this putative basis for stream segregation based on sound level differences is a monaural effect not requiring an interaural comparison although the task could have been accomplished by either of the two ears.

We attempted to estimate the contribution of head-shadow level cues to our rhythm discrimination task by disrupting the information carried by sound levels. We did that by introducing burst-by-burst random variation in the levels of the sound bursts that constituted the rhythmic patterns. The level variation (in dB) followed a Gaussian distribution that was quantified by the standard deviation of level, σ , tested parametrically at $\sigma = 0$ (i.e., no variation), 2 dB, and 4 dB. The burst-by-burst level variation would have disrupted any spatial cues derived from proximal sound levels

differing between target and masker, whereas they should have had no effect on ITDs, ILDs, or spectral shape cues.

Experiment 1C was completed by five of the seven listeners that completed Experiments 1A and 1B. The stimuli were rhythmic sequences of 20-ms noise bursts, identical to those used in Experiment 1B except for the random level variation. The pass bands were broadband, low-band, or high-band as indicated.

An example of the ability of one listener to discriminate rhythms on the basis of level differences alone in the broadband condition is shown in Fig. 6. In that demonstration, target and masker were co-located at 0°, and mean differences in sound levels were the only cues for segregating target and masker. When there was no burst-by-burst level variation (upper line with circles), this listener reached criterion rhythm discrimination at a target-to-masker mean level difference of 3 dB; the median across five listeners in this task was 3.0 dB and the range was 2.5–4.3 dB. Introduction of a burst-by-burst level variation with $\sigma = 2$ dB increased the target-to-masker ratio needed for criterion performance by 1.4 dB (median threshold shift across five listeners: 1.4 dB; range: 0.5–2.8 dB), and a 4-dB level variation made the task impossible for this listener across the range of target-to-masker ratios that was tested (median threshold shift relative to the no-level-variation condition across five listeners: 5.3 dB; range: 2.6 to > 6 dB). We estimated that at the threshold target/masker azimuth separations in our usual task, the level differences would have been no more than a few decibels. For that reason, we predicted that introduction of a 2- or 4-dB level variation would introduce a substantial impairment in performance if the listeners were relying on level differences to do the task.

Distributions of listeners’ performance in broadband, low-band, and high-band conditions with varying amounts of level variation are shown in Fig. 7. In some cases, one can see what appears to be a slight trend toward increasing thresholds associated with greater level variation, but that

FIG. 6. Rhythm discrimination by one listener based on sound level alone. Target and masker were co-located at 0°. The three curves distinguished by symbol type represent three amounts of burst-by-burst variation in sound level as indicated. Performance generally improved with decreasing masker level (i.e., with increasing difference between target and masker levels) and was impaired by increasing level standard deviation.

trend was not significant for any pass-band or target-location condition ($P = 0.14\text{--}0.94$ depending on pass-band and on target location; Kruskal–Wallis test). It was not surprising that there was little or no effect of level variation in the broadband and low-band conditions. Any level effect would have been greater at high than at low frequencies, and tests of the low-band stimulus in Experiment 1B indicated that

elimination of high-frequency cues did not impair performance compared to broadband conditions. The lack of a significant impact of level variation in the high-band condition [Figs. 7(E) and 7(F)] is particularly interesting. It demonstrates that the performance that was obtained in the absence of low-frequency ITD_{fs} cues was not impaired by disruption of monaural level cues, suggesting that stream segregation in the higher frequency range relies on interaural difference cues.

IV. EXPERIMENT 2: INTERAURAL DIFFERENCE CUES AT HIGH FREQUENCIES

Studies of sound localization demonstrate that ILD's are the dominant cue for localization of high-frequency sounds (e.g., Strutt, 1907; Macpherson and Middlebrooks, 2002). Nevertheless, listeners can detect interaural differences in the timing of the envelopes of high-frequency sounds (ITD_{env}), raising the possibility that ITD_{env} cues might have contributed to performance in our spatial segregation task. We evaluated the relative contributions of ILD and ITD_{env} cues to stream segregation by adapting the rhythm discrimination task to conditions of high-band (4–16 kHz) dichotic stimuli presented through headphones.

Experiment 2 was completed by eight listeners (ages 19–26 yr, five female); seven of those listeners had previously completed Experiment 3, which measured RMR thresholds with free-field stimuli varied in azimuth and elevation. High-band stimuli were presented through Sennheiser HD-265 circumaural headphones at a comfortable listening level. Aside from the dichotic presentation, the sequences of target and masker sounds were identical to the high-band stimuli used in Experiment 1B: target and masker each consisted of 16 20-ms bursts at an aggregate rate of 10/s. The target was always presented with 0 dB ILD and 0 μ s ITD. The ILD and ITD of the masker were varied independently among trials; the signs of all non-zero values of ILD and ITD were such that sounds were lateralized toward the listener's right side. The ITDs of the entire waveforms were varied, including both ITD_{env} and ITD_{fs}, but we assume that listeners were insensitive to fine-structure delays in these 4- to 16-kHz noise bands. Tested ITDs ranged from 0 to 700 μ s; 700 μ s is approximately the maximum envelope ITD available to an average-sized listener from a far-field source (e.g., Middlebrooks and Green, 1990). Tested ILDs ranged from 0 to 8 dB; they are smaller than the maximum ≥ 20 dB ILDs available to a listener from a far-field source at frequencies ≥ 4 kHz (e.g., Shaw, 1974).

When masker ILD and ITD were small, listeners generally heard a uniform train of noise bursts localized near the center of the head. Sufficient increases in the interaural difference of the masker tended to distinguish masker from target, permitting discrimination of the rhythm. Generally, a large interaural difference produced an intracranial image of the masker displaced to the right side of the midline, but detection of change in perceived lateralization was not a requirement of the task.

Figures 8(A) and 8(B) show individual examples of rhythm discrimination as a function of masker ILD; these examples represent the range from the most sensitive

FIG. 7. Distributions of RMR thresholds for various conditions of varying level, pass band, and target location. Plot format is as described for Fig. 3.

FIG. 8. Rhythm discrimination under headphones. The panels represent the most (A) and least (B) sensitive subject and the mean across nine listeners (C). The curves indicated by symbols represent conditions of various fixed interaural time differences. Error bars are standard errors of the means.

[Fig. 8(A)] to around the least sensitive [Fig. 8(B)] individual listener. Means and standard errors across all eight listeners are shown in Fig. 8(C). The various lines and symbols indicate various parametric values of ITD. In all cases, discrimination performance was at chance, around $d' = 0$, when the masker ILD was 0 dB. That indicates that no amount of ITD alone was sufficient to permit segregation of target and masker noise-burst sequences when target and masker ILD were equal. Introduction of a 4- or 8-dB ILD in the masker, however, consistently resulted in above-threshold rhythm discrimination. Generally, the lines representing large ITD values lie above the low-ITD lines, indicating that there was some synergy between ILD and ITD_{env} cues. Across all conditions, a two-way analysis of variance showed a strong main effect of ILD ($P < 0.0001$) but a main effect of ITD that only approached significance ($P = 0.066$) and no significant two-way interaction ($P = 0.78$). When ceiling and floor effects were minimized by limiting the two-way analysis of variance to ILD's of 4 and 8 dB, however, there were significant main effects of both ILD ($P < 0.0001$) and ITD ($P < 0.02$); again, there was no significant interaction ($P = 0.68$). Overall, the headphone results support the hypothesis that ILD is the dominant binaural cue for spatial stream segregation of high-frequency sounds but indicate that there could be some small additional contribution from ITD_{env} cues.

V. EXPERIMENT 3: SPATIAL STREAM SEGREGATION IN THE VERTICAL MIDLINE

Our measures of stream segregation with target and masker separated in horizontal location demonstrated that the task was accomplished primarily by spatial information provided by interaural difference cues: predominantly by ITD_{fs} in low-frequency sounds and, when low frequencies were not present, by ILD in high-frequency sounds with some contribution of ITD_{env} . It is well known that ITD and ILD signal the locations of sound sources in the horizontal dimension (e.g., Strutt, 1907; Middlebrooks and Green, 1991), and it might be that our listeners were segregating streams on the basis of perceived target and masker location, irrespective of particular acoustic cues. Alternatively, it might have been that the interaural difference cues themselves formed the basis for stream segregation and that localization *per se* was neither necessary nor sufficient. We attempted to distinguish the influences of location from those of interaural cues by testing spatial stream segregation for targets and maskers separated by elevation in

the vertical midline. In the vertical midline, sound sources cannot be discriminated by interaural difference cues, but listeners can localize sounds on the basis of spectral shape cues provided by the direction-dependent filtering by the head and external ears (reviewed by Middlebrooks and Green, 1991).

Pilot experiments demonstrated that most subjects were unable to do the RMR task when the stimulus patterns consisted of sequences of noise bursts that were 10 ms in duration. For that reason, we tested three conditions that differed in the durations of the noise bursts that constituted the stimulus sequences: Experiments 3A–3C used noise bursts that were 10, 20, and 40 ms in duration, respectively; 20 ms was the burst duration that was used for tests in the horizontal dimension in Experiments 1 and 2. Ten (9 female), listeners completed Experiment 3A, and nine (seven female) completed Experiments 3B and 3C. Six listeners completed all three series. None of these listeners participated in Experiment 1. In all three conditions, the target was fixed at 0° azimuth and elevation, and the masker was presented from varying locations in the horizontal plane at 0° elevation or from varying locations in the vertical midline at 0° azimuth; horizontal-plane and vertical-midline masker locations were varied among trials in random order. The purpose of testing locations in the horizontal plane was to test whether the performance of the listeners in Experiment 3 overlapped with those in Experiment 1 and to evaluate the influence of burst duration on RMR thresholds in azimuth. For the RMR measurements, subjects sat 1.2 m from the loudspeakers. Vertical locations of maskers in Experiment 3A were -10°, 0°, 10°, 20°, and 30°. In Experiments B and C, vertical locations were -30°, -20°, -10°, -5°, 0°, 5°, 10°, 20°, 30°, 40°, and 50°. In Experiments 3B and 3C, the results showed no significant difference between upward (i.e., masker located above the target) and downward RMR thresholds ($P > 0.05$; signed rank test). For that reason, data from upward and downward maskers were combined in the distributions, and each listener was represented in the illustrations and in statistical tests by two data points; in Experiment 3A, only upward maskers were tested with a tall enough distribution of masker locations. We also measured MAA in the vertical midline for the three noise-burst durations. For the MAA measurements using 10, 20-, and 40-ms noise-burst durations, listeners sat 2.4 m from the speakers, resulting in speaker separations of 2.5° and 5°. Sound A was fixed at 0° elevation and sound B ranged from -15° to +25°. The MAA measurements with 10-ms bursts were conducted after

FIG. 9. Distribution of RMR thresholds in the horizontal dimension for various experiment blocks and for stimuli consisting of patterns of 10-, 20-, or 40-ms noise bursts.

Experiment 3C and they were completed only by the six listeners who completed Experiment 3A and who were also were available at the end of the study.

The distributions of RMR thresholds in the horizontal dimension obtained in Experiments 3A, 3B, and 3C were similar to each other and corresponded closely with those obtained with a different population of listeners in Experiment 1 (Fig. 9). The medians varied across the four groups that are illustrated ($P < 0.0005$; Kruskal–Wallis). Pair-wise comparison with Bonferonni adjustment, however, showed that the significant difference was due to the threshold for the 40-ms condition being lower than that for any of the other conditions ($P < 0.05$). There were no significant pair-wise differences between any of the conditions with burst durations of 10 and 20 ms ($P > 0.05$).

In contrast to the situation in azimuth, rhythmic masking release in elevation differed markedly among the three burst-duration conditions in Experiments 3A, 3B, and 3C. The performance of one listener in all three conditions is shown in Fig. 10. In the 10-ms-burst condition, that listener was unable to reach the criterion of $d' = 1$ for any target/masker separation up to 30° , which was the greatest separation tested in Experiment 3A. Performance improved in the 20-ms-burst condition, although it hovered around the $d' = 1$ criterion across a wide range of negative masker elevations. Only for the 40-ms-burst condition did performance increase reliably with increasing target/masker separation. Results from all the listeners are summarized in Fig. 11(A); Fig. 11 also shows summary data from the 20-ms azimuth conditions in Experiment 1 for comparison. In elevation, performance was consistently poor for the 10-ms elevation condition: 7 of 10 listeners failed to reach the $d' = 1$ criterion for target/masker separations up to 30° . In the 20-ms-burst-duration condition, six of the nine listeners did well in the vertical 20-ms condition with RMR thresholds of 13.5° or narrower for maskers located above and/or below the target, whereas the other three listeners could barely do the task with upward and/or downward RMR thresholds $> 30^\circ$. The broad distribution of thresholds in the 20-ms-burst condition was consistent with the tendency of listeners to show somewhat inconsistent performance for that burst duration, as was the case for the example in Fig. 10(B). In the 40-ms-burst condition, all the listeners could do the task, with RMR thresholds tightly

FIG. 10. Rhythm discrimination in the vertical midline by one listener for stimuli consisting of patterns of noise bursts that were 10-ms (A), 20-ms (B), or 40-ms (C) in duration.

clustered around a median threshold of 7.1° . Distributions of RMR thresholds varied significantly with burst duration ($P < 0.0001$, Kruskal–Wallis), and all pairs of burst-duration conditions were significantly different ($P < 0.05$ for 10- versus 20-ms and for 20- versus 40-ms; $P < 0.01$ for 10- versus 40-ms; pair-wise comparisons with Bonferonni adjustment).

We also measured MAAs in the vertical dimension for various stimulus burst durations [Fig. 11(B)]; again, summary data from the 20-ms azimuth conditions are shown for comparison. Contrary to the situation for RMR, MAAs were relatively insensitive to stimulus burst duration; note the difference in vertical scales between Figs. 11(A) and 11(B). There was a significant narrowing of vertical MAAs with increasing burst duration ($P < 0.005$, Kruskal–Wallis) but that was due to the significant difference between the 10- and 40-ms conditions ($P < 0.005$; pair-comparison with Bonferonni adjustment); all other paired comparisons showed no significant difference ($P > 0.05$).

VI. GENERAL DISCUSSION

Listeners could segregate interleaved sequences of otherwise-identical sounds on the basis of differing locations of target and masker. In some cases, spatial acuity in the RMR paradigm approached the localization acuity represented by the listeners' MAAs. We begin this section by considering the present results in the context of previous objective measures of stream segregation that did or did not show strong spatial or binaural sensitivity. We then evaluate the spatial cues that listeners might use for spatial stream segregation, and we consider the possible influence of the reported sluggish temporal dynamics of binaural systems. We consider performance in terms of particular cues versus location discrimination, and we discuss what that might tell us about brain systems for spatial hearing. Finally, we consider the implications of the present results for design of auditory prostheses.

A. Obligatory and voluntary spatial stream segregation

Somewhat surprisingly, many previous studies employing objective measures of stream segregation have failed to

FIG. 11. Summary of RMR thresholds (A) and MAAs (B) for three pass-band conditions in azimuth and three burst-duration conditions in elevation, as indicated. Plot format is as described for Fig. 3.

show robust segregation on the basis of differences in target and masker location or binaural cue. The measures that demonstrated little or no spatial effect include detection of a silent gap between two sound markers (referred to here as A and B) that could differ in dichotic laterality or in free-field location and detection of temporal asymmetry of a sound burst (B) positioned in time between two marker sounds (A), again with A and B possibly differing in laterality or location (e.g., Phillips, 2007; Stainsby *et al.*, 2011). In those studies, a substantial increase in threshold, indicating perceptual segregation of A and B sounds, was observed only when A and B sounds were presented to opposite ears. This “ear of entry” difference, of course, would have produced a conspicuous difference in the perceived laterality of A and B, but it also would have restricted A and B to differing peripheral channels. Application of an ITD of $500\ \mu\text{s}$ or more, opposite in sign to A and B sounds, also would have produced large laterality differences, but the effects of such large ITDs on gap detection or temporal asymmetry detection either were not significant or were significant only for ITDs larger than the range that a listener would encounter with natural free-field stimuli (Phillips, 2007; Stainsby *et al.*, 2011). For free-field stimuli, gap thresholds were elevated (i.e., A and B are segregated) only when the A and B sources were located in opposite sound hemifields or when A was located on the midline and B was located far to the side. When both sources were in the same hemifield, A-B separations of 60° or larger had little influence on gap threshold (e.g., Phillips, 2007).

The gap-detection and temporal-asymmetry-detection results demonstrate perceptual stream segregation primarily in conditions in which the two sound sources were processed by different peripheral channels (i.e., different ears) or were represented, presumably, in different cortical hemispheres. Those studies, however, all tested what has been called *obligatory streaming*. In a test of obligatory streaming, good performance requires the listener to *fuse* information between the two (A and B) sounds to, for instance, evaluate the time between offset of A and onset of B. When perceptual differences

between A and B are sufficiently large, A and B are forced into separate perceptual streams, and evaluation of timing between A and B is impaired. The available data indicate that location or binaural cues produce obligatory streaming only in conditions in which those cues send A and B sounds largely to separate peripheral channels or to separate cortical hemispheres.

In contrast to obligatory streaming, *voluntary streaming* is observed in experimental tasks that encourage the listener to *segregate* perceptual streams, as in the present study. Most tests of speech-on-speech masking contain some element of voluntary streaming, as discussed in Sec. VIB. Three other studies using speech or musical stimuli specifically addressed voluntary stream segregation. Darwin and Hukin (1999) tested the influence of ITD in binding successive words in a sentence. Two carrier sentences were presented simultaneously with differing ITDs. Within each sentence, target words, “bird” and “dog,” were presented so that each word matched in ITD one or the other of the carrier sentences. The listener was asked to attend to the sentence “Could you please write the word X down now” (where X could be bird or dog) and to report which target word was heard with that sentence. On a substantial majority of trials, listeners reported the target word for which the ITD matched that of the attended sentence when the targets (and sentences) differed in ITD by as little as $90\ \mu\text{s}$. That result predicts voluntary segregation of speech streams at free-field separations of $\sim 10^\circ$. Hartmann and Johnson (1991) asked listeners to identify pairs of simple melodies that were played simultaneously. In various conditions, the two melodies were distinguished by a variety of spectral, temporal, and interaural parameters that enabled varying degrees of voluntary stream segregation. Lateralization of the two melodies by application of ITDs of $\pm 500\ \mu\text{s}$ resulted in high levels of correct identification of the melodies that were only slightly (and not significantly) lower than the strongest factor, which was ear of entry (i.e., one melody presented to each ear); smaller ITDs were not tested in that study. Hartmann and Johnson concluded that “peripheral channeling is of

paramount importance” in [voluntary] stream segregation but suggested that the definition of periphery be broadened to include “low-level binaural interaction.” [Saupe and colleagues \(2010\)](#) presented musical phrases simultaneously from three locations; the phrases were synthesized with timbres characteristic of three different musical instruments. Listeners were asked to detect a large descending pitch interval in the phrase played by a particular target. A spatial separation of 28° resulted in substantial improvement in performances compared to a condition of co-located sources. Those authors discussed possible contributions of better-ear listening, binaural unmasking, and external-ear-related amplification to account for their results.

Our objective measure of stream segregation was inspired by the report by [Sach and Bailey \(2004\)](#). That study used sequences of 500-Hz tone pips presented through headphones. Listeners attempted to discriminate between two target rhythms in the presence of masker pulses that varied randomly in time. Performance was at chance when target and masker both had 0 dB ILD and 0 μ s ITD, but performance improved when an ITD of around 100–200 μ s (depending on the listener) was applied to the masker. In far-field stimulus conditions, an ITD of 200 μ s in a 500-Hz tone would correspond to a location slightly less than 15° off the midline ([Kuhn, 1977](#); his Fig. 11, empirical clothed torso condition). For comparison, listeners in the present study showed a median horizontal RMR threshold of 8.1° . Sach and Bailey tested another condition in which a 4-dB ILD was applied to the target and the masker ILD was maintained at 0 dB; an ILD of 4 dB at 500 Hz would correspond to a far-field location $\sim 30^\circ$ off the midline ([Shaw, 1974](#)). In that condition, discrimination of the target rhythms was nearly perfect when the masker ITD was -600 to +300 μ s. When the masker ITD was +600 μ s, however, discrimination was reduced to random-chance levels. The 4-dB target ILD and the 600- μ s masker ITD likely would have lateralized both target and masker toward the right headphone. The conclusion by Sach and Bailey was that the equal lateralization of target and masker defeated segregation of the target and masker streams. It is difficult to interpret the Sach and Bailey result in terms of far-field sound sources because in the condition in which stream segregation was poor, listeners were exposed to spatially implausible stimuli: a target with midline ITD and $\sim 30^\circ$ (4-dB) ILD compared with a masker with far-lateral ITD and midline (0-dB) ILD. It is clear from the results, however, that the listeners somehow confused target and masker in that condition, whereas a 4-dB difference in ILD between target and masker produced stream segregation in other masker ITD conditions.

B. Spatial segregation of speech streams

Speech recognition in typical multi-talker environments and in most speech-on-speech masking experiments challenges the listener with masking by concurrent sounds as well as potential confusion among interleaved sequences of sounds. Spatial separation of target and masker(s) can markedly improve recognition in such situations (e.g., [Edmonds and Culling, 2005](#); [Ihlefild and Shinn-Cunningham, 2008](#);

[Marrone et al., 2008](#)). Many reports of improvement of speech recognition by spatial conditions have emphasized the phenomenon of spatial release from concurrent masking (e.g., [Edmonds and Culling, 2005](#); [Marrone et al., 2008](#)), whereas others have noted that spatial cues can enhance streaming of successive sounds from the same talker ([Ihlefild and Shinn-Cunningham, 2008](#); [Kidd et al., 2008](#)). We consider here some of the studies that emphasized spatial and binaural effects on streaming.

A study by [Ihlefild and Shinn-Cunningham \(2008\)](#) examined patterns of errors in a speech-on-speech masking paradigm. A 90° separation of target and masker sources was particularly effective in reducing errors in which the listener confused elements of the target and masker sentences; that is, spatial separation reduced the percentage of errors in which the listener reported words from the masker instead of from the target. Conversely, the spatial configuration had relatively little effect on the rate of errors in which the listener reported a word that was contained in neither target nor masker. That result is consistent with the view that spatial separation of target and masker aided in linking the successive words in the target sentences. Kidd and colleagues ([Kidd et al., 2008](#)) constructed sentence stimuli in which target and masker words alternated in time, such that there was no temporal overlap between target and masker. That study also demonstrated a strong effect of ITD (in the range of 150–750 μ s) as a “linkage variable” that aided the listener in streaming together the words of the target sentence.

A study by [Marrone and colleagues \(2008\)](#) evaluated the spatial acuity of spatial release from masking. The target was located at 0° , and two maskers were positioned symmetrically to the left and right. Threshold target/masker ratio decreased substantially when the target/masker separation was increased from 0° to $\pm 15^\circ$ and showed a small further decrease for separations of $\pm 45^\circ$ and $\pm 90^\circ$. The authors fit a rounded exponential filter to the data, estimating a 3-dB spatial release from masking at a target/masker separation averaging 8.7° across listeners. That scale of spatial resolution agrees well with the spatial resolution that we obtained with our non-speech RMR paradigm. Note that the speech-on-speech stimulus set used by Marrone and colleagues presumably included some combination of concurrent masking and streaming of successive words, whereas our experiment design avoided temporal overlap of target and masker pulses.

C. Acoustic cues for stream segregation

The locations of sound sources are not mapped directly at the auditory periphery. Instead, locations must be computed by the central auditory system from spatial cues that arise from the interaction of incident sound with the head and external ears (reviewed by [Middlebrooks and Green, 1991](#)). The spatial cues that are available differ between low (below ~ 1.5 kHz) and high (primarily above ~ 3 kHz) frequencies and between horizontal and vertical dimensions, and we exploited those properties to attempt to identify the spatial cues that listeners might use for spatial stream segregation. As noted in the preceding text, previous studies of obligatory stream segregation show robust spatial or binaural effects only in conditions in

which masker and target are largely segregated to different peripheral channels. For that reason, our discussion of spatial cues will focus on voluntary stream segregation.

In the present study, RMR thresholds in the horizontal dimension obtained with low-band sounds were not significantly different from those obtained with broadband sounds, whereas high-band thresholds were substantially wider. Interaural time differences, primarily ITD_{fs} (e.g., Buell *et al.*, 1991), are essentially the only spatial cues that are available in the 400- to 1600-Hz frequency range for a far-field sound in anechoic conditions. Our result indicates that low-frequency cues alone can enable stream segregation with higher spatial acuity than is possible with any of the high-frequency cues. For that reason, we conclude that in broadband conditions, listeners appear to rely on the ITD_{fs} cues that are available at low frequencies. In a speech-on-speech masking task, Kidd and colleagues (2010) also showed that spatial release from masking was greater for low- than for mid- or high-frequency bands, although in that study the broadband stimulus gave the best masking release of all; as in most speech tasks, that study measured some combination of concurrent masking and stream segregation. The dominance of low-frequency ITD cues in our RMR task and in the speech study by the Kidd group mirrors the dominance of ITD_{fs} in localization judgments (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002).

Listeners in the present study showed robust spatial stream segregation for high-frequency sounds, albeit with coarser spatial acuity than for low frequencies. The classic work by Lord Rayleigh (Strutt, 1907), which has come to be known as the “duplex theory,” demonstrated that ILDs are the dominant cues for localization of high-frequency pure tones; ILD dominance has been confirmed for the high-frequency components of broadband sounds (Macpherson and Middlebrooks, 2002). In the present Experiment 1B, the median RMR threshold for the 0° target in the high-band condition was 14.6° . The ILD produced by a masker at 14.6° would be 5-10 dB across most of the range of our 4- to 16-kHz high-band stimulus (e.g., Shaw, 1974). Those ILDs are substantially larger than the ILD detection thresholds of 0.5-1 dB obtained under headphones (Mills, 1960). That the high-frequency RMR thresholds corresponded to ILDs that were well above ILD detection thresholds accords with our observation that high-frequency RMR thresholds tended to be considerably larger than high-frequency MAAs.

Another potential spatial cue at high frequencies would be the interaural delays in the envelopes of waveforms (ITD_{env}). Macpherson and Middlebrooks (2002) showed that for high-pass sounds, listeners give substantially less weight in their localization judgments to ITD_{env} than to ILD. That study also showed, however, that adding amplitude modulation to the stimulus waveforms tended to increase the weight given to ITD_{env} ; those amplitude-modulated waveforms had envelopes similar to those of the stimuli in the present study. In Experiment 2 of the present study, sounds were presented through headphones so that high-frequency ILD and ITD_{env} could be manipulated independently. Stream segregation in high frequencies clearly was dominated by ILD, but synergy with ITD_{env} was evident.

In the present study, performance for the 40° target was nearly as good as that for the 0° target (median RMR thresholds of 11.2° at 40° compared with 8.1° at 0°). Several authors have argued that fine spatial acuity around the frontal midline might be accomplished by comparison between populations of hemifield-tuned neurons in the two cerebral hemispheres (e.g., Stecker *et al.*, 2005; Phillips, 2007). In the 40° condition in the present study, however, the target and all the maskers were located in the same (right) hemifield, and those locations presumably were represented primarily by activity in the left cortical hemisphere. Based on the results obtained in that stimulus configuration, one must entertain the possibility that listeners achieved reasonably good stream segregation by relying on computations within a single hemisphere. The small but significant difference in RMR thresholds measured for maskers on the right (further from the midline) side of the 40° target than on the left side is consistent with the general decrease in the spatial gradient of ITDs and ILDs associated with greater distance from the midline (e.g., Kuhn, 1977; Shaw, 1974).

Most studies of spatial release from masking have emphasized the “head-shadow effect” in which the head casts a shadow such that the target-to-noise ratio at the ear closer to the target (and at which the masker is attenuated) is greater than at the far ear. As described in the Results (Sec. III C), our experimental design eliminated the head-shadow effect in the usual sense of one ear having a better target-to-noise ratio because the test rhythms could be discriminated by attending either to the target or the masker. There was, however, a related sound-level cue at high frequencies consisting of alternating higher and lower sound levels from the direct-path target and the head-shadowed masker. The head-shadow level effect, if it contributed at all to our RMR task, clearly had poorer spatial acuity than did ITD for the reason that the head shadow would be greatest at high frequencies, whereas elimination of high frequencies (i.e., the low-band condition) had no effect on RMR thresholds. At high frequencies, disruption of level cues by introduction of burst-by-burst level variation had no significant effect on high-frequency RMR thresholds. For that reason, we conclude that listeners relied more on binaural cues, in that case ILDs, than on the head-shadow level cue. We note that our experimental situation was unnatural in that useful information could be gleaned both from the target and from the masker. In a more typical situation, there is a genuine signal of interest that must be heard above uninformative noises. For that reason, we might have underestimated the importance of head-shadow cues. On the other hand, our experimental design exaggerated the information that might have been available from level cues by holding target and masker source levels constant, at least in the conditions in which no burst-by-burst level variation was introduced (i.e., the $\sigma = 0$ condition). More typically, target and masker vary in level from syllable to syllable, for instance, as in our varying-level conditions. Our results suggest that in most everyday situations, binaural ILDs would offer higher-resolution cues for stream segregation in the horizontal dimension than would the head-shadow level effect.

When target and masker both were presented from the vertical midline, differences in binaural cues between target

and masker would have been negligible. The principal spatial cues for broadband sound sources located in the vertical midline are the spectral envelopes resulting from the direction-dependent filtering of the broadband source spectrum by the acoustical properties of the head and external ears (reviewed by [Middlebrooks and Green, 1991](#)). Given a noise-burst duration of 40 ms (or even 20 ms for some listeners), listeners could utilize spectral cues to perform the RMR task in the vertical dimension about as well as they could with a full complement of binaural cues in the horizontal dimension. Performance under conditions of noise bursts as short as 10 ms, however, was substantially worse. We consider the influence of burst duration on vertical stream segregation in the next section.

D. “Sluggish” spatial hearing

The binaural auditory system is widely regarded as “sluggish” in its responses to changing binaural parameters. [Blauert \(1972\)](#), for instance, found that listeners showed a “lag” in their ability to track “in detail” excursions of binaural parameters: performance decayed when the durations of excursions were less than an average of 207 ms for ITD and less than 162 ms for ILD. Similarly, [Grantham and Wightman \(1978\)](#) reported that thresholds for detection of varying ITD increased markedly with increases in the rate of variation from 2.5 to 20 Hz. In the present experiment, the rate of presentation of noise bursts (10/s) was selected on the basis of previous reports of the rate dependence of stream segregation (e.g., [Bregman, 1990](#)). Also, that rate corresponds approximately with the rate of phoneme production in typical speech. Our choice of presentation rate, however, is in the range in which binaural sluggishness would have limited listeners’ ability to track the changing locations of noise bursts. The RMR task did not require explicit tracking of locations, but it is reasonable to assume that binaural sluggishness might have influenced performance in the RMR task. In our MAA task, the time from sound onset to onset was 300 ms in contrast to the 100-ms onset-to-onset time in the RMR task. Considering binaural sluggishness, the longer time in the MAA task might account in part for the observations that, on average, MAAs were narrower than were RMR thresholds. Binaural sluggishness cannot account entirely for the differences between MAAs and RMR thresholds, however. Specifically, that difference was considerably greater for high-band (presumably ILD-based) conditions than for low-band (ITD-based) conditions, whereas [Blauert \(1972\)](#) reported that ILD sensitivity showed less sluggishness than did ITD sensitivity.

We are not aware of explicit tests of the ability of listeners to track changes in vertical location. We observed in the present study, however, that vertical RMR thresholds narrowed markedly with increases in noise-burst duration from 10 to 40 ms. There have been previous reports of an influence of noise-burst duration on the accuracy with which listeners could identify the locations of sounds presented in the vertical midline ([Hartmann and Rakerd, 1993](#); [Hofman and Van Opstal, 1998](#); [Macpherson and Middlebrooks, 2000](#)). Those studies showed impaired vertical localization when

stimuli were brief (i.e., clicks or 3-ms noise bursts) and were presented at high sound levels (>45 dB sensation level). [Hofman and Van Opstal \(1998\)](#) showed vertical localization improving with increases in burst duration from 3 ms to up to 40 or 80 ms. That range of durations encompassed the range over which we observed improving RMR thresholds. The strong influence of burst duration in the present study was somewhat unexpected for the reason that our sound levels were generally around 40-dB sensation level, below the level at which localization was impaired in the previous studies; [Hofman and Van Opstal \(1998\)](#) used an A-weighted level of 70 dB, and [Macpherson and Middlebrooks \(2000\)](#) showed performance only beginning to decay with levels increasing above 45 dB sensation level. Also, the MAAs in the vertical dimension measured in the present study were largely insensitive to noise-burst duration, indicating that listeners were able to discriminate locations in all of the tested burst-duration conditions. Spatial hearing based only on spectral cues generally is less robust than that based on interaural difference cues (e.g. see review by [Middlebrooks and Green, 1991](#)). Further study is needed, but our interpretation based on our present understanding is that performance that is already challenged by brief stimuli can break down when further challenged by stimuli that vary rapidly in location.

E. Discrete pathways for location discrimination and spatial stream segregation?

The present results illustrate some characteristics of spatial stream segregation that are shared with the discrimination of locations of static sound sources (i.e., the MAA task), but there are other characteristics that differ. Location discrimination and spatial stream segregation almost certainly derive input from common brain stem nuclei that analyze binaural and spectral-shape cues, but it is not clear whether location discrimination and stream segregation involve common or discrete forebrain pathways. We offer here two working hypotheses. Our “common” hypothesis is that spatial cues are integrated into a common spatial representation that is utilized by both the MAA and RMR tasks (and, presumably any other spatial-hearing task). We note that numerous auditory cortex studies in animals suggest that such a spatial representation (or representations) would be highly distributed rather than in the form of a point-to-point map (reviewed by [King and Middlebrooks, 2011](#)). Targets would be represented, hypothetically, in some sort of spatial coordinates rather than by cue (e.g., ITD in μ s, ILD in dB) values. That means that various spatial tasks might differ in acuity among conditions that emphasize differing spatial cues, but the common hypothesis predicts that such cue dependence would be correlated across spatial tasks. In favor of the common hypothesis is the observation that spatial stream segregation can be observed in the horizontal and vertical dimensions even though the spatial cues are very different, i.e., binaural cues for horizontal and spectral shape cues for vertical. A contrary observation is that RMR thresholds varied dramatically across various pass-band and burst-duration conditions even though MAAs were relatively constant across those conditions—the results violate the prediction of

correlated acuity across spatial tasks. That failure in correlation between MAAs and RMR thresholds might be accounted for in part by binaural sluggishness as considered in the preceding text. As noted, however, the difference between RMR thresholds in the high- and low-band conditions is opposite to the predictions based on [Blauert's \(1972\)](#) observations of lags in ILD and ITD tracking.

The “discrete” hypothesis, an alternative that we favor, is that spatial stream segregation involves one or more brain pathways that are distinct from the pathway(s) utilized for location discrimination. One could account for the wide variation in RMR thresholds across conditions by speculating that these stream-segregation pathways do not utilize the representations of various spatial cues as uniformly and as efficiently as does the putative localization pathway. The Discrete hypothesis receives support from the observations from multiple laboratories that neurons in multiple cortical areas show qualitatively similar spatial sensitivity (reviewed by [King and Middlebrooks, 2011](#)) yet functional inactivation of only a subset of those cortical areas leads to localization deficits (e.g., [Malhotra et al., 2004](#)). We speculate that the spatial information that is available in some of the “non-localization” cortical areas might be used for spatial stream segregation or for other spatial functions not requiring overt localization. Two speech-on-speech masking studies that used dichotic stimulation support the notion that localization is not necessary for spatial stream segregation ([Edmonds and Culling, 2005](#)). In those studies, masker ITDs and ILDs were 0 μ s and 0 dB. The target had ITDs and/or ILDs in two different frequency bands. Essentially the same spatial unmasking of speech by ITD and/or ILD was observed whether the interaural differences in the two bands signaled the same or opposite hemifields. That is, spatial release from masking, which presumably involved some component of spatial stream segregation, was dissociated from localization. Another speech study could account for spatial release from masking in the free field entirely in terms of better-ear listening and binaural (ITD) unmasking “without involving sound localization” ([Culling et al., 2004](#)).

F. Implications for auditory prostheses

Cochlear-implant users show an advantage in sound localization and in speech reception in noise in binaural compared to monaural conditions (e.g., [van Hoesel and Tyler, 2003](#); [Litovsky et al., 2009](#)); present-day binaural configurations include bilateral (i.e., two cochlear implants) and bimodal (one cochlear implant and one hearing aid). Localization by those users in the horizontal dimension is thought to rely primarily on ILD cues (e.g., [Van Hoesel and Tyler, 2003](#)). Speech reception in noise shows benefits from binaural summation and the head-shadow effect. The “squench effect,” which reflects binaural unmasking on the basis of ILD and ITD, is relatively weak in those users, varying among studies and among listeners from little to no significant effect (e.g., [van Hoesel and Tyler, 2003](#)).

The present results emphasize the importance of ITD_{fs} cues for spatial stream segregation. Unfortunately, cochlear-implant users show little or no sensitivity to ITD_{fs}. That is

because the large majority of cochlear-implant speech processors transmit only envelope timing, replacing the temporal fine structure of sound with that of a fixed-rate electrical pulse train (i.e., [Wilson et al., 1991](#)). Moreover, even in laboratory studies with special processors, sensitivity to temporal fine structure in implant users is impaired (e.g., [Zeng, 2002](#)). A recent study in animals ([Middlebrooks and Snyder, 2010](#)) has demonstrated that low-frequency brain stem pathways transmit temporal fine structure with greater acuity than do mid- and high-frequency pathways, which normally receive temporal information only from waveform envelopes. That result suggests that the poor ITD_{fs} sensitivity of implant users might be due to the failure of present-day implant to achieve frequency-specific stimulation of low-frequency pathways.

The present results indicate that head-shadow level effects, which would be available to bilateral or bimodal implant users, might provide some stream segregation but indicate that the spatial resolution from that cue is coarser than that available from either ILD or ITD. The ILD cue potentially could support spatial stream segregation in prosthesis users, but effective use of ILD cues would require careful equalization of the automatic gain control and compression functions of the left- and right-ear devices.

The present demonstration of ITD_{fs} as the dominant cue for stream segregation, combined with the poor sensitivity of cochlear-implant users to temporal fine structure, could help explain the difficulties that implant users have in speech recognition in complex auditory scenes. There are efforts underway to improve the use of ITD_{fs} cues by implant users: by transmitting temporal fine structure through processors (e.g., [Arnoldner et al., 2007](#)), by use of pseudo-monophasic waveforms intended to stimulate the apical cochlea ([Macherey et al., 2011](#)), and by use of penetrating auditory-nerve electrodes that could stimulate apical fibers selectively ([Middlebrooks and Snyder, 2010](#)). The present demonstration of the importance of ITD_{fs} cues for spatial stream segregation offers further motivation for those efforts.

VII. SUMMARY AND CONCLUSIONS

This study tested in isolation the ability of listeners to use spatial cues to segregate temporally interleaved sequences of sounds. This stream segregation is an important component of hearing in complex auditory scenes. Small target/masker separations produced robust spatial stream segregation, with thresholds <10° in the horizontal plane, that is demonstrated both by performance in the RMR task and by reports of the subjective experience of one or two streams. Compared to broadband stimulus conditions, elimination of high frequencies had no effect on segregation thresholds, whereas elimination of low frequencies produced a marked degradation in performance. Those results suggest the high-spatial-acuity performance that we observed was based on listeners' use of interaural delays in temporal fine structure (ITD_{fs}). Our procedure tested “voluntary” stream segregation in which a listener could exploit spatial stream segregation to improve his or her performance as is the case when stream segregation improves speech reception in a complex

auditory scene. Previous measures, of “obligatory” stream segregation, have shown little or no influence of ITD_{fs}.

The results demonstrated a dissociation between location-discrimination acuity and the spatial acuity of stream segregation. The acuity of localization discrimination, represented by MAAs, was roughly equal for broad-, low-, and high-band conditions in the horizontal plane and across all conditions of varying noise-burst durations in the vertical midline. Nevertheless, performance in the RMR task varied markedly across those conditions. The results indicate that stream segregation is based on differences in the spatial cues corresponding to target and masker locations not necessarily on perceived difference in locations of target and masker. We favor the hypothesis that the various spatial cues are detected in the brain stem and that those brain stem nuclei project to one or more central pathways for spatial stream segregation that are distinct from those for location discrimination.

Cochlear implant users who have two implants (“bilateral”) or one implant and one hearing aid (“bimodal”) show a binaural benefit for recognition of speech in noise that is primarily due to a head-shadow effect. They show little or no binaural unmasking based on ITD or ILD (i.e., no “binaural squelch”). Our demonstration of the importance of ITD_{fs} cues for spatial stream segregation combined with previous demonstrations of impaired ITD_{fs} sensitivity by implant users might account for the limited binaural unmasking in implant users. The present results give additional motivation to efforts to improve fine-structure sensitivity in users of auditory prostheses.

ACKNOWLEDGMENTS

We thank Ewan Macpherson for useful conversations concerning the selection of the RMR measure and Peter Bremen, Lauren Javier, Virginia Richards, Justin Yao, and Shen Yi for helpful comments on the manuscript. This work was funded by NIH Grant No. RO1 DC000420.

- Arnoldner, C., Riss, D., Brunner, M., Baumgartner, W. D., and Hamzavi, J. S. (2007). “Speech and music perception with the new fine structure speech coding strategy: Preliminary results,” *Acta Otolaryngol.* **127**, 1298–1303.
- Blauert, J. (1972). “On the lag of lateralization caused by interaural time and intensity differences,” *Audiology* **11**, 265–270.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 1–773.
- Buell, T. N., Trahiotis, C., and Bernstein, L. R. (1991). “Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays,” *J. Acoust. Soc. Am.* **90**, 3077–3085.
- Cherry, C. E. (1953). “Some experiments on the recognition of speech, with one and two ears,” *J. Acoust. Soc. Am.* **25**, 975–979.
- Culling, J. F., Hawley, M. L., and Litovsky, R. (2004). “The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources,” *J. Acoust. Soc. Am.* **116**, 1057–1065.
- Darwin, C. J., and Hukin, R. W. (1999). “Auditory objects of attention: The role of interaural time differences,” *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 617–629.
- Edmonds, B. A., and Culling, J. F. (2005). “The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay,” *J. Acoust. Soc. Am.* **117**, 3069–3078.
- Grantham, D. W., and Wightman, F. L. (1978). “Detectability of varying interaural temporal differences,” *J. Acoust. Soc. Am.* **63**, 511–523.
- Green, D. M., and Swets, J. A. (1988). *Signal Detection Theory and Psychophysics*, reprint ed. (Peninsula, Los Altos, CA), pp. 1–505.
- Hartmann, W. M., and Johnson, D. (1991). “Stream segregation and peripheral channeling,” *Music Percept.* **9**, 155–184.
- Hartmann, W. M., and Rakerd, B. (1993). “Auditory spectral discrimination and the localization of clicks in the sagittal plane,” *J. Acoust. Soc. Am.* **94**, 2083–2093.
- Hofman, P. M., and Van Opstal, J. A. (1998). “Spectro-temporal factors in two-dimensional human sound localization,” *J. Acoust. Soc. Am.* **103**, 2634–2648.
- Ihlefeld, A., and Shinn-Cunningham, B. (2008). “Spatial release from energetic and informational masking in a selective speech identification task,” *J. Acoust. Soc. Am.* **123**, 4369–4379.
- Kidd, G., Jr., Best, V., and Mason, C. R. (2008). “Listening to every other word: Examining the strength of linkage variables in forming streams of speech,” *J. Acoust. Soc. Am.* **124**, 3793–3802.
- Kidd, G., Jr., Mason, C. R., Best, V., and Marrone, N. (2010). “Stimulus factors influencing spatial release from speech-on-speech masking,” *J. Acoust. Soc. Am.* **128**, 1965–1978.
- King, A. J., and Middlebrooks, J. C. (2011). “Cortical representation of auditory space,” in *The Auditory Cortex*, edited by J. A. Winer and C. E. Schreiner (Springer Science+Business Media, New York), pp. 329–341.
- Kuhn, G. F. (1977). “Model for the interaural time differences in the azimuthal plane,” *J. Acoust. Soc. Am.* **62**, 157–167.
- Licklider, J. C. R., Webster, J. C., and Hedlum, J. M. (1950). “On the frequency limits of binaural beats,” *J. Acoust. Soc. Am.* **22**, 468–473.
- Litovsky, R., Parkinson, A., and Arcaroli, J. (2009). “Spatial hearing and speech intelligibility in bilateral cochlear implant users,” *Ear Hear.* **30**, 419–431.
- Macherey, O., Deeks, J. M., and Carlyon, R. P. (2011). “Extending the limits of place and temporal pitch perception in cochlear implant users,” *J. Assoc. Res. Otol.* **12**, 233–251.
- Macpherson, E. A., and Middlebrooks, J. C. (2000). “Localization of brief sounds: Effects of level and background noise,” *J. Acoust. Soc. Am.* **108**, 1834–1849.
- Macpherson, E. A., and Middlebrooks, J. C. (2002). “Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,” *J. Acoust. Soc. Am.* **111**, 2219–2236.
- Malhotra, S., Hall, A. J., and Lomber, S. G. (2004). “Cortical control of sound localization in the cat: Unilateral cooling deactivation of 19 cerebral areas,” *J. Neurophysiol.* **92**, 1625–1643.
- Marrone, N., Mason, C. R., and Kidd, G., Jr. (2008). “Tuning in the spatial dimension: Evidence from a masked speech identification task,” *J. Acoust. Soc. Am.* **124**, 1146–1158.
- Michéyl, C., and Oxenham, A. J. (2010). “Objective and subjective psychophysical measures of auditory stream integration and segregation,” *J. Assoc. Res. Otol.* **11**, 709–724.
- Middlebrooks, J. C., and Green, D. M. (1990). “Directional dependence of interaural envelope delays,” *J. Acoust. Soc. Am.* **87**, 2149–2162.
- Middlebrooks, J. C., and Green, D. M. (1991). “Sound localization by human listeners,” *Ann. Rev. Psychol.* **42**, 135–159.
- Middlebrooks, J. C., and Snyder, R. (2010). “Selective electrical stimulation of the auditory nerve activates a pathway specialized for high temporal acuity,” *J. Neurosci.* **30**, 1937–1946.
- Mills, A. W. (1960). “Lateralization of high-frequency tones,” *J. Acoust. Soc. Am.* **32**, 132–134.
- Moore, B. C. J., and Gockel, H. (2002). “Factors influencing sequential stream segregation,” *Acta Acust.* **88**, 320–332.
- Phillips, D. P. (2007). “A perceptual architecture for sound localization in man,” *Hear. Res.* **238**, 124–132.
- Sach, A. J., and Bailey, P. J. (2004). “Some characteristics of auditory spatial attention revealed using rhythmic masking release,” *Percept. Psychophys.* **66**, 1379–1387.
- Saupe, K., Koelsch, S., and Rübsamen, R. (2010). “Spatial selective attention in a complex auditory environment such as polyphonic music,” *J. Acoust. Soc. Am.* **127**, 472–480.
- Shaw, E. A. G. (1974). “Transformation of sound pressure level from the free field to the eardrum in the horizontal plane,” *J. Acoust. Soc. Am.* **56**, 1848–1861.
- Shinn-Cunningham, B. G. (2005). “Influences of spatial cues on grouping and understanding sound,” in *Proceedings of the Forum Acusticum*, August 29, Budapest, Hungary, pp. 1539–1544.
- Stainsby, T. H., Fullgrabe, C., Flanagan, H. J., and Waldman, S. K. (2011). “Sequential streaming due to manipulation of interaural time differences,” *J. Acoust. Soc. Am.* **130**, 904–914.

- Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. (2005). "Location coding by opponent neural populations in the auditory cortex," *PLoS Biol* **3**:e78, 520–528.
- Strutt, J. W. (Lord Rayleigh) (1907). "On our perception of sound direction," *Philos. Mag.* **13**, 214–232.
- Turgeon, M., Bregman, A. S., and Ahad, P. A. (2002). "Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises," *J. Acoust. Soc. Am.* **111**, 1819–1831.
- van Hoesel, R. J. M., and Tyler, R. S. (2003). "Speech perception, localization, and lateralization with bilateral cochlear implants," *J. Acoust. Soc. Am.* **113**, 1617–1630.
- Wightman, F. L., and Kistler, D. J. (1992). "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.* **91**, 1648–1661.
- Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," *Nature* **352**, 236–238.
- Zeng, F. G. (2002). "Temporal pitch in electric hearing," *Hear. Res.* **174**, 101–106.
- Zhou, B., Green, D. M., and Middlebrooks, J. C. (1992). "Characterization of external ear impulse responses using Golay codes," *J. Acoust. Soc. Am.* **92**, 1169–1171.