

# Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients

Robert S. Hong<sup>a)</sup> and Christopher W. Turner

Department of Otolaryngology-Head & Neck Surgery, Department of Speech Pathology & Audiology, University of Iowa, 121B WJSHC, Iowa City, Iowa 52242-1012

(Received 20 October 2005; revised 20 April 2006; accepted 21 April 2006)

This study examined the ability of cochlear implant users and normal-hearing subjects to perform auditory stream segregation of pure tones. An adaptive, rhythmic discrimination task was used to assess stream segregation as a function of frequency separation of the tones. The results for normal-hearing subjects were consistent with previously published observations (L.P.A.S van Noorden, Ph.D. dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands 1975), suggesting that auditory stream segregation increases with increasing frequency separation. For cochlear implant users, there appeared to be a range of pure-tone streaming abilities, with some subjects demonstrating streaming comparable to that of normal-hearing individuals, and others possessing much poorer streaming abilities. The variability in pure-tone streaming of cochlear implant users was correlated with speech perception in both steady-state noise and multi-talker babble. Moderate, statistically significant correlations between streaming and both measures of speech perception in noise were observed, with better stream segregation associated with better understanding of speech in noise. These results suggest that auditory stream segregation is a contributing factor in the ability to understand speech in background noise. The inability of some cochlear implant users to perform stream segregation may therefore contribute to their difficulties in noise backgrounds. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2204450]

PACS number(s): 43.66.Ts, 43.66.Mk, 43.66.Lj, 43.71.Ky [JHG]

Pages: 360–374

## I. INTRODUCTION

Auditory stream segregation is the process used to separate a complex sound into different perceptual streams, often corresponding to the different individual sources from which the sound is derived (Bregman, 1990). For example, it is used when people selectively listen to the melodies (streams) played by different instruments (sources) in the presence of an orchestral accompaniment. It also allows people to listen in on different conversations (streams) with different people (sources), one at a time, at a cocktail party. Compared to normal-hearing individuals, cochlear implant patients have tremendous difficulty accomplishing the tasks described in both of the preceding examples; cochlear implant recipients perform significantly worse than normal-hearing individuals on tests of complex song recognition (Gfeller *et al.*, 2005) and speech recognition in steady and modulated noise (Fu *et al.*, 1998; Nelson *et al.*, 2003; Stickney *et al.*, 2004). Thus, a better understanding of auditory stream segregation in implant users may be relevant for improving their ability to hear in complex acoustic environments.

The study of auditory streaming is commonly performed in the laboratory with sequences composed of two different tones (tones A and tone B) that alternate rapidly in time. When these tones are close together in frequency, they are often heard in a single perceptual stream (fusion) fluctuating between tones A and B. As the frequencies become farther apart, this percept changes to that of two different streams (fission), one composed of repeating tone A's and the other of

repeating tone B's. Such a dependence of auditory stream segregation on frequency separation has been observed in both normal-hearing (Miller and Heise, 1950; Van Noorden, 1975) and hearing-impaired individuals (Rose and Moore, 1997; Mackersie *et al.*, 2001).

Van Noorden (1975) varied the frequency separation between tones A and B and found that the boundary between fusion and fission varied depending on the instructions given to the subjects. If the subjects were asked to listen for a single stream, and the frequency difference between the tones was gradually increased until this was no longer possible, one boundary was measured which was termed the temporal coherence boundary. If the subjects instead were asked to listen for two streams, with the frequency difference between tones decreased until this was no longer possible, a different boundary was defined called the fission boundary. In analyzing the data, van Noorden also found a region of frequency separation of the tones between the two boundaries where either fusion or fission could be perceived (depending on the instructions given); this has been referred to as the ambiguity region.

Studies of auditory streaming comparing the ability of normal-hearing and hearing-impaired subjects to stream pure tones have demonstrated that, in general, the hearing-impaired subjects have a reduced ability to perform stream segregation. Rose and Moore (1997) and Mackersie *et al.* (2001) examined the frequency separation for pure tones at the fission boundary and determined that hearing-impaired ears require a greater frequency separation than normal-hearing ears at this boundary, though there were a few exceptions. A number of studies have been performed in an

<sup>a)</sup>Electronic mail: robert-hong@uiowa.edu

attempt to explain the reason for this difference in performance between normal-hearing and hearing-impaired listeners. One theory proposed by Beauvois and Meddis (1996) is that auditory streaming can be explained by the degree of overlap of excitation patterns in the cochlea in response to an acoustic stimulus, with less overlap leading to greater stream segregation. This theory implies that the broader auditory filters of hearing-impaired individuals are responsible for their reduced ability to perform stream segregation. However, Rose and Moore (1997) and Mackersie *et al.* (2001) found that frequency selectivity alone is not a good predictor of auditory streaming ability in hearing-impaired subjects. Another theory suggests that it is the clarity of pitch sensations evoked by the pure tones that determines if streaming occurs. It has been suggested that because hearing-impaired individuals have poorer pure-tone frequency discrimination than normal-hearing subjects, they have a weaker pitch sensation associated with pure tones, and thus are less able to stream pure tones (Rose and Moore, 2005). Rose and Moore (2005) found evidence of statistically significant correlations between frequency discrimination and fission boundary in hearing-impaired subjects. However, these correlations were not particularly strong, suggesting that differences in frequency discrimination are also not enough to fully explain much of the variation in streaming ability in the hearing-impaired. The previous studies suggest that frequency selectivity and frequency discrimination ability can contribute to a person's ability to perform pure-tone stream segregation, but are not sufficient for a good prediction of streaming ability. This is not particularly surprising given that central processes, such as attention, are also thought to be important in stream segregation (Carlyon *et al.*, 2001), suggesting that subject-to-subject variation in such central processes is also important to consider in predicting streaming ability.

Cochlear implant recipients have poorer frequency selectivity and frequency discrimination ability via their speech processors than normal-hearing subjects (Dorman *et al.*, 1996; Gfeller *et al.*, 2002). Therefore, we would expect that as a group, their pure-tone stream segregation ability would be "worse" than that of normal-hearing subjects, assuming that the relevant central processing abilities of both groups are similar. (By "worse" streaming ability, we mean that if, for a given frequency difference between two rapidly alternating pure tones, subject 1 hears one perceptual stream but subject 2 hears two streams, then subject 1 has "worse" streaming ability because he cannot utilize the frequency difference cue to segregate the tones into different streams.) We may also see some cochlear implant subjects who have similar (or even better) stream segregation abilities than the worst-streaming normal-hearing subjects. One can imagine that a cochlear implant patient with extremely good central processing abilities could overcome a degradation of frequency cues (if the cues are not too degraded) and outperform a normal-hearing subject with poor central processing but intact frequency cues.

In the present study, we examined the ability of normal-hearing and cochlear implant subjects to perform stream segregation of acoustically presented pure tones as a function of frequency separation. The cochlear implant subjects listened

through their cochlear implant speech processor. This allowed a direct comparison with normal-hearing subjects, who also listened in the sound field. Furthermore, the cochlear implant subjects in this study performed all experiments with the MAPs they used on an everyday basis. This permitted us to determine if their performance on the stream segregation task was related to their everyday ability to understand speech in background noise, which was also measured. Additionally, streaming ability was assessed at multiple base frequencies (200, 800, and 2000 Hz) to determine if streaming abilities varied within a subject, as might be expected if there were uneven patterns of nerve survival across the cochlea in cochlear implant users. This also allowed us to determine if the ability to stream in lower frequency regions, such as those corresponding to the fundamental frequency of talkers, was more strongly related to the ability to understand speech in noise than streaming at higher frequencies. This might be expected because fundamental frequency differences between the target and masker speech are thought to be important cues for talker segregation.

## II. EXPERIMENT 1. PURE-TONE STREAM SEGREGATION

### A. Participants

Seven normal-hearing subjects (ages 21–35) and eight cochlear implant subjects (ages 39–78) participated in this study. All normal-hearing subjects had pure-tone thresholds of less than 20 dB HL across octave audiometric frequencies (0.25–8.0 kHz), with the exception of one subject who had a threshold of 50 dB HL at 8000 Hz. All cochlear implant subjects were tested using their everyday signal processing strategies and had at least 1 year of experience with their device at the time of testing. The age, implant type, signal processing strategies, and stimulation modes of the cochlear implant subjects who participated in experiment 1 are shown in Table I. This study received prior approval from the Institutional Review Board at the University of Iowa.

### B. Stimuli and procedures

In this study, we use a rhythmic discrimination task based on the one introduced by Roberts *et al.* (2002) to assess auditory stream segregation. We will refer to this task as the streaming rhythm task. In this method, subjects are presented two sequences of rapidly alternating tones and asked to identify which sequence has an irregular rhythm. The task is based on the premise that the irregular rhythm is most easily identified when tone A and tone B are heard together in the same stream, as opposed to individually in different streams. Thus, this task measures stream segregation at the temporal coherence boundary, which is the boundary assessed when subjects are asked to hear all tones in the same stream.

The sequences used in this experiment are identical in rhythm to those described by Roberts *et al.* (2002). All stimuli were pure tones of 60-ms duration, which included 10-ms linear onset and offset ramps. Subjects were presented with two sequences of tones alternating in an AB fashion, with each sequence containing 12 AB cycles. In one se-

TABLE I. Cochlear implant subject demographics. The age, device, signal processing strategy, and stimulation mode (MP=monopolar; BP=bipolar) of each of the 16 cochlear implant subjects who participated in this study are shown. Also indicated are the specific experiments (Exp. 1, 2, and/or 3) in which each participated.

Subject	Age	Device	Strategy	Mode	Experiment no.		
					1	2	3
CI1	49	Nucleus CI24M	ACE	MP	x		x
CI2	51	Nucleus CI24R	CIS	MP	x	x	x
CI3	78	Nucleus CI24M	ACE	MP	x		x
CI4	39	Nucleus CI24R	ACE	MP	x		x
CI5	75	Nucleus CI24R	ACE	MP	x		x
CI6	76	Nucleus CI24M	ACE	MP	x		x
CI7	65	Clarion CII HF	HiRes	MP	x		x
CI8	48	Clarion CI (spiral)	CIS	MP	x		x
CI9	78	Clarion CII HF	HiRes	MP		x	x
CI10	33	Clarion 90K	HiRes	MP		x	x
CI11	64	Clarion CII HF	HiRes	MP			x
CI12	58	Nucleus CI22	SPEAK	BP <sup>a</sup>			x
CI13	67	Clarion CII HF	HiRes	MP			x
CI14	46	Clarion CII HF	HiRes	MP			x
CI15	54	Clarion CI (spiral)	CIS	MP			x
CI16	44	Clarion CII HF	HiRes	MP			x

<sup>a</sup>Electrode 3 served as the ground for all electrode pairs in bipolar configuration.

quence, tone A and tone B alternated back and forth regularly, separated by a 40-ms silence between each tone. In the other sequence, tone A and tone B alternated such that the first six AB cycles were regularly spaced (identical to the first sequence), the next four AB cycles (transition region) had an increasing delay (with the magnitude of the increase in delay being equal from one cycle to the next) imposed on

the onset of tone B, and the final two AB cycles maintained the accumulated delay. The length of each sequence was 2.4 s. In Fig. 1, with “\*” = tone A and “#” = tone B, the components of the regular and irregular rhythm sequences are shown.

In the streaming rhythm task, it was to the listener’s advantage to hear all the tones in the same perceptual stream, because the delay in the irregular rhythm sequence was more difficult to detect when they were heard in different streams. Thus, as the two tones in the sequence became further apart in frequency, if a listener experienced stronger segregation of the two tones into different streams, then the rhythmic discrimination task became more difficult.

Subjects were presented a two-interval forced choice task (2IFC) task with feedback in which they were asked to identify which of two sequences of alternating tones results in a rhythm that sounds “unsteady and irregular.” The irregular sequence was randomly presented in either the first or second interval from trial to trial. Each sequence was numbered according to the order presented (“1” or “2”) and displayed as labeled buttons on a touch-screen (MicroTouch). The task measured the smallest delay in tone B that resulted in a detectable irregularity of rhythm (Fig. 1), utilizing a three-down one-up adaptive staircase to converge on the 79.4% correct point on the psychometric function for time delay (Levitt, 1971). The maximum delay that was imposed on tone B to avoid overlap of the tones was 40 ms. The initial size of the accumulated delay was 32 ms. The adaptive step size for the first two reversals was a step size of 8 ms. The last four reversals used a step size of 4 ms. The time delay threshold for each run was taken as the mean of the last four reversals. Thresholds for subjects who made four total incorrect identifications at the maximum time de-

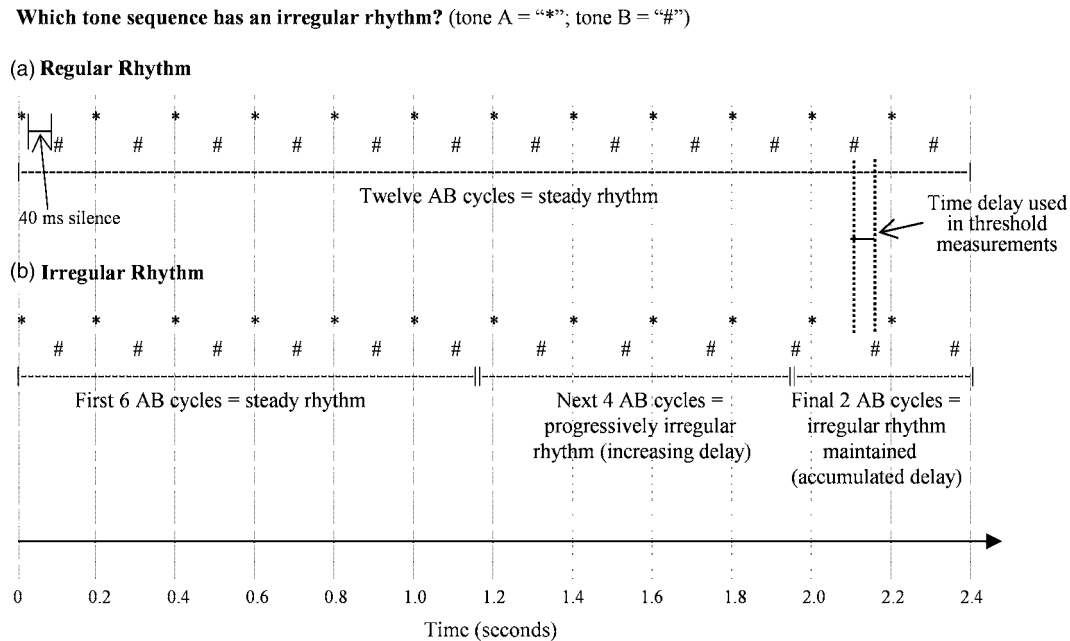


FIG. 1. Test stimuli used in rhythmic discrimination task to assess auditory stream segregation. (a) The regular rhythm is composed of 24 evenly spaced, alternating pure tones in the ABAB format. Tones A and B are always separated by 40 ms of silence. (b) The irregular rhythm is composed of 24 alternating tones that are evenly spaced over the first six cycles, have a progressively increasing delay in tone B over the next four cycles, and maintain the accumulated delay over the final two cycles. The time delay used to calculate threshold measurements for detection of the irregular rhythm is the delay in tone B in the final, accumulated delay section of that sequence.

lay were nominally taken as the 40-ms maximum delay, and the adaptive procedure was terminated in those cases.

The stimuli were grouped into three conditions, based upon the frequency of tone A: (1) low base frequency, with tone A=200 Hz; (2) medium base frequency, with tone A=800 Hz; and (3) high base frequency, with tone A=2000 Hz. For each base frequency (tone A), tone B was chosen to give a Weber fraction  $[(\text{frequency B} - \text{frequency A}) / \text{frequency A}]$  of 0, 0.01, 0.1, 0.5, 1.0, and 3.0. The exception was that for the 2000-Hz base frequency, a Weber fraction of 3.0 was approximated with a tone B of 7800 Hz instead of the calculated 8000 Hz, because 8000 Hz is above the frequency range presented by signal processing strategies to cochlear implants. Normal-hearing subjects were tested at all Weber fractions for each base frequency. Cochlear implant subjects were tested at the Weber fractions that appeared to be most relevant for defining the shape of the psychometric function at each base frequency.

Testing was completed for one base frequency condition before proceeding to the next. The order of conditions was randomized across subjects. At each base frequency condition, the first Weber fraction tested was 0 (frequency of tone A=tone B). The order of presentation of all other Weber fractions was randomized within each condition. Prior to the collection of data, each subject was given three practice runs at the first base frequency condition to be tested, with one run corresponding to a Weber fraction of 0, and the other two runs corresponding to a Weber fraction of 0.01. If a subject was not able to obtain an average time delay threshold below 15 ms with the two tones equal in frequency (Weber fraction=0), the subject was excluded from further testing, since this suggested a lack of ability to perform rhythmic discrimination exclusive of stream segregation. One normal-hearing subject was unable to meet this criterion at any of the base frequencies. Additionally, one cochlear implant subject (subject CI8) was able to meet this criterion for two base frequencies (200 and 2000 Hz) but not the third (800 Hz), and was excluded from testing at 800-Hz base frequency only. At each Weber fraction for each base frequency condition, subjects were tested with three consecutive adaptive runs. A fourth run was added if the standard deviation of the average of the three trials was greater than 6 ms, or if the standard deviation of the reversals of any run was greater than 8 ms. For each subject, the value for time delay threshold at a particular Weber fraction and base frequency was taken as the average of all three (or four) runs performed at that test condition.

All stimuli were generated digitally using MATLAB and stored on a Macintosh G4 computer. Stimuli were output through a 16-bit digital-to-analog converter (Audiomedia III, Digidesign, Inc.) at a sampling rate of 44.1 kHz and smoothed by a 20-kHz antialiasing low-pass filter. The stimuli were presented via a loudspeaker situated directly in front of the listener in a double-walled sound-attenuated booth. For normal-hearing subjects, all stimuli were presented at 80 dB SPL, with their right ear plugged with an ear plug. For cochlear implant subjects, the low base frequency set was presented at 95 dB SPL, and the medium and high base frequency sets were presented at 90 dB SPL. The ear

without the cochlear implant was plugged. Cochlear implant subjects were allowed to adjust their microphone settings so that the sound was at a comfortable level.

## C. Results and discussion

### 1. Streaming abilities of normal-hearing and cochlear implant subjects

The normalized results of both normal-hearing and cochlear implant subjects on this task are shown in Fig. 2 for three different base frequencies (tone A=200, 800, or 2000 Hz). Our approach to eliminating any possible effects of a subject's basic ability to discriminate rhythms upon the question of interest (ability to separate frequencies into different streams) was to normalize the raw scores by dividing each value by the threshold at Weber fraction=0 (i.e., when all of the tones are identical). This particular method of normalization was selected, because performance on temporal tasks in normal-hearing subjects (such as gap discrimination or detection) has been shown to be relatively linear with respect to frequency separation (of the tones bordering the gap), when both factors are plotted on a logarithmic scale (e.g., Neff *et al.*, 1982; Formby and Forrest, 1991). [We also tested an alternative method of normalization wherein the values of threshold at Weber fraction=0 were subtracted from the raw scores. This method yielded similar conclusions (data not shown).] Thus, the normalized thresholds may be interpreted as a relative measure of auditory stream segregation, measuring how many times more difficult the streaming rhythm task became compared to baseline (when all the tones were identical) as the frequency difference between alternating tones was increased.

The average normalized results of six normal-hearing (NH) subjects on the task are indicated by the heavy line in Fig. 2. As shown in the figure, the smallest average threshold for normal-hearing subjects was achieved when there was no frequency difference between the tones, which is as expected since all of the tones were identical in frequency and thus heard in the same stream. As the frequency separation increased between tone A and tone B, thresholds increased, suggesting that subjects more strongly heard two streams. A single value for the overall streaming ability of normal-hearing subjects was derived from the slope of a regression line with a y intercept at 1 (since at a frequency difference of 0, normalization by the threshold at Weber fraction=0 results in a value of 1) that was fit through the data shown in Fig. 2. Data where ceiling effects were reached at larger frequency differences (i.e., when subjects were unable to identify the irregular rhythm at the largest time delay possible) were not included in the regression analysis, since inclusion of such data may result in underestimation of the slope. Higher values for the slope correspond to larger increases in difficulty of the task as the frequency difference between alternating tones is widened; thus, larger slopes correspond to more auditory stream segregation with increasing frequency separation. The value of the slope for the average normal-hearing (NH) subject at each of the three base frequencies and the 95% confidence interval for each slope distribution ( $\pm 2$  standard deviations about the mean) are shown in Fig. 3. As can

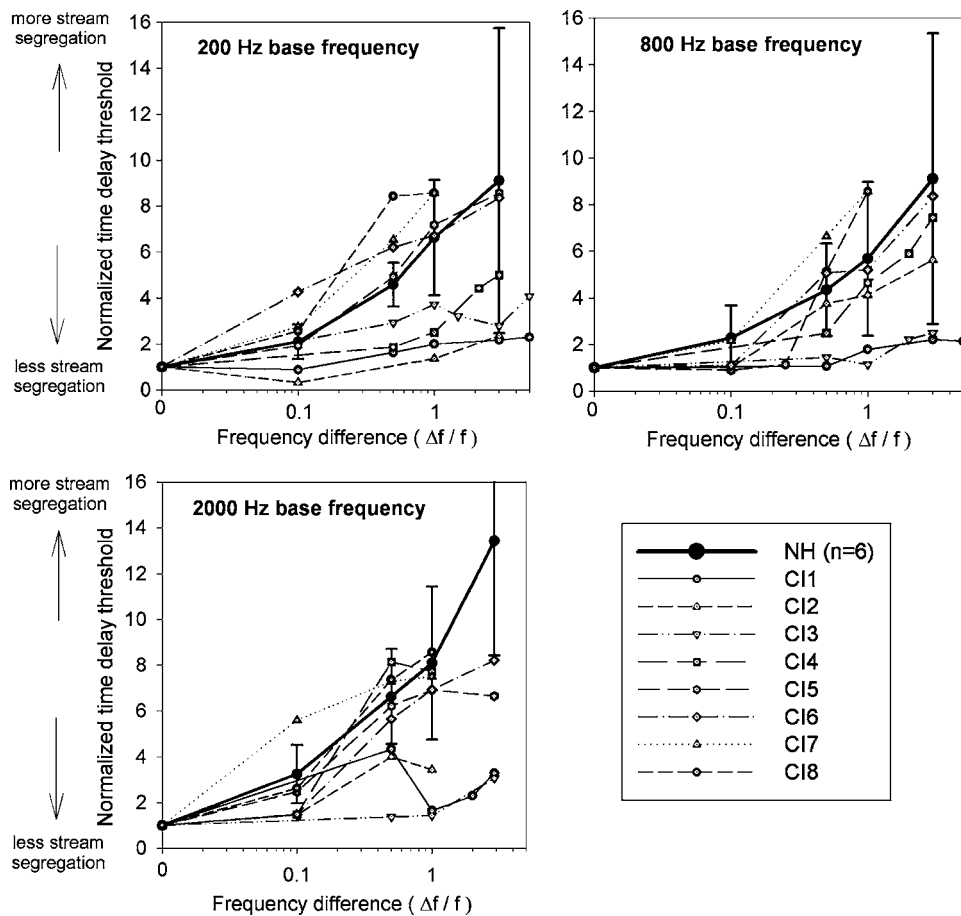


FIG. 2. Normalized performance on the streaming rhythm task for normal-hearing and cochlear implant subjects. The average performance of six normal-hearing subjects (NH) is shown in the heavy line at various frequency differences between tone A and tone B, with the error bars representing one standard deviation above and below the mean. The individual performance of seven to eight cochlear implant (CI) subjects is also shown. Measures are made at each of three base frequencies (tone A=200, 800, or 2000 Hz). The time delay thresholds are normalized by baseline performance, giving a normalized threshold that represents the factor by which the task increases in difficulty at different frequency separations of the alternating tones. A larger value of normalized threshold is consistent with greater auditory stream segregation.

be seen from this figure, there is a relatively large range of pure-tone streaming abilities among normal-hearing subjects as measured by this task.

Figure 2 also shows the individual data from eight cochlear implant (CI) subjects on the streaming rhythm task. With cochlear implant recipients, the baseline performance (Weber fraction=0) was generally comparable to that of normal-hearing subjects on the task, but there was variation in performance among cochlear-implant subjects with increasing frequency separation. Statistical analysis was again performed by fitting a regression line in a similar manner as discussed for normal-hearing subjects through the data for each cochlear implant subject. To determine the performance of cochlear implant relative to normal-hearing subjects, we compared the regression slopes of each cochlear implant subject with that of the average normal-hearing subject at each of the three base frequencies (Fig. 3). Values for individual cochlear implant subjects that fall outside the dotted lines (representing the 95% confidence interval for the normal-hearing slope distribution) are interpreted as significantly different from those of the average normal-hearing subject. At a base frequency of 200 Hz, four cochlear implant subjects have slopes that are significantly lower than the average normal-hearing subject, while at 800 and 2000 Hz, two and three cochlear implant subjects, respectively, have slopes that are significantly lower. These results suggest that a number of cochlear implant subjects experience significantly less auditory stream segregation with increasing frequency separation than normal-hearing listeners. The effect of base fre-

quency on streaming was also analyzed via repeated-measures ANOVA for the regression slopes, with no significant differences observed within cochlear implant subjects between base frequencies [ $F_{1,17,6}=4.08$  (Greenhouse-Geisser adjustment for lack of sphericity);  $p > 0.05$ ].

## 2. Influence of electrode separation (place pitch cues) on streaming

The influence of electrode separation on the streaming rhythm task was also examined for the six Nucleus CI24 cochlear implant recipients. The goal of this analysis was to determine if performance on the task could be explained primarily by the distance between the electrodes presenting tones A and B; if this were the case, for example, we would expect two tones presented by adjacent electrodes to result in similar streaming performance whether the base frequency was 200, 800, or 2000 Hz. We focused on the Nucleus subjects because we wanted our measure of electrode separation to reflect the role of place pitch cues for auditory streaming. The low-pass temporal envelope cutoff frequency for the signal processing strategies of the Nucleus subjects was  $\sim 125$  Hz (personal communication with Bom-Jun Kwon, Cochlear Corporation), which was below the lowest frequency tested (200 Hz), suggesting that temporal pitch cues would not confound our analysis. In contrast, for the Clarion subjects, the low-pass envelope cutoff frequency was high enough that temporal pitch cues may have been available to these subjects.

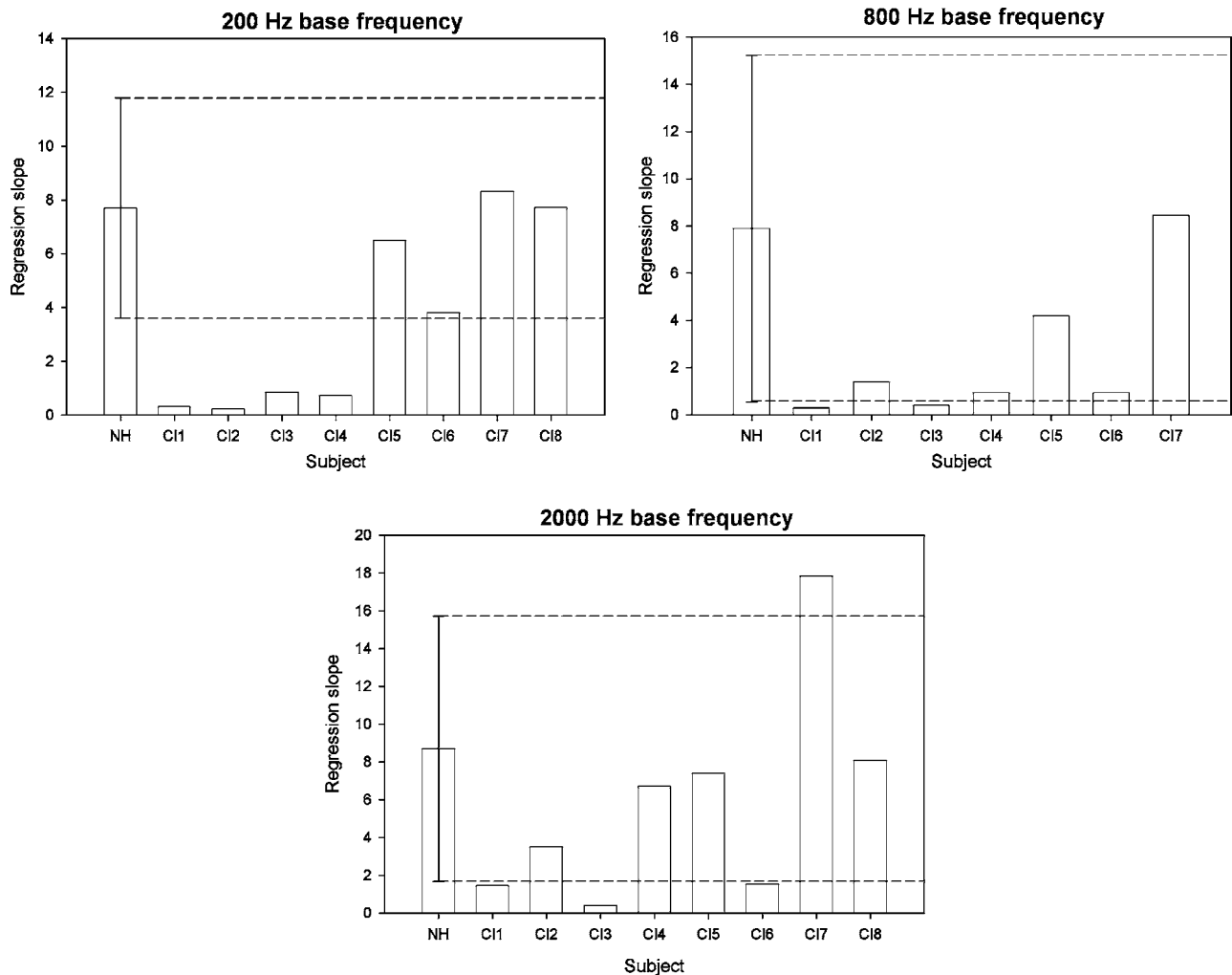


FIG. 3. Overall streaming ability in normal-hearing and cochlear implant subjects. Overall streaming ability is indicated by the slopes of regression lines (with a common y intercept) fit through the data for the streaming rhythm task at each of three base frequencies (tone A=200, 800, or 2000 Hz). NH represents the average slope of six normal-hearing subjects, with the 95% confidence interval for the normal-hearing slope distribution ( $\pm 2$  standard deviations about the mean) denoted by the dashed lines. The slopes of each individual cochlear implant (CI) subject are shown by the other bars.

To analyze the role of electrode separation on streaming, the values for frequency difference shown in Fig. 2 were replotted as electrode separation on separate graphs for each individual subject (with two such examples shown in Fig. 4). To convert frequency difference to electrode separation, each stimulus tone (tone A or tone B) was assigned to the single electrode that was programmed to represent that stimulus frequency in the patient's clinical MAP (according to the cutoff frequencies for each electrode). Then, the number of electrodes separating the stimulating electrodes corresponding to tone A and tone B was determined. This conversion only provided an estimate of the electrode separation for two tones, since the slope of each filter at the cutoff frequency was not infinite. Nevertheless, this was useful in providing a general picture of the influence of electrode separation on streaming. The conversion was performed for each subject at all three base frequencies, with the resulting values plotted versus normalized threshold onto a single graph for each subject. Linear regression was then performed for the data across all three base frequencies for each subject. A statistically significant value for the slope of the regression line was interpreted as evidence that performance on the streaming

rhythm task could be explained by electrode separation, regardless of the region of the cochlea that the stimuli were presented.

Four cochlear implant subjects (subjects CI4, CI1, CI5, and CI2) showed evidence of a similar streaming ability based on electrode separation across all frequency ranges. Figure 4(a) shows the data for one of these four subjects, subject CI4. As depicted by the figure, the linear regression line for the normalized time delay thresholds versus electrode separation across the three base frequency conditions was statistically significant ( $p < 0.001$ ), with slope=0.278 and  $r=0.655$ . The linear regressions for the other three subjects (not shown in figure) were also statistically significant—subject CI1 (slope=0.117;  $r=0.396$ ;  $p < 0.005$ ), subject CI5 (slope=1.388;  $r=0.841$ ;  $p < 0.001$ ), and subject CI2 (slope=0.339;  $r=0.792$ ;  $p < 0.001$ )—suggesting that electrode separation could also explain a significant portion of their performance on the streaming rhythm task in different stimulus frequency ranges.

In contrast, the other two cochlear implant subjects (subjects CI3 and CI6) appeared to have different streaming

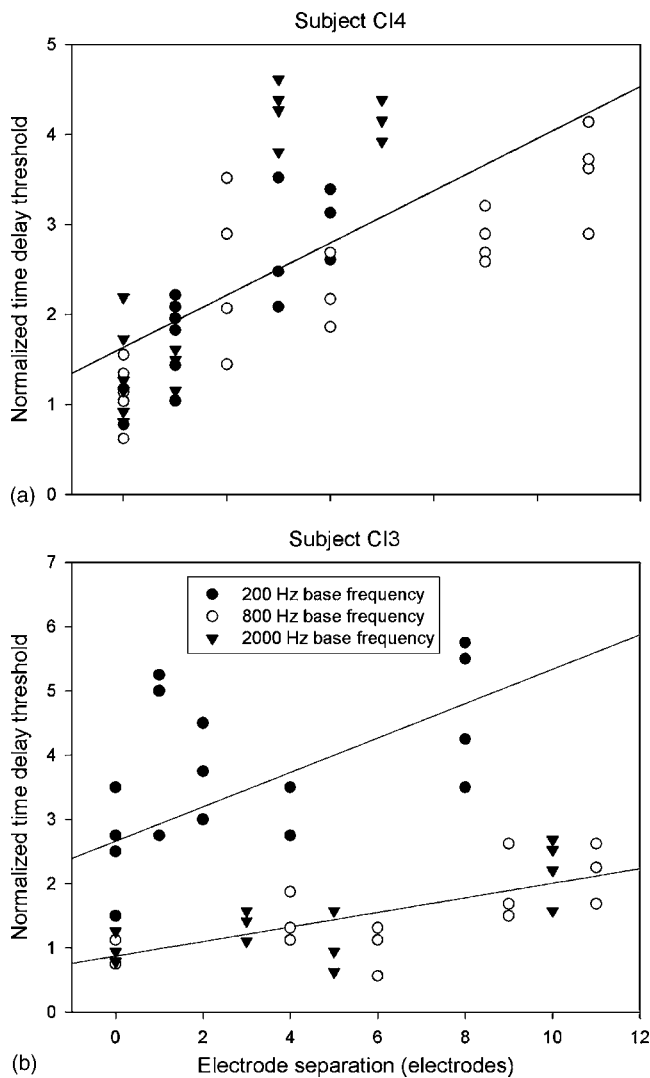


FIG. 4. Performance on streaming rhythm task with respect to electrode separation in cochlear implant subjects. Data from all three base frequencies of the streaming rhythm task are converted into electrode separation and plotted on the same graph for each subject. The resulting graphs of two subjects, which represent the two general types of data patterns observed, are shown. (a) Subject CI4 has a moderate, statistically significant correlation between task performance across all frequency ranges and electrode separation ( $r=0.655$ ;  $p<0.001$ ). (b) Subject CI3 shows no evidence of a correlation between task performance and electrode separation for data combined across all three base frequencies ( $r=0.140$ ;  $p>0.1$ ; regression line not shown). However, correlations were significant when one regression line was drawn through the 200-Hz data ( $r=0.570$ ;  $p<0.02$ ), and a different regression line was drawn through the combined 800- and 2000-Hz data ( $r=0.714$ ;  $p<0.001$ ).

abilities across electrodes in different regions of the cochlea. Figure 4(b) shows an example of one subject (subject CI3) who appeared to have different streaming abilities in different frequency ranges, after accounting for electrode separation, since a single regression line fit the data very poorly ( $r=0.140$ ;  $p>0.1$ ). In fact, if two regression lines are drawn through the data, one through the 200-Hz base frequency data points and the other through the 800- and 2000-Hz base frequency points, the correlations are statistically significant (for 200 Hz line, slope=0.268,  $r=0.570$  and  $p<0.02$ ; for the 800–2000-Hz line, slope=0.113,  $r=0.714$  and  $p<0.001$ ). The data for subject CI6 also fit a single regression line

poorly ( $r=0.053$ ;  $p>0.5$ ). Two regression lines were again required to better describe the data, with one line for the 800- and 2000-Hz base frequencies (slope=0.176;  $r=0.723$ ;  $p<0.001$ ) and another for the 200-Hz base frequency ( $r=0.264$ ;  $p>0.1$ ), though the latter correlation was not statistically significant.

### 3. General discussion

The results from Fig. 3 suggest that there is a range of streaming abilities for both normal-hearing and cochlear implant users. Nevertheless, despite the variability in performance on the streaming rhythm task within the normal-hearing group, this study demonstrates that some cochlear implant users stream significantly worse than normal-hearing subjects. These differences in pure-tone streaming between cochlear implant subjects likely represent a combination of differences in peripheral (frequency resolution) and central processing. The results from Fig. 4 suggest that in some cases, within a cochlear implant subject, streaming ability based on place pitch cues can vary from one region of the cochlea to another. Within a subject, the contribution of central processing to streaming is presumably similar across all stimulus frequencies, and thus differences in streaming for different electrode locations reflect differences in peripheral processing. One possible explanation for why some subjects may have different streaming abilities for electrode separation in different regions of the cochlea is that there may be different amounts of nerve survival throughout the cochlea, leading to differences in the perceptual distance of pitch attributed to adjacent electrodes and thus differences in pure-tone streaming.

There is a possibility that differences in loudness, as opposed to differences in pitch, could be responsible for some of the auditory stream segregation observed in this experiment. All of the stimuli were presented at equal dB SPL and thus were not loudness balanced. We believe that the effects of loudness differences in streaming in our experiment were minimal for a number of reasons. First, for normal-hearing listeners, there are minimal differences in loudness across much of the intensity and frequency range tested [ISO 226: 2003 (Normal equal-loudness-level contours) of the International Organization for Standardization]. Second, the presence of any loudness cues for all subjects was minimized by the presentation of stimuli in a free field, because the head position of listeners was not fixed with respect to the speaker, which allowed for potential random variations in sound intensity over a few decibels with sporadic changes in head position during the course of testing. Third, none of the subjects reported that loudness differences hindered performance on the task. Fourth, there is no clear evidence that loudness differences can be used to obligatorily segregate sounds in the same powerful way as frequency differences (Bregman, 1990, pp. 126–127). Finally, it is interesting to note that even if loudness cues were present, there were clearly some cochlear implant subjects that had little ability to stream at any frequency difference based on either loudness or frequency cues (e.g., subject CI1 or CI3 in Fig. 3). For the other cochlear implant subjects with better streaming abilities, if they were able to use loudness as a cue

for streaming, this would suggest that our results may overestimate their ability to segregate pure tones into different streams.

There is also the possibility that age differences between the normal-hearing and cochlear implant groups in this experiment may be a confounding factor in the analysis of the results with respect to streaming. The cochlear implant subjects were older than the normal-hearing subjects (ages 39–78 compared to ages 21–35), and there is evidence in the literature that performance on temporal tasks declines with age (for a review, see Pichora-Fuller, 2003). However, there are a number of reasons why we believe that differences in age between the two groups did not affect our overall conclusions. First, as stated earlier, the time delay thresholds from the streaming rhythm task used in our analysis of streaming ability were all normalized with respect to each individual subject's baseline ability to discriminate rhythms (baseline assessed at Weber fraction=0). By doing this, we attempted to control for differences in basic temporal perception ability between individuals, including those due to differences in age. Second, we did not find any statistically significant correlations between age and performance on the streaming rhythm task (for normalized thresholds and regression slopes), both within and across subject groups. Finally, assuming that temporal ability differences due to age were responsible for differences in performance on the task, we would expect that the older cochlear-implant group would have higher thresholds than the younger normal-hearing group. However, our results indicated the opposite: the time delay thresholds of the cochlear-implant group tended to be lower than those of the normal-hearing group. Thus, it does not appear that age differences between the two groups confounded our conclusion that the streaming ability of many cochlear implant subjects was worse than that of normal-hearing individuals.

### III. EXPERIMENT 2. AUDITORY STREAM SEGREGATION OR GAP DISCRIMINATION?

#### A. Rationale

It is possible that the results obtained on the streaming rhythm task may not reflect auditory streaming, but merely gap discrimination. For example, one can imagine a subject who performs the streaming rhythm task by focusing only on the end of the alternating-tone sequence to determine the sequence with the irregular rhythm. Taken to the extreme, it is possible that the subject may ignore the entire sequence except for the final three tones of the sequence, where the largest and smallest gaps are present, turning the task into one that looks very similar to gap discrimination.

It is difficult to determine based on the results of experiment 1 whether gap discrimination or stream segregation is the dominant phenomenon. Studies of gap detection and stream segregation in normal-hearing subjects have demonstrated that both are affected similarly by differences in frequency between tones: it becomes more difficult to perceive a gap and more difficult to hear tones in the same stream as the frequency separation widens between different tones (Neff *et al.*, 1982; Phillips *et al.*, 1997). If we presume that

the increasing frequency separation affects performance by increasing the perceptual pitch distance between tones, then a similar relationship might also be seen with gap detection in cochlear implants. It has been found that gap detection in cochlear implants worsens with increasing electrode distance (place pitch effects) and increasing rate differences (temporal pitch effects) between the two tones which border the gap to be detected (Hanekom and Shannon, 1997; Chatterjee *et al.*, 1998; van Wieringen and Wouters, 1999). Furthermore, any correlations found in this study (described in experiment 3) between the streaming rhythm task and performance on a speech perception in noise task also will not lend insight into whether the streaming task measures stream segregation or gap discrimination. While correlations between speech perception in noise have been found with auditory streaming (Mackersie *et al.*, 2001), correlations have also been found between gap detection and speech perception in noise (Tyler *et al.*, 1982; Dreschler and Plomp, 1985), although such correlations with gap detection are not universally present (Strouse *et al.*, 1998; Snell and Frisina, 2000).

Although there are a number of similarities between auditory stream segregation and gap discrimination, there are also a number of ways to distinguish the two. First, auditory stream segregation is affected by the presentation rate of the alternating tones, whereas gap discrimination is not (Neff *et al.*, 1982). If performance on a task designed to assess auditory streaming varies with presentation rate, this lends supportive evidence that the task measures streaming ability. Second, auditory streaming is known to build up over time for alternating tones of moderate frequency separation, such that increasing amounts of segregation are seen over the first 10–30 s of listening to such tones (Anstis and Saida, 1985). This observation can also be used to obtain supportive evidence that differences in performance on the streaming rhythm task employed in this study reflect differences in streaming ability, and this latter approach is the one that we chose to take in experiment 2.

The goal of experiment 2 is to determine if the streaming rhythm task used in experiment 1 measures auditory stream segregation. For this experiment, we use shortened versions of the alternating-tone sequences presented in the streaming rhythm task and again measure the time delay threshold required to hear the irregular rhythm. We refer to this task as the short rhythm task. If there is no difference in relative performance between the short rhythm task and the streaming rhythm task when the frequency separation in tones is increased, this suggests that the streaming rhythm task is merely a measure of gap discrimination. However, if, as we hypothesize, subjects perform relatively worse on the streaming rhythm task, then this suggests that the streaming rhythm task measures streaming ability: the build-up of streaming induced by the longer sequences in the streaming rhythm task makes it more difficult to detect the irregular rhythm.

#### B. Participants

Three normal-hearing subjects who participated in experiment 1 also participated in this experiment. Additionally,

three cochlear implant (subjects CI2, CI9, and CI10) subjects were randomly selected to participate in this experiment, with their demographics found in Table I.

### C. Stimuli and procedures

The task in this experiment is identical in all respects to the streaming rhythm task presented in experiment 1, except for the stimuli. In this experiment, the rhythmic sequences are shortened so that they contain only three pure tones (instead of 24 pure tones). The three pure tones of the regular rhythm sequence in this experiment are derived from the first three tones of the regular rhythm sequence of the streaming rhythm task. The three pure tones of the irregular rhythm sequence in this experiment are identical to the first three tones of the “accumulated delay” region (described in Fig. 1) of the irregular rhythm sequence of the streaming rhythm task. The time delay used for rhythmic discrimination threshold measurements is thus the same for the two experiments. As with the streaming rhythm task, subjects listened to two sequences and are asked to identify the irregular, unsteady rhythm in a 2-IFC adaptive task that converges on the 79.4% correct point for time delay threshold. Subjects were tested at each of three base frequencies (200, 800, and 2000 Hz) at conditions corresponding to a Weber fraction=0 and Weber fraction=0.5. Overall thresholds were determined from the average of the thresholds from three (or four) runs at each test condition. These specific conditions were chosen because they were the ones where correlations were subsequently performed between the streaming rhythm task and speech perception in noise (described in experiment 3), allowing insight into whether any correlations observed at those specific frequency differences reflected stream segregation or gap discrimination.

### D. Results and discussion

The baseline ability (Weber fraction=0) of each listener to perform the streaming rhythm task and the short rhythm task at each of three base frequencies (200, 800, and 2000 Hz) is shown in Fig. 5 for both normal-hearing [Fig. 5(a)] and cochlear implant [Fig. 5(b)] subjects. In every case, when all of the tones were identical in frequency, it was easier to detect the irregular rhythm in the streaming rhythm task (24-tone sequences) than in the short rhythm task (three-tone sequences), suggesting that the streaming rhythm task is a fundamentally easier task (irrespective of streaming considerations). There are a number of possible reasons to explain this result. First, the 24-tone sequences contain three repetitions of the irregular rhythm used in the short rhythm task, which may provide more chances for the subject to hear the irregular rhythm in the streaming rhythm task, making the task easier. Second, the 24-tone sequences contain a region of progressively increasing delay in the rhythm (Fig. 1) not present in the three-tone sequences. This may provide an additional cue to subjects taking the streaming rhythm task for detecting the irregular rhythm.

Figure 6 shows the performance of normal-hearing [Fig. 6(a)] and cochlear implant [Fig. 6(b)] subjects on the streaming rhythm task and the short rhythm task at a Weber fraction

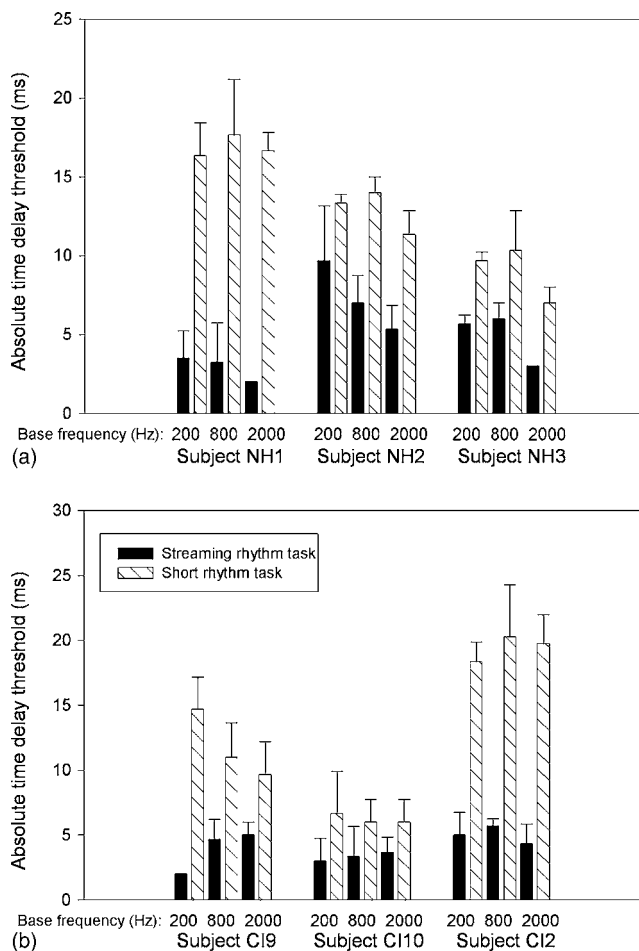


FIG. 5. Baseline performance on streaming rhythm task versus short rhythm task for normal-hearing and cochlear implant subjects. The baseline performance (tone A = tone B) on the streaming rhythm task (filled bars) and short rhythm task (hatched bars) is shown at all three base frequencies (200, 800, and 2000 Hz) for (a) three normal-hearing and (b) three cochlear implant subjects. Error bars represent one standard deviation.

of 0.5. The raw scores for time delay threshold have been divided by baseline performance thresholds (i.e., threshold at Weber fraction=0) for each test to give the resulting normalized thresholds shown on the ordinate in Fig. 6. This normalized threshold represents the performance on each task when the frequency difference of the alternating tones is at a Weber fraction of 0.5, after taking into account the baseline performance on each task shown in Fig. 5. The normalization procedure allows us to directly compare the relative performance on the streaming and short rhythm tasks at a Weber fraction of 0.5 to determine if auditory stream segregation plays a role in performance on the streaming rhythm task. The reason that normalized thresholds as opposed to absolute thresholds are compared between the two tasks is that the normalized threshold is the relevant measure used to assess streaming in the streaming rhythm task; our goal is to determine if a measure derived from the short rhythm task in a similar manner leads to similar results. If normalized performance on the two tasks is the same, this suggests that the difference in sequence length for the two tasks does not affect performance and that the streaming rhythm task does not measure stream segregation. In contrast, if normalized per-

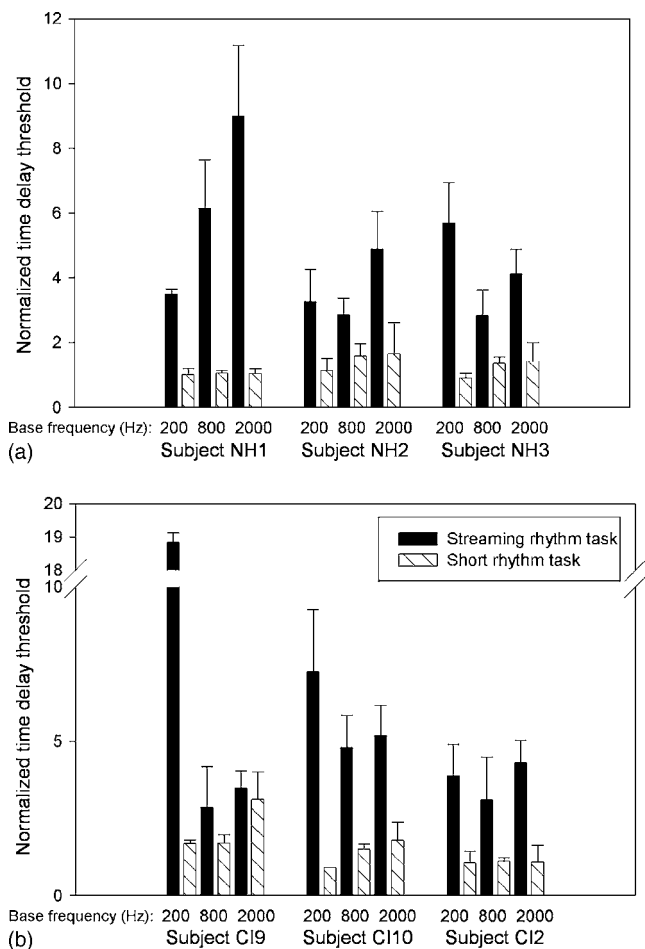


FIG. 6. Performance on streaming rhythm task versus short rhythm task for normal-hearing and cochlear implant subjects. Normalized time delay thresholds for the streaming rhythm (filled bars) and short rhythm (hatched bars) tasks are shown, with the alternating tones separated by a frequency difference corresponding to a Weber fraction of 0.5. Data are shown at all three base frequencies (200, 800, and 2000 Hz) for (a) three normal-hearing and (b) three cochlear implant subjects. Error bars represent one standard deviation.

formance is worse with the streaming rhythm task, this suggests that the longer sequence in the streaming rhythm task has induced a build-up of stream segregation to make it more difficult to detect the irregular rhythm in this task, and that the streaming rhythm task does measure stream segregation.

The values of the normalized threshold for the short rhythm task for both normal-hearing and cochlear implant subjects across all base frequencies generally fell in the range of 1 to 2, suggesting that increasing the frequency separation of the tones from baseline (no difference in frequency) to one corresponding to a Weber fraction of 0.5 resulted in either no change in difficulty of the task to a two times increase in difficulty of the task. For the streaming rhythm task, the normalized threshold values at a Weber fraction of 0.5 ranged from 2.83 to 9 for normal-hearing subjects and from 2.86 to 19 for cochlear implant subjects. This suggests that the streaming rhythm task was at least 2.8 times more difficult for all subjects in this experiment when the frequency separation was increased from baseline to one corresponding to a Weber fraction of 0.5.

To determine the contribution of streaming to this in-

creased difficulty, we compared the normalized thresholds of the streaming rhythm task with those of the short rhythm task. Figure 6 shows that in every case, for both normal-hearing and cochlear implant subjects, the normalized threshold was significantly higher for the streaming rhythm task. For normal-hearing subjects, the mean of the threshold on the streaming rhythm task was 4.69, compared to 1.23 for the short rhythm task ( $t_{df=8}=4.85$ ;  $p=0.001$ ). For cochlear implant subjects, the mean of the threshold on the streaming rhythm task was 5.96, compared to 1.54 for the short rhythm task ( $t_{df=8}=2.62$ ;  $p=0.03$ ). If subjects were performing the streaming rhythm task by ignoring the first part of the 24-tone sequence and only focusing on the end of the sequence, such that streaming was irrelevant to the task, we would expect the normalized thresholds for the two tasks to be similar. This is not the case. Instead, the higher values of normalized threshold for the streaming rhythm task suggest a greater build-up of streaming with the longer sequence, making the streaming rhythm task relatively more difficult than the short rhythm task. Therefore, the results suggest that the normalized threshold in the streaming rhythm task reflects auditory stream segregation and not just gap discrimination abilities.

Our findings are consistent with those in the literature that suggest an increase in separation in pitch between the tones bordering a gap may result in a more difficult gap discrimination task for both normal-hearing (Neff *et al.*, 1982; Phillips *et al.*, 1997) and cochlear implant subjects (Hanekom and Shannon, 1997; Chatterjee *et al.*, 1998; van Wieringen and Wouters, 1999). Furthermore, the differences in performance we observed between the streaming rhythm task and short rhythm task are consistent with the results of Grose and Hall (1996), who showed that the build-up of auditory stream segregation through the use of longer sequences can make a gap discrimination task more difficult; in our experiment, the greater difficulty of the streaming rhythm task compared to the short rhythm task provides evidence that the streaming rhythm task measures auditory stream segregation. Further evidence for the streaming rhythm task as a measure of streaming has been provided by Roberts *et al.* (2002). This study compared the results of the streaming rhythm task with results from a different task that has been used to assess streaming and found similar results for both tasks, suggesting that the present task does measure auditory stream segregation.

In conclusion, experiment 2 provides evidence that the streaming rhythm task is a measure of auditory stream segregation and not merely gap discrimination. Having examined this, we will now move forward to experiment 3 and determine the implications of performance on our measure of stream segregation on the understanding of speech in background noise for cochlear implant users.

#### IV. EXPERIMENT 3. AUDITORY STREAMING AND SPEECH PERCEPTION IN NOISE

##### A. Rationale

A better ability to perform auditory stream segregation should be associated with a better ability to understand

speech in various backgrounds; to understand speech in noise, it is helpful to segregate the target speech into one attended perceptual group and the noise into a different, ignored group. Such a relationship between streaming and speech perception in noise has been found in hearing-impaired individuals (Mackersie *et al.*, 2001), but, to our knowledge, has yet to be studied in cochlear implant patients. We are interested in determining if a relationship exists between stream segregation and speech perception in noise in cochlear implant patients, because such a relationship would suggest that measures of stream segregation may be useful for comprehending the problems that implant users have in understanding speech in noise. Our hypothesis is that cochlear implant patients who have difficulty segregating pure tones into separate streams will have more trouble understanding speech in various backgrounds.

Additionally, we have chosen to study two different types of noise—steady-state noise and two-talker babble—to see if streaming ability is more strongly related to one or the other. For example, studies have shown that in normal-hearing subjects, the fundamental frequency of target speech is an important cue for segregating it from background talkers (Brokx and Nootboom, 1982; Assmann and Summerfield, 1990). This leads us to hypothesize that the strongest correlations may occur for cochlear implant subjects between low-frequency pure-tone streaming ability and speech perception in babble, assuming fundamental frequency is also an important cue for talker segregation in cochlear implant users. Furthermore, by studying both types of noise, we can determine if the relative performance on speech perception in steady-state noise versus two-talker babble is related to streaming ability in cochlear implant users. Normal-hearing subjects perform better on tasks of speech perception in a competing talker background compared to steady-state noise (Duquesnoy, 1983; Festen and Plomp, 1990). In contrast, cochlear implant subjects experience an increase in masking with competing talker(s) compared to steady-state noise (Qin and Oxenham, 2003; Turner *et al.*, 2004), which is the opposite of the trend seen for normal-hearing subjects. One reason that cochlear implant subjects may have more trouble understanding speech in a competing talker background is that they may not be able to segregate the target from background speech in a way that allows them to correctly identify the target speech. Thus, they are not able to take advantage of the temporal and spectral gaps in the competing talker noise to obtain additional glimpses of the target speech and experience the release from masking observed in normal-hearing listeners. This leads us to hypothesize that this inability of cochlear implant subjects to experience a release from masking may also be correlated to pure-tone streaming ability.

## B. Participants

All cochlear implant subjects who participated in experiment 1 also participated in this experiment. All subjects were tested on speech perception in steady noise and in multi-talker babble. Eight additional cochlear implant subjects (ages 33–78) were recruited to undertake an abridged version

of the pure-tone streaming task described in experiment 1 in order to increase the sample size for the correlations of streaming ability with speech perception in noise. All cochlear implant subjects had at least 1 year of experience with their device at the time of testing. The age, type of implant, signal processing strategy, and stimulation mode of each participating cochlear implant subject are shown in Table I.

## C. Stimuli and procedures

### 1. Pure-tone streaming task (abridged version)

The full version of the streaming rhythm task used in experiment 1 allowed for the measurement of time delay thresholds at many different frequency differences; these data were subsequently fit with a regression line, with the slope interpreted as an overall measure of streaming ability. However, such testing was extremely time consuming, requiring about 6 h per subject. To shorten testing time in experiment 3, we opted to use an abridged version of the streaming rhythm task, where the normalized value of the time delay threshold at a Weber fraction of 0.5 was taken as the streaming metric. This normalized threshold represents how many times more difficult the streaming rhythm task was at a Weber fraction of 0.5 than at baseline (Weber fraction=0). The use of this metric allowed us to more efficiently increase the sample size for the correlation analysis performed in this experiment. The abridged version of the streaming rhythm task is identical to the full version described in experiment 1, except the only conditions tested were those where tones A and B differed by frequency differences corresponding to Weber fractions of 0 and 0.5. The frequency difference corresponding to a Weber fraction =0.5 was chosen to be tested because from experiment 1, it appeared that the results from cochlear implant subjects at this condition demonstrated the largest variation across subjects and exhibited minimal ceiling effects. Large variation across subjects in the rhythmic discrimination threshold was desirable because we wanted to subsequently correlate our measure of streaming ability with speech perception in noise. Furthermore, values for the regression slope streaming metric used in experiment 1 were strongly correlated with normalized thresholds at a Weber fraction of 0.5 in experiment 3 (200 Hz base frequency:  $r=0.881$ ,  $p<0.005$ ; 800 Hz base frequency:  $r=0.830$ ,  $p<0.025$ ; 2000 Hz base frequency:  $r=0.906$ ,  $p<0.0025$ ), suggesting consistency between the two measures of streaming in cochlear implant users.

Testing was once again performed at all three base frequencies (200, 800, and 2000 Hz) in a randomized order. At each base frequency, the first condition tested was that where the Weber fraction was 0, followed by the condition where the Weber fraction was 0.5. The exclusion criterion of an average raw score time delay threshold below 15 ms with the two tones equal in frequency (Weber fraction=0) was maintained for the abridged version of the test. Two subjects did not meet this criterion at one out of three base frequencies: both subjects CI11 and CI15 were unable to meet this criterion at the 2000-Hz base frequency condition.

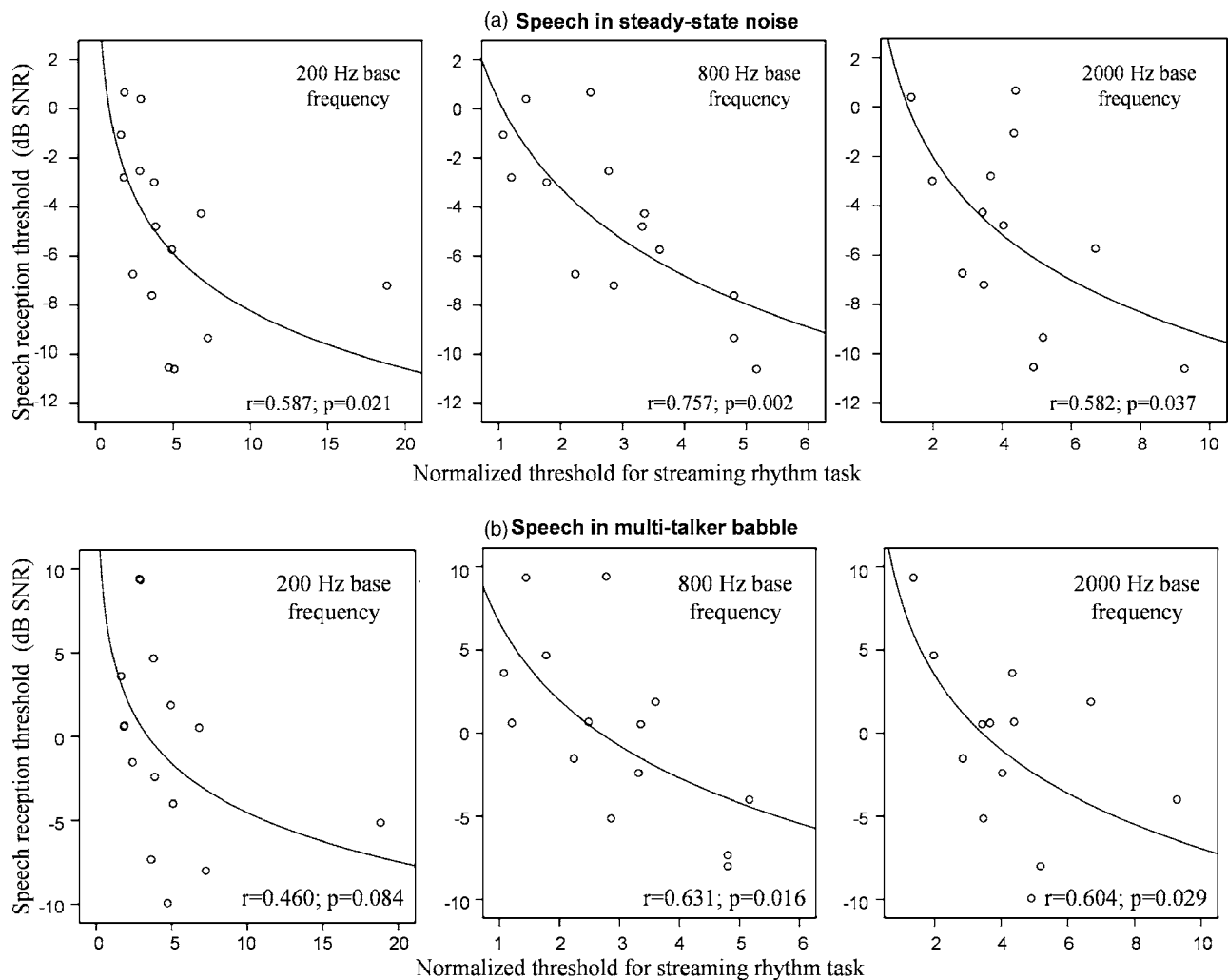


FIG. 7. Correlations between streaming rhythm task and speech in noise for cochlear implant subjects. Normalized time delay thresholds for the streaming rhythm task at a Weber fraction of 0.5 at each of three base frequencies (200, 800, and 2000 Hz) are plotted against performance on (a) speech in steady-state noise and (b) speech in two-talker babble.

## 2. Speech perception in steady-state noise and multi-talker babble

The tasks of speech perception in steady-state noise and multi-talker babble used in this study are the same as those previously described by Turner *et al.* (2004). The speech in these tasks was presented via a loudspeaker situated directly in front of the listener at 70 dB SPL. The most pertinent details of these tasks will be described in the following paragraph. For a more complete description of the tasks, the reader may refer to Turner *et al.* (2004).

Subjects were asked to identify a randomly chosen, previously recorded spondee from a fixed set of 12 homogeneously difficult spondees. The spondee was spoken by a female talker (fundamental frequency range: 212–250 Hz) in the presence of varying levels of background sound. Two different backgrounds were employed. In the first task, the background was a steady-state white noise, low-pass filtered at  $-12$  dB per octave above 400 Hz to resemble the long-term spectrum of speech. In the second task, the background was two simultaneously presented sentences from the SPIN test (Bilger, 1984), with one sentence spoken by a male talker (fundamental frequency range: 81–106 Hz) and the

other sentence spoken by a female talker (fundamental frequency range: 149–277 Hz). Within each task, the same noise sample was used for the background from trial to trial. An adaptive procedure with 2-dB step sizes (for the varying background level) was used to determine the 50%-correct point (in terms of signal-to-noise ratio) for the speech recognition threshold (SRT) for spondees in noise. Each run was composed of 14 reversals, with the mean of the last ten reversals taken as threshold for that run. Each subject completed four or five runs, with the mean of the last three runs taken as SRT for that subject.

## D. Results and discussion

This experiment was designed to determine if a correlation exists between pure-tone streaming ability and speech perception in noise. The normalized time delay threshold (calculated in the same manner as for experiment 1; see Fig. 2) for the rhythmic discrimination streaming task was correlated using exponential fits with performance on speech in steady-state noise and speech in multi-talker babble for all cochlear implant subjects. The results are shown in Fig. 7 for

each of the three base frequencies of the rhythm streaming task. Figure 7(a) shows the correlations with speech in steady-state noise, and Fig. 7(b) shows the correlations with speech in multi-talker babble. In each case, there was a negative correlation between performance on the streaming rhythm task and speech in noise, suggesting that cochlear implant subjects who had larger time delay thresholds (and therefore better streaming abilities) performed better on speech recognition in noise. The correlations were of moderate strength, with  $r$  values ranging from 0.460 to 0.757, as shown in Fig. 7. All of the correlations were statistically significant ( $p < 0.05$ ), except for the correlation between the streaming rhythm task at 200-Hz base frequency and speech in multi-talker babble ( $p = 0.084$ ). Exponential fits were chosen to better account for the one data point at the 200-Hz base frequency on the streaming rhythm task that was much higher than all of the others, with a value close to 18 for the normalized threshold. However, all of the data were also fit with linear regression curves (not shown), which yielded similarly moderate correlations and statistically significant results. Additionally, statistical analysis performed to determine if the correlation coefficients were significantly different from each other revealed no significant differences, whether comparing  $r$  values within each speech perception in noise task at different base frequencies ( $p > 0.4$ ), or comparing  $r$  values between the two speech perception tasks at the same base frequency ( $p > 0.5$ ).

The cochlear implant subjects in this study perform on average 4.5 dB worse (standard deviation of 3.53 dB) on the speech perception in two-talker babble task than the speech perception in steady-state noise task. For comparison's sake, Turner *et al.* (2004) has reported for the same tasks that normal-hearing subjects perform about 13 dB better on speech perception in babble. Correlations performed between the difference in speech recognition thresholds (i.e., examining the lack of masking release) for the two tasks and the normalized time delay thresholds of the streaming rhythm task at each of three base frequencies revealed no significant correlations ( $p > 0.35$  for all correlations).

The results of the streaming rhythm task may underestimate the amount of stream segregation when the subject cannot hear the irregular rhythm even at the largest possible time delay. However, this problem occurred only in a few cases. Of the 42 unique values for threshold obtained at a Weber fraction=0.5 (across subjects and base frequencies), only four of them were assigned the nominal (ceiling) value of 40 ms for threshold. Thus, ceiling effects did not appear to be a major problem with our task at the frequency difference where the correlations with speech perception in noise were performed.

The correlation between pure-tone stream segregation and speech perception in noise found in cochlear implant subjects is consistent with findings in the literature that suggest that frequency resolution is an important determinant of the ability to perceive speech in noise (Fu *et al.*, 1998). Diminished frequency resolution likely leads to both poorer pure-tone auditory streaming and poorer speech perception in noise. However, it is likely that other common factors, beyond peripheral processes, are also involved in auditory

stream segregation and speech perception in noise. For example, common central processes such as selective attention have been shown to influence auditory stream segregation (Carlyon *et al.*, 2001) and certainly also play a role in the ability to understand speech in noise. We propose that the normalized thresholds from the streaming rhythm task are a measure of the summation of relevant peripheral and central processing abilities within a subject. For example, a cochlear implant subject with poor peripheral frequency resolution and extremely good cognitive processing may perform equally well on pure-tone streaming as a different cochlear implant subject with better frequency resolution but poorer cognitive processing. The shared importance of these peripheral and central processes with speech perception in noise could thus explain why a correlation is observed between streaming and speech perception in noise in cochlear implant subjects.

The observation that the weakest correlation between streaming and speech perception in noise occurred between the 200-Hz base-frequency streaming condition and speech perception in two-talker babble was unexpected. We had hypothesized that we may find the strongest correlation here, because the 200-Hz region is where the fundamental frequencies of the different talkers are found, and the ability to segregate talkers based on fundamental frequencies is thought to be important for understanding speech in noise. However, there may be a number of reasons for the discrepancy between our hypothesis and our results. First, the ability of cochlear implant users to stream pure tones (as measured by the streaming rhythm task) may be different from their ability to stream more complex, speechlike stimuli. This is because cochlear implant users may have differing abilities to extract the fundamental frequencies from more complex tones. Thus, pure-tone streaming at a base frequency of 200 Hz may not be an accurate measure of streaming based on fundamental frequency. Second, a number of recent studies have suggested that current cochlear implant users who rely solely on electric hearing are deficient in or receive little to no benefit from fundamental frequency cues for understanding speech in a competing talker background (Qin and Oxenham, 2005; Kong *et al.*, 2005). This suggests that cues other than fundamental frequency may currently be more important for talker segregation by cochlear implant users. Thus, correlations between speech perception in noise and streaming ability across all frequency regions may be similar, because frequency cues in all regions may be used by cochlear implant users to segregate speech from noise. Finally, the correlations between speech perception and streaming at different base frequencies may be similar because they are correlated with some common underlying factor. For example, central processing abilities may be related to both auditory stream segregation and speech perception in noise.

We also hypothesized that the increase in masking experienced by cochlear implant subjects with speech perception in two-talker babble compared to steady-state noise may be correlated with streaming ability. One possible reason for the lack of such correlations in our results is that the ability to stream speech from different talkers may be different from the ability to correctly identify which speech belongs to the

target and which to the masker(s). In order to take advantage of the temporal and spectral gaps in the competing talker noise to understand speech, one must be able to do both. Thus, variability in the ability to correctly identify the target speech among subjects with similar streaming abilities may account for the lack of correlation between streaming and relative performance on the two speech in noise tasks.

## V. CONCLUSIONS

- (i) The results of this study suggest that the streaming rhythm task (adapted from Roberts *et al.*, 2002) is a test paradigm that can be used to assess auditory stream segregation in cochlear implant subjects. The task is sensitive to detection of a range of pure-tone streaming abilities that is present within the cochlear implant population.
- (ii) There is a range of streaming abilities both within individual cochlear implant subjects (comparing different regions of the cochlea) and between cochlear implant subjects. Some cochlear implant subjects perform comparably to normal-hearing subjects, while others experience much less streaming than normal-hearing subjects as the frequency separation of two alternating tones is increased.
- (iii) The variability in pure-tone streaming abilities across a wide range of frequencies among different cochlear implant users correlates moderately well with their ability to perceive speech in both steady-state noise and multi-talker babble.
- (iv) For normal-hearing listeners, many cues other than frequency have been shown to be available for stream segregation, including those based on amplitude spectra (Iverson, 1995), temporal envelope (Grimault *et al.*, 2002), and spatial location (Dowling, 1973). It will be important in the future to determine which of these cues available to normal-hearing individuals are also available to cochlear implant users for auditory stream segregation, as well as the utility of these additional streaming cues for the perception of speech in complex acoustical backgrounds.

## ACKNOWLEDGMENTS

This research was supported by Grant Nos. R01 DC000377, P50 DC00242, and 5 T32 DC000040-13 from the National Institutes of Health. The authors thank Anna Yao and Sheryl Erenberg for help in collecting the data, and Arik Wald for help with programming the experiments. We also appreciate the thoughtful comments and suggestions on this manuscript from John Grose, Fan-Gang Zeng, and one anonymous reviewer.

Anstis, S., and Saida, S. (1985). "Adaptation to auditory streaming of frequency-modulated tones," *J. Acoust. Soc. Am.* **11**, 257–271.  
 Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **88**, 680–697.  
 Beauvois, M. W., and Meddis, R. (1996). "Computer simulation of auditory

stream segregation in alternating-tone sequences," *J. Acoust. Soc. Am.* **99**, 2270–2280.

Bilger, R. C. (1984). "Speech recognition test development," ASHA Reports no. 14, pp. 2–15.  
 Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).  
 Brox, J. P. L., and Nootboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics* **10**, 23–36.  
 Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (2001). "Effects of attention and unilateral neglect on auditory stream segregation," *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 115–127.  
 Chatterjee, M., Fu, Q.-J., and Shannon, R. V. (1998). "Within-channel gap detection using dissimilar markers in cochlear implant listeners," *J. Acoust. Soc. Am.* **103**, 2515–2519.  
 Dorman, M. F., Smith, M., Smith, L., and Parkin, J. L. (1996). "Frequency discrimination and speech recognition by patients who use the Ineraid and continuous interleaved sampling cochlear-implant signal processors," *J. Acoust. Soc. Am.* **99**, 1174–1184.  
 Dowling, W. L. (1973). "The perception of interleaved melodies," *Cognit. Psychol.* **5**, 322–337.  
 Dreschler, W. A., and Plomp, R. (1985). "Relations between psychophysical data and speech perception for hearing-impaired subjects. II," *J. Acoust. Soc. Am.* **78**, 1261–1270.  
 Duquesnoy, A. J. (1983). "Effect of a single interfering noise or speech source on speech intelligibility," *J. Acoust. Soc. Am.* **74**, 739–743.  
 Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.  
 Formby, C. and Forrest, T. G. (1991). "Detection of silent temporal gaps in sinusoidal markers," *J. Acoust. Soc. Am.* **89**, 830–837.  
 Fu, Q.-J., Shannon, R. V., and Wang, X. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.  
 Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (2002). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," *Cochlear Implants Int.* **3**, 29–53.  
 Gfeller, K. E., Olszewski, C., Rychener, M., Sena, K., Knutson, J., Witt, S., and Macpherson, B. (2005). "Recognition of 'real-world' musical excerpts by cochlear implant recipients and normal-hearing adults," *Ear Hear.* **26**, 237–250.  
 Grimault, N., Bacon, S. P., and Micheyl, C. (2002). "Auditory stream segregation on the basis of amplitude-modulation rate," *J. Acoust. Soc. Am.* **111**, 1340–1348.  
 Grose, J. H., and Hall, J. W. (1996). "Perceptual organization of sequential stimuli in listeners with cochlear hearing loss," *J. Speech Hear. Res.* **39**, 1149–1158.  
 Hanekom, J., and Shannon, R. V. (1997). "Gap detection as a measure of electrode interaction in cochlear implants," *J. Acoust. Soc. Am.* **104**, 2372–2384.  
 Iverson, P. (1995). "Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes," *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 751–763.  
 Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (2005). "Speech and melody recognition in binaurally combined acoustic and electric hearing," *J. Acoust. Soc. Am.* **117**, 1351–1361.  
 Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.  
 Mackersie, C. L., Prida, T. L., and Stiles, D. (2001). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," *J. Speech Lang. Hear. Res.* **44**, 19–28.  
 Miller, G. A. and Heise, G. A. (1950). "The trill threshold," *J. Acoust. Soc. Am.* **22**, 637–638.  
 Neff, D. L., Jesteadt, W., and Brown, E. L. (1982). "The relation between gap discrimination and auditory stream segregation," *Percept. Psychophys.* **31**, 493–501.  
 Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.  
 Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (1997). "Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," *J. Acoust. Soc. Am.* **101**, 3694–3705.

- Pichora-Fuller, M. K. (2003). "Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension," *Int. J. Audiol.* **42**, 559–567.
- Qin, M. K., and Oxenham, A. J. (2003). "Effects of simulated cochlear-implant processing on speech perception in fluctuating maskers," *J. Acoust. Soc. Am.* **114**, 446–454.
- Qin, M. K., and Oxenham, A. J. (2005). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," *Ear Hear.* **26**, 451–460.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). "Primitive stream segregation of tone sequences without difference in fundamental frequency or passband," *J. Acoust. Soc. Am.* **112**, 2074–2085.
- Rose, M. M., and Moore, B. C. J. (1997). "Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **102**, 1768–1778.
- Rose, M. M., and Moore, B. C. J. (2005). "The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects," *Hear. Res.* **204**, 16–28.
- Snell, K. B., and Frisina, D. R. (2000). "Relationships among age-related differences in gap detection and word recognition," *J. Acoust. Soc. Am.* **107**, 1615–1626.
- Stickney, G. S., Zeng, F-G., Litovsky, R., and Assman, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Strouse, A., Ashmead, D. H., Ohde, R. N., and Wesley, D. G. (1998). "Temporal processing in the aging auditory system," *J. Acoust. Soc. Am.* **104**, 2385–2399.
- Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (2004). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," *J. Acoust. Soc. Am.* **115**, 1729–1735.
- Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (1982). "Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners," *J. Acoust. Soc. Am.* **72**, 740–753.
- van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.
- van Wieringen, A., and Wouters, J. (1999). "Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees," *J. Acoust. Soc. Am.* **106**, 1925–1939.