Sequential stream segregation using temporal periodicity cues in cochlear implant recipients^{a)}

Robert S. Hong^{b)} and Christopher W. Turner

Department of Otolaryngology–Head and Neck Surgery and Department of Communication Sciences and Disorders, 121B Wendell Johnson Speech and Hearing Center, University of Iowa, Iowa City, Iowa 52242-1012

(Received 19 June 2008; revised 1 May 2009; accepted 1 May 2009)

Sequential stream segregation involves the ability of a listener to perceptually segregate two rapidly alternating sounds into different perceptual streams. By studying auditory streaming in cochlear implants (CIs), one can obtain a better understanding of the cues that CI recipients can use to segregate different sound sources, which may have relevance to such everyday activities as the understanding of speech in background noise. This study focuses on the ability of CI users to use temporal periodicity cues to perform auditory stream segregation. A rhythmic discrimination task involving sequences of alternating amplitude-modulated (AM) noises is used. The results suggest that most CI users can stream AM noise bursts at relatively low modulation frequencies (near 80 Hz AM), but that this ability diminishes at higher modulation frequencies. Additionally, the ability of CI users to perform streaming using temporal periodicity cues appears to be comparable to that of normal-hearing listeners. These results imply that CI subjects may in certain contexts (i.e., when the talker has a low fundamental frequency voice) be able to use temporal periodicity cues to segregate and thus understand the voices of competing talkers.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3140592]

PACS number(s): 43.66.Mk, 43.66.Ts, 43.66.Lj, 43.66.Hg [RYL]

Pages: 291–299

I. INTRODUCTION

Cochlear implants (CIs) have been remarkable in their ability to restore hearing to deaf individuals. Many CI users experience near-total to total understanding of everyday speech in quiet conditions. However, their ability to understand speech deteriorates significantly in conditions with background noise. The understanding of speech in background noise is one of the biggest challenges facing CI listeners today. One approach (the approach advocated by this study) to examining the factors that may be most relevant to improving the understanding of speech in noise by CI users is to examine different auditory cues individually in a laboratory setting—specifically, by using psychoacoustic tests that assess auditory stream segregation. Auditory stream segregation is the process used to separate a complex sound into different perceptual streams, corresponding to the different individual sources from which the sound is derived. The ability to stream (pure tones) has been shown to be correlated with the understanding of speech in noise in both hearing-impaired (Mackersie *et al.*, 2001) and CI listeners (Hong and Turner, 2006), consistent with the idea that a better ability to segregate sounds leads to a better understanding of speech in competing backgrounds. Auditory streaming has been studied extensively in normal-hearing (NH) listeners, with the suggestion that any salient difference between sounds can

potentially be used as a cue for segregating sounds (Moore and Gockel, 2002). In contrast, few studies have been performed to date on auditory streaming in CI users (Hong and Turner, 2006; Chatterjee *et al.*, 2006; Cooper and Roberts, 2007), with these studies generally suggesting that streaming may occur with CIs but on a limited basis.

The study of sequential stream segregation is performed in the laboratory with sequences composed of two different sounds (sound A and sound B) that alternate rapidly in time. When these sounds are similar, they are often heard in a single perceptual stream fluctuating between sounds A and B. As the sounds become more different, this percept changes to that of two different streams, one composed of repeating sound A's and the other of repeating sound B's. The traditional method of assessing sequential stream segregation in NH subjects is to ask subjects to subjectively indicate whether they hear one or two streams when listening to a sequence of alternating sounds. While such a method may be practical for NH subjects, it is unclear if such a method can accurately assess streaming in CI subjects, since sound is perceived differently through a CI than through NH ears. Thus, CI subjects may not understand what is meant by hearing one stream versus two streams, since they may not be able to perceive or describe such a difference. As a result, they may respond inaccurately with respect to streaming, because they may use some other percept (unrelated to streaming) to guide their decisions.

The research in this study attempts to bypass the concerns associated with the previously described methodology by using rhythmic discrimination procedures to assess sequential stream segregation in CIs. Roberts *et al.* (2002) and

^{a)} Portions of this work were presented in “Auditory stream segregation using temporal periodicity cues in cochlear implants,” Midwinter Research Meeting of the Association for Research in Otolaryngology, Denver, CO, 2007.

^{b)} Author to whom correspondence should be addressed. Electronic mail: robert-hong@uiowa.edu

Stainsby *et al.* (2004) compared results obtained using the traditional subjective method to assess streaming and one based on a rhythmic discrimination task, with similar results obtained from each methodology. With respect to CI subjects, there are two primary advantages to using rhythmic discrimination to measure streaming. First, the concept of rhythm (unlike that of streaming) is likely to be familiar to CI users, since the rhythmic perception ability of CI users is similar to that of NH subjects (Gfeller and Lansing, 1991). Thus, CI users will more likely use the appropriate perceptual criteria to perform the task. Second, a rhythmic discrimination streaming task has a clear right and wrong answer, unlike the earlier described tasks where listeners are asked to indicate if they think they hear one or two streams. This allows an adaptive procedure to be implemented that can converge on a particular streaming threshold, providing the increased efficiency and reliability enjoyed by adaptive procedures (Leek, 2001) to the proposed tasks.

The ability of CI recipients to use rate pitch cues (also referred to in this paper as temporal envelope or temporal periodicity cues) for sequential stream segregation is not known. Shannon (1992) suggested that CI users have a poor ability to detect temporal envelope cues greater than 300 Hz. Thus, one may expect that the ability of CI users to use rate pitch cues for streaming, if present, would be limited to cues less than 300 Hz. However, even a potentially limited range of temporal envelope cues to low frequencies (<300 Hz) for streaming by CI users is important to study. Low-frequency temporal envelope cues correspond to the fundamental frequencies of many human voices, and fundamental frequency is one of the most important cues for the segregation of speech from background noise, at least in NH listeners (Brokx and Nootboom, 1982). The ability to stream using low-frequency rate pitch cues is particularly important to examine in CI users because this may be the dominant cue used by CI listeners for streaming. Because place pitch cues are relatively unreliable compared to rate pitch cues for conveying pitch in current CI users (Geurts and Wouters, 2001; Laneau *et al.*, 2004), rate pitch cues may take on a relatively more important role for sound segregation. Our study uses sinusoidal amplitude-modulated (AM) noise bursts to determine the extent to which CI recipients can use rate pitch cues for sequential stream segregation. If low-frequency rate pitch cues are usable by CI listeners for streaming, then this would suggest that future speech processing strategies designed to enhance these cues may improve the perception of human vocal fundamental frequency and speech in background noise.

In our previous study of sequential stream segregation in CI recipients (Hong and Turner, 2006), we utilized the method requiring subjects to detect an irregular rhythm described by Roberts *et al.* (2002) to successfully assess auditory streaming of pure tones in CI users. However, attempts to adapt this method to measure streaming with temporal envelope cues using AM noise bursts were problematic. It appeared that subjects were able to perform the task while ignoring the envelope differences for different AM noise bursts; the common timbre of the noise, which can be a strong cue for grouping (Dannenbring and Bregman, 1976;

Cusack and Roberts, 2000), allowed subjects to integrate all the sounds in the same stream and hear the irregular rhythm, irrespective of AM rate. This observation is also consistent with that of Vliegen *et al.* (1999), who, using a similar rhythmic discrimination task to that used in our previous CI streaming experiments, found little evidence of streaming for complex tones with unresolved harmonics, presumably for a similar reason. Thus, it was necessary to develop a different method to assess the streaming with temporal periodicity cues in CI users, one that required the successful segregation of sounds into two different streams as opposed to integration of sounds into a single stream to perform the task. This methodology is similar to that used by Hartmann and Johnson (1991) and Dowling (1973) to assess streaming, except it is based on interleaved rhythms as opposed to interleaved melodies. The interleaved rhythmic discrimination task that we developed to assess the sequential stream segregation of AM noise bursts is described in Sec. II.

Two types of streaming experiments involving interleaved rhythmic discrimination are used in this study to probe the AM streaming abilities of CI users. The goal of the first type of experiment is to assess the AM streaming ability of CI users under “optimal” conditions—that is, when the temporal modulation cue should be the strongest. Shannon (1992) demonstrated that CI users are most sensitive to temporal modulations at lower modulation frequencies (80–100 Hz AM), and so a modulation frequency of 80 Hz AM is chosen as a base frequency for this experiment. Furthermore, all the stimuli used for this experiment are 100% modulated, since the ability to detect modulations decreases as modulation depth decreases. This experiment thus measures the ability of listeners to stream a fully-modulated 80 Hz AM stimulus apart from a fully-modulated AM stimulus at a higher AM rate to determine if CI users have any ability to stream AM stimuli.

The second type of streaming experiment is designed to provide insight into the AM frequencies at which temporal envelope cues can be used for streaming. This second type of experiment measures the modulation depth at which listeners are no longer able to stream an 80, 200, or 300 Hz AM stimulus apart from an unmodulated noise. If CI users are able to perform AM streaming, this ability may decrease as the modulation AM frequency is increased from 80 to 300 Hz AM, given the increasingly poor ability of CI users to detect modulation at AM frequencies higher than 300 Hz (Shannon, 1992). Extending this hypothesis to the real world setting of speech perception in noise, such a trend would suggest that temporal envelope cues may be more useful to CI users for segregating a male speaker ($F_0=80\text{--}120$ Hz) than a child speaker ($F_0\approx$ approximately 300 Hz) from a competing background.

II. METHODS

A. Participants

10 CI and 12 NH subjects participated in these experiments. All of the CI listeners used the Hi-Resolution™ strategy, a strategy that allows transmission of the temporal periodicity cues of interest in these experiments, with an

TABLE I. CI subject demographics. The age, duration of profound deafness, length of CI use at time of testing, device, and stimulation rate for all CI subjects is shown. All of the subjects used the Hi-Resolution™ strategy.

Subject	Age (years)	Duration of deafness (years)	Length of CI use (years)	Device	Stimulation rate (pulses/s channel)
CI1	30	11	4.6	Clarion CII HF	2520
CI2	79	0.6	3.4	Clarion CII HF	2320
CI3	71	12	1.8	Clarion 90K	5156
CI4	46	1	3.4	Clarion CII HF	2047
CI5	64	10	3.2	Clarion CII HF	2320
CI6	68	1	3	Clarion CII HF	5156
CI7	41	36	4.3	Clarion 90K	5156
CI8	52	46	6.4	Clarion CII HF	5156
CI9	47	23	4.1	Clarion CII HF	4679
CI10	56	0.1	3	Clarion 90K	2184

Advanced Bionics CI (Sylmar, CA). Additionally, the rate of stimulation on each channel was required to be greater than 1200 pulses/s, to allow for the accurate representation of temporal periodicity information up to 300 Hz (McKay *et al.*, 1994; Wilson, 1997), as shown in Table I.

To minimize effects of non-streaming related factors on performance, musicians with college level musical training were excluded from testing, since these subjects may be extraordinarily good at rhythm-related tasks. Age may also have an effect on the ability of subjects to perform temporally-related tasks, with a decline in performance observed by some studies for older subjects (Pichora-Fuller, 2003; Strouse *et al.*, 1998) but not others (Takahashi and Bacon, 1992). To prevent age from playing a potentially confounding role in the assessment of streaming, the NH subjects were roughly age-matched to the CI users; the average age of NH listeners is 47.9 years and CI listeners is 55.4 years.

Pure-tone audiometry was performed on all NH listeners to confirm that they had NH. All subjects had NH (<20 dB HL) across octave frequencies (from 250 to 8000 Hz), with the exception of one listener who had a 40 dB HL at 8000 Hz and a second listener who had a 25 dB HL at 4000 Hz.

B. Stimuli and procedures

1. Interleaved rhythmic discrimination streaming task

In this study, a three-interval two-alternative forced choice, interleaved rhythmic discrimination task was used to assess auditory streaming based on temporal envelope cues. In this task, subjects were presented a target rhythm and asked to identify which of two subsequent sequences of alternating sounds contained the target rhythm. If sound A and sound B were sufficiently different along any perceptual dimension, allowing listeners to stream sound A apart from sound B, then listeners were able to hear the target rhythm (composed of sound A's) in one of the two sequences (Fig. 1). Incidentally, the listener could either listen to the sound A's alone or the sound B's alone to hear the target rhythm, since the rhythm could be found among either sound, even though the target rhythm was played for the listener during the task only with sound A.

All of the stimuli used for sound A and sound B were AM, broadband noise bursts. Each noise burst was 125 ms in duration, corresponding approximately to the duration of syllables in speech, with 10 ms linear ramps at the onset and offset of each stimulus. The starting phase of sinusoidal amplitude modulation was random for each noise burst. The target rhythm was composed of four sound A's, separated by either a long silence (175 ms duration) or a short silence (25 ms duration), as shown in Fig. 1. Each of the two sequences was composed of a unique arrangement of eight sounds (four sound A's and four sound B's), with each sound separated by 25 ms of silence, as shown in the figure. The difference between the two sequences is denoted by the arrows, where sound A and sound B are reversed. The figure shows each sequence to be composed of a set of eight sounds played two times in a row, for a total duration of 2.4 s. In the actual task, each set of eight sounds was played four times in a row, such that each sequence actually lasted 4.8 s (twice as long as shown in Fig. 1). If the listener could perceptually stream sound A apart from sound B, the listener should be able to hear that sequence 2 contained the target rhythm (shown in bold and underlined in Fig. 1).

The stimuli were presented in a double-walled sound attenuated booth through a single loudspeaker situated directly in front of the listener at 72 dB SPL, with a 3 dB loudness rove from one sound to the next within each sequence. The root-mean-square power of each noise burst was equated before applying the loudness rove. The loudness rove was used to prevent listeners from using loudness cues, as opposed to rate pitch cues, to stream AM noise bursts, with a 3 dB range chosen to cover the range of different loudness that may be perceived by subjects for stimuli of different AM rates (Zhang and Zeng, 1997). All stimuli were stored at 44.1 kHz and passed through a low-pass 20 kHz filter before presentation to avoid aliasing. Feedback was provided on all tests.

To address the possibility that some subjects may not be able to perform the AM streaming task because of difficulties with discriminating rhythms in general, irrespective of streaming, all subjects were first screened with an initial task that required subjects to correctly identify the target rhythm in the absence of an interleaved rhythm. This test was iden-

Which sequence contains the target rhythm?

Target rhythm:

FIG. 1. Interleaved rhythmic discrimination task. The target rhythm is composed of four sound A's played in the rhythm shown. Each sequence is composed of a unique arrangement of eight sounds (four sound A's and four sound B's), which in the figure are repeated twice. The difference between the two sequences is denoted by the arrows, where sound A and sound B are reversed. If the listener can perceptually stream sound A apart from sound B, the listener will be able to hear that sequence 2 contains the target rhythm (shown in bold and underlined).

tical to the interleaved rhythm task, except all of the sound B's in the streaming task (as shown in Fig. 1) were replaced with silent gaps equal in duration to sound B. To pass the basic rhythm discrimination screening task, subjects needed to correctly identify the sequence that matched the target rhythm in five consecutive trials. Subjects that could not perform the simple rhythm discrimination task were excluded from further testing. In fact, one NH and one CI subject (not shown in Table I) did not pass the simple rhythm discrimination task and thus did not participate in any of the streaming experiments.

The streaming of AM noise bursts was assessed using four different conditions. In the first condition, the ability to stream was measured as a function of separation in amplitude-modulation rate between sound A and sound B, with both sounds fully modulated. Sound A was always an 80 Hz AM noise burst. Sound B was always a noise burst with a higher rate of AM compared to sound A, with sound B varied in logarithmic steps (sound B = 80.5, 81, 82, 84, ..., 1104 Hz AM) to determine the AM frequency separation needed for auditory stream segregation. In the other three conditions, the ability to stream was measured as a function of modulation depth between a modulated sound A (fixed at 80, 200, or 300 Hz AM) and an unmodulated sound B. Sound A was varied in linear steps of modulation depth to determine the modulation depth needed to stream a modulated AM noise burst from an unmodulated one. In all four conditions, a two-down one-up adaptive staircase that converged on the 70.7% correct point of the psychometric function was used to determine the difference needed between sound A and sound B that resulted in a detectable target rhythm. In the first condition, where the task adapted based on AM frequency, there were eight total reversals in each run. The first two reversals were taken as practice and had a multiplicative step size of 4. The geometric

mean of the final six reversals, with a multiplicative step size of 2, was taken as the threshold for a single run. In the other three conditions, where the task adapted based on modulation depth, there were ten total reversals in each run. The first four reversals were taken as practice, with a linear step size of 20% modulation depth for the first two steps and 10% modulation depth for the next two steps. The arithmetic mean of the final six reversals, with a linear step size of 4% modulation depth, was taken as the threshold for a single run. In each of the four conditions, a total of four consecutive runs were performed, with the mean of the final three runs taken as the threshold for each condition for each subject. Additionally, during the first practice run, it was sometimes necessary to introduce a loudness cue of up to 20 dB to sound A relative to sound B, to help the subjects understand the task. This cue was incrementally reduced until no longer present during the practice run, and it was not present in any of the subsequent three runs, where actual data were collected. Finally, subjects who exceeded the upper bound of the task after completing the required practice run for a particular condition were recorded as being unable to perform the task at that condition.

2. Amplitude-modulation discrimination tasks

Each of the four conditions tested in the AM interleaved rhythmic discrimination experiments had a corresponding AM discrimination test condition. The AM discrimination task, like the interleaved rhythm task, is a three-interval two-alternative forced choice task that adapts based on a two-down one-up rule. In this task, three consecutive AM noise bursts were played. Listeners were asked to decide if the second or third noise burst sounded different than the first, ignoring differences in loudness (since a 3 dB loudness rove was implemented among the noise bursts). The AM discrimi-

FIG. 2. Group performance on amplitude-modulation streaming and rate discrimination for tasks adapting based on AM frequency. The average results of the NH and CI groups are shown. All AM stimuli are fully modulated. The base frequency (sound A) is 80 Hz AM. The results demonstrate the threshold difference in AM frequency between sound A and sound B (with sound B always greater than 80 Hz AM) needed to stream or discriminate the two sounds. Error bars represent one standard deviation above the mean.

nation task that corresponded to the AM interleaved rhythm task adapting on AM frequency can be thought of as an AM rate discrimination task, where the base frequency was a fully-modulated 80 Hz AM noise burst. The three AM discrimination tasks that corresponded to the three AM interleaved rhythm tasks adapting on modulation depth can be thought of as three different AM detection experiments, with base frequencies of 80 Hz AM, 200 Hz AM, and 300 Hz AM, respectively. The composition of the different AM noise bursts, the step sizes, and the number of reversals used were identical to those found in the corresponding AM interleaved rhythmic discrimination task, as this allowed for a direct comparison of the ability to discriminate between AM noise bursts versus the ability to use such a perceptual difference between AM noise bursts for auditory stream segregation.

III. RESULTS

The average results for NH and CI subjects on the AM streaming task between fully-modulated AM noise bursts are shown in Fig. 2. In this task, sound A was always an 80 Hz AM noise burst. Sound B was a noise burst of a higher AM frequency, and thus the AM frequency difference threshold was the number of Hz above 80 Hz AM needed to be able to perform the task. As the figure demonstrates, the abilities of NH and CI subjects to stream AM stimuli on this task are comparable ($t_{df=20}=0.264$; $p>0.5$). However, there was a wide range of abilities among different individuals of both groups for AM streaming (Fig. 3). Some were excellent at streaming AM stimuli (e.g., NH 7 or CI1), while others could not perform the task at all (e.g., NH 4, NH 6, NH 10, and CI 10). For purposes of statistical analysis, all values for the AM frequency difference threshold above 300 Hz, including those where subjects were not able to perform the task (out of range=OOR), were taken as 300 Hz. This statistical cutoff

FIG. 3. Individual performance on amplitude-modulation streaming and rate discrimination for tasks adapting based on AM frequency. The results of each of the (a) 12 NH and (b) 10 CI listeners are shown. All AM stimuli are fully modulated. The base frequency (sound A) is 80 Hz AM. The results demonstrate the threshold difference in AM frequency between sound A and sound B (with sound B always greater than 80 Hz AM) needed to stream or discriminate the two sounds. Error bars represent one standard deviation above the mean. "OOR" indicates that the results are "out of range," indicating that the listener is unable to perform the task, even at the largest AM frequency differences. Error bars represent one standard deviation above the mean.

was chosen since 300–400 Hz AM corresponds to the range reported by others where the ability to detect amplitude modulation, and thus presumably to use such cues for streaming, is severely limited for both NH and CI listeners (Shannon, 1992). (This poor ability to use a 300 Hz AM temporal cue for streaming was confirmed by other experiments in this study, as will be shown later in Fig. 4.)

Figure 2 also demonstrates that the ability of both NH and CI subjects to perform AM rate discrimination was significantly better than the ability to perform AM streaming (for NH, $t_{df=11}=3.604$, $p<0.005$; for CI, $t_{df=9}=3.378$, $p<0.01$). This suggests that it was not the ability to detect differences in AM rate that limited the ability of listeners to perform AM streaming. An analysis of the individual data shown in Fig. 3 further confirmed this point. All four sub-

FIG. 4. Group performance on amplitude-modulation streaming and detection for tasks adapting based on modulation depth. The average results of the NH and CI groups are shown for stimuli at 80, 200, or 300 Hz amplitude modulation. The results demonstrate the threshold modulation depth for sound A that is needed to stream or discriminate it apart from an unmodulated sound B at each of these three AM rates. Error bars represent one standard deviation above the mean.

jects who could not perform the AM streaming task (NH 4, NH 6, NH 10, and CI 10) were able to perform the AM rate discrimination task with relative ease.

The average results of both NH and CI subjects for AM streaming at different AM base frequencies as a function of modulation depth are shown in Fig. 4. There was no significant difference between NH and CI subjects on AM streaming at base frequencies of 80 Hz AM, 200 Hz AM, or 300 Hz AM ($p > 0.2$ for each of the three conditions). However, a repeated-measures analysis of variance (ANOVA) demonstrated that there was a significant difference within groups across base frequencies for both NH ($F_{2,22}=19.009$; $p < 0.001$) and CI ($F_{2,18}=23.46$; $p < 0.001$) listeners. Post-hoc pair-wise comparisons (Bonferroni) showed that for NH listeners, AM streaming at 80 Hz AM was significantly better than at 200 Hz AM ($p < 0.005$) and also than at 300 Hz AM ($p < 0.001$); however, AM streaming at 200 Hz AM was not significantly better than at 300 Hz AM ($p = 0.165$). For CI subjects, AM streaming results were similar to those of NH listeners, with streaming at 80 Hz AM significantly better than at 200 Hz AM ($p < 0.005$) and also than at 300 Hz AM ($p < 0.005$), and no significant difference between streaming at 200 Hz AM and 300 Hz AM ($p > 0.5$).

The individual AM streaming results for NH and CI listeners at different base frequencies are shown in Fig. 5. Within both groups, there was a range of streaming abilities. At a base frequency of 80 Hz AM, the majority of both NH and CI listeners were able to perform streaming as a function of modulation depth (9 out of 12 NH and 9 out of 10 CI subjects). In contrast, at a base frequency of 300 Hz AM, only one listener (NH 2) out of all the subjects tested (either NH or CI) was able to perform AM streaming. The ability to perform streaming at a base frequency of 200 Hz AM varied among both NH and CI subjects.

FIG. 5. Individual performance on amplitude-modulation streaming for tasks adapting based on modulation depth. The individual streaming results of each of the (a) 12 NH and (b) 10 CI listeners are shown for stimuli at 80, 200, or 300 Hz amplitude modulation. The results demonstrate the threshold modulation depth for sound A that is needed to stream it apart from an unmodulated sound B at each of these three AM rates. Values extending beyond the horizontal dotted line indicate that the subject was not able to perform the task for that particular condition.

Comparison of AM streaming versus AM detection abilities at different base frequencies suggests that both NH and CI subjects were able to detect smaller differences in modulation depth than they were able to use for streaming (Fig. 4). Paired t-tests demonstrated this to be true for NH subjects across each of the three base frequencies ($p < 0.005$), although for CI subjects this was the case only at 200 Hz AM ($p < 0.05$), though such a difference approached significance at 80 Hz AM ($p = 0.064$). Analyses of the individual results for AM streaming (Fig. 5) and for AM detection (Fig. 6) further highlight the point that the ability to stream AM stimuli did not appear to be limited by the ability to detect AM stimuli. For example, 8 out of the 11 NH listeners who could not perform AM streaming at 300 Hz AM could perform AM detection at 300 Hz AM, while for CI listeners, the same held true for 3 out of the 8 CI listeners at 200 Hz AM. Finally, it is noted that the lack of significant difference between streaming and detection for CI listeners at 300 Hz AM was because most CI subjects could neither detect [Fig. 6(b)] nor stream at this base frequency [Fig. 5(b)]; for these subjects at this test condition, it is unclear if

FIG. 6. Individual performance on amplitude-modulation detection for tasks adapting based on modulation depth. The individual results of each of the (a) 12 NH and (b) 10 CI listeners are shown for stimuli at 80, 200, or 300 Hz amplitude modulation. The results demonstrate the threshold modulation depth needed to detect the amplitude modulation at each of these three AM rates. Values extending beyond the horizontal dotted line indicate that the subject was not able to perform the task for that particular condition.

their ability to stream was limited by their ability to discriminate, since they could not perform either task.

Correlations were also performed for NH and CI subjects between age and each of the measures of AM streaming/discrimination/detection. It has been reported that age may play a role in performance on temporally-based tasks (e.g., [Strouse et al., 1998](#)). Because the age range of the listeners from each of the two groups was wide (NH = 22–68 years old; CI = 30–79 years old), correlations could be performed to examine if age also affected the results of the tasks in this study, which involved temporal modulations of the envelope. The results demonstrated significant positive correlations for NH subjects between age and AM streaming thresholds at 200 Hz base frequency ($r=0.809$; $p<0.001$), AM detection thresholds at 200 Hz ($r=0.723$; $p<0.01$), and AM detection thresholds at 300 Hz ($r=0.683$; $p<0.02$). No correlation was expected or found for AM streaming at 300 Hz base frequency, since only 1 out of 12 NH subjects could perform the task, though interestingly, the one who could perform the task was the youngest subject tested. None of the other correlations were significant for NH subjects. Additionally, no significant correlations were found for the CI group.

IV. DISCUSSION

The first objective of this study is to determine whether or not CI recipients can use temporal envelope cues for auditory stream segregation. The results from these experiments (Figs. 2–5) demonstrate that the large majority of CI recipients (nine out of ten subjects) can to some extent use such cues for streaming. This is the first study to the knowledge of the authors that has demonstrated that CI users can use temporal periodicity cues for auditory stream segregation. Additionally, these results suggest that at least some of the ability of CI recipients to perform pure-tone auditory stream segregation (as observed by [Hong and Turner, 2006](#)), specifically in the lower frequency regions (less than about 200 Hz), can be attributed to their ability to use temporal periodicity cues for streaming.

The next objective, given that most CI users could perform auditory streaming with temporal periodicity cues, is to determine how the AM streaming abilities of CI listeners compare to those of NH listeners. The results suggest that as a group, the AM streaming abilities of NH and CI subjects were similar (Figs. 2 and 4), though there was considerable variation among individuals of each group. Such a result is consistent with other findings in literature, which have suggested that, in general, the temporal perception abilities of CI recipients are similar to those of NH listeners ([Moore and Glasberg, 1988](#); [Shannon, 1989, 1992](#)).

Another question of interest is to determine the extent to which AM cues may be useful for segregating sounds in situations where such cues are degraded (i.e., when the modulation depth is not 100%). Temporal envelope cues in the real world, such as those corresponding to the fundamental frequency of a talker, are often degraded by background noise and reverberation. In examining the abilities of CI users to perform streaming using AM cues, it is clear that listeners are better able to stream using degraded AM cues at lower AM frequencies (80 Hz AM) than at higher AM frequencies (200 Hz AM), and that this ability is essentially nonexistent at the highest AM frequencies tested (300 Hz AM). This finding suggests that temporal envelope cues may be more useful in the real world for segregating the voice of a male (lower Fo) than a female or child (higher Fo) from competing backgrounds. However, further research is needed to establish the relevance of our findings to the real world setting, because the perceived strength of temporal periodicity cues represented via broadband AM noise bursts (as used in this study) may be stronger than that found in speech signals with a more limited bandwidth. Thus, the extent to which our results can be extrapolated to the streaming of limited bandwidth speech remains to be determined.

The results of this study suggest that the age of listeners may influence performance on AM detection and AM streaming tasks; significant correlations among NH listeners between age and AM detection/streaming at higher AM rates (200 and 300 Hz AM) were observed, with older subjects performing more poorly on these tasks. With respect to AM detection, our findings differ from those of [Takahashi and Bacon \(1992\)](#), which suggested that age was largely unrelated to AM detection; we do not have a good explanation for

these differences. With respect to AM streaming, we suspect that the observed effects of age may be due to the temporal nature of the rhythmic discrimination task used in this study, since performance on temporally-based tasks decreases with age (Trainor and Trehub, 1989; Strouse *et al.*, 1998). An alternative interpretation is that auditory streaming ability with temporal periodicity cues diminishes with age; however, this is less likely, since a number of studies have suggested that sequential stream segregation is preserved in older adults (Trainor and Trehub, 1989; Snyder and Alain, 2007). It is important to note, however, that age does not factor into the primary comparison made in this study (i.e., examining the ability of NH versus CI listeners to perform AM streaming), since the average age of each group was controlled to be similar (47.9 versus 55.4 years old).

Grimault *et al.* (2002) measured the ability of NH listeners to use AM cues for auditory streaming. The test paradigm and stimuli used by Grimault *et al.* (2002) were different than those of this study, as they used a traditional test paradigm asking listeners to judge whether they heard one or two streams. Nevertheless, the results of the two studies appear to be relatively consistent with respect to the ability of NH listeners to perform AM streaming, providing evidence that the interleaved rhythmic discrimination task provides a valid assessment of streaming ability. Grimault *et al.* (2002) found that for fully-modulated stimuli, the streaming threshold occurred when sound A was 100 Hz AM and sound B was roughly 200 Hz AM. Similarly, in this study, the streaming threshold for fully-modulated stimuli occurred when sound A was 80 Hz AM and sound B was roughly 215 Hz AM (or 163 Hz AM, if you exclude the results of the three NH subjects who could not perform the task). Additionally, both Grimault's study and this study demonstrate that performance on streaming worsened with shallower modulation depths, when amplitude-modulation cues are weaker.

Comparisons between AM rate discrimination (and AM detection) thresholds with AM streaming thresholds obtained for both NH and CI listeners suggest that for both groups of listeners, it is easier to discriminate (or detect) AM stimuli than it is to stream them. In fact, in a number of cases, listeners were able to perform the discrimination/detection task but not the streaming task with the same stimuli (Figs. 3, 5, and 6). The finding that discrimination/detection thresholds are smaller than streaming thresholds is consistent with those of other authors. For example, Grimault *et al.* (2002) observed that NH listeners had smaller thresholds on AM rate discrimination and AM detection tasks than AM streaming tasks. Similarly, Rose and Moore (2005) observed that both NH and hearing-impaired listeners had smaller thresholds for pure-tone frequency discrimination than pure-tone streaming. The observation that streaming thresholds are larger than discrimination/detection thresholds suggests that the ability to stream is not necessarily limited by the ability to discriminate stimuli, implying that auditory stream segregation involves more complex neural processing than that needed for auditory discrimination.

A recent study by Gaudrain *et al.* (2008) examined the ability of listeners to perform obligatory stream segregation of vowel sequences using noise-excited envelope vocoders

simulating CIs. Obligatory stream segregation is measured with test paradigms that require integration of rapidly alternating sounds into a single stream to successfully perform the task. They did not find any convincing evidence of obligatory streaming using temporal periodicity cues in these CI simulations. Our results differ from those of Gaudrain *et al.* (2008) in that our results suggest that CI listeners can use temporal periodicity cues for streaming, at least up to ~200 Hz. There are a few possible reasons for this discrepancy between our results and those of Gaudrain *et al.* (2008). First, our study measures schema-based stream segregation as opposed to obligatory stream segregation—that is, the test paradigm used in our study requires segregation of sounds into two different streams as opposed to integration of sounds into a single stream to complete the task. Thresholds for schema-based versus obligatory stream segregation are known to differ, as they represent two distinct boundaries of sequential streaming (van Noorden, 1975), and thus it is not theoretically inconsistent that temporal periodicity cues may be useful for CI recipients for schema-based but not obligatory stream segregation. Second, our study uses sinusoidal AM noise bursts while Gaudrain *et al.* (2008) utilized noise-vocoder vowels. It is possible that temporal periodicity cues are stronger in AM noise bursts than in noise-vocoder vowels, leading to the different conclusions of the two studies. Finally, our study examines CI recipients directly while Gaudrain *et al.* (2008) used CI simulations in NH individuals. While CI simulations are useful for obtaining a rough estimate of CI performance, they generally do not reflect the wide range of listening abilities found in actual CI users, leading to a third possible reason for the differing results of the two studies.

The results of this study suggest that temporal periodicity cues may be helpful for CI recipients in segregating sounds. It is important to note, however, that the CI users that participated in these tasks were carefully selected to all use the Hi-Res strategy with stimulation rates of at least 1200 pulses/s per channel—conditions that ensured that temporal periodicity cues were available and accurately presented to CI listeners. It is unlikely that all current CI recipients would be able to take advantage of AM cues for auditory stream segregation, because of the following two possible limitations to their implant. First, these cues are not even transmitted by some CIs; for example, in some Nucleus signal-processing strategies, the low-pass cutoff for the temporal envelope is effectively 125 Hz, preventing rate pitch information above 125 Hz from being transmitted by the implant (Bom Jun Kwon, Cochlear Corporation, personal communication). Second, even if temporal envelope cues are transmitted, they are often distorted as many common signal-processing strategies do not provide electrical stimulation at fast enough rates to accurately convey the full range of perceivable rate pitch cues [i.e., up to ~300 Hz rate pitch (Edgington *et al.*, 1978; Shannon, 1983; Zeng, 2002)]. It has been demonstrated that the electrical carrier pulse rate must be at least four to five times the frequency of the desired rate pitch for the temporal envelope cue to be accurately represented by electrical stimulus (McKay *et al.*, 1994; Wilson, 1997), suggesting that a carrier rate of at least 1200 pulses/

s channel ($=4 \times 300$ Hz) is necessary. The results of these experiments encourage signal-processing strategies to include more information in the temporal domain, as this may be helpful for segregating sounds in the real world, potentially leading to better speech perception in background noise by CI recipients.

V. CONCLUSIONS

A novel adaptive rhythmic discrimination task requiring successful segregation of sounds into two different streams to assess sequential stream segregation is described in this study.

This is the first demonstration to our knowledge that CI listeners can utilize purely temporal cues for sequential stream segregation.

CI and NH listeners demonstrate similar abilities to use temporal periodicity cues for streaming.

The ability of both CI and NH listeners to stream using temporal periodicity cues is poorer than their ability to detect/discriminate such cues. This suggests that CI signal-processing strategies designed to improve speech perception in noise need to consider not only improvements in signal detection but also signal streaming, since listeners in some instances can detect certain cues but not use such cues for sound segregation.

Improvements in the encoding by CIs of temporal periodicity cues up to ~ 200 Hz may be useful for CI listeners in segregating sounds and, by extension, for understanding speech in background noise.

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 Hong for help in collecting the data, and Arik Wald for help with programming the experiments.

Broxk, J. P. L., and Nootboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics* **10**, 23–36.

Chatterjee, M., Sarampalis, A., and Oba, S. I. (2006). "Auditory stream segregation with cochlear implants: A preliminary report," *Hear. Res.* **222**, 100–107.

Cooper, H. R., and Roberts, B. (2007). "Auditory stream segregation of tone sequences in cochlear implant listeners," *Hear. Res.* **225**, 11–24.

Cusack, R., and Roberts, B. (2000). "Effects of differences in timbre on sequential grouping," *Percept. Psychophys.* **62**, 1112–1120.

Dannenbring, G. L., and Bregman, A. S. (1976). "Stream segregation and the illusion of overlap," *J. Exp. Psychol. Hum. Percept. Perform.* **2**, 544–555.

Dowling, W. L. (1973). "The perception of interleaved melodies," *Cogn. Psychol.* **5**, 322–337.

Eddington, D. K., Dobelle, W. H., Brackman, D. E., Mladejovsky, M. G., and Parkin, J. L. (1978). "Auditory prostheses research with multiple channel intracochlear stimulation in man," *Ann. Otol. Rhinol. Laryngol.* **87**, 1–59.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J.-C. (2008). "Streaming of vowel sequences based on fundamental frequency in cochlear-implant simulations," *J. Acoust. Soc. Am.* **124**, 3076–3087.

Geurts, L., and Wouters, J. (2001). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Am.* **109**, 713–726.

Gfeller, K. E., and Lansing, C. (1991). "Melodic, rhythmic, and timbral perception of adult cochlear implant users," *J. Speech Hear. Res.* **34**, 916–920.

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.

Hartmann, W. M., and Johnson, D. (1991). "Stream segregation and peripheral channeling," *Music Percept.* **9**, 155–184.

Hong, R. S., and Turner, C. W. (2006). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," *J. Acoust. Soc. Am.* **120**, 360–374.

Laneau, J., Wouters, J., and Moonen, M. (2004). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," *J. Acoust. Soc. Am.* **116**, 3606–3619.

Leek, M. R. (2001). "Adaptive procedures in psychophysical research," *Percept. Psychophys.* **63**, 1279–1292.

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.

McKay, C. M., McDermott, H. J., and Clark, G. M. (1994). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," *J. Acoust. Soc. Am.* **96**, 2664–2673.

Moore, B. C. J., and Glasberg, B. R. (1988). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears," *J. Acoust. Soc. Am.* **83**, 1093–1101.

Moore, B. C. J., and Gockel, H. (2002). "Factors influencing sequential stream segregation," *Acust. Acta Acust.* **88**, 320–333.

Pichora-Fuller, M. K. (2003). "Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension," *Int. J. Audiol.* **42**, S59–S67.

Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," *J. Acoust. Soc. Am.* **112**, 2074–2085.

Rose, M. M., and Moore, B. C. J. (2005). "The relationship between stream segregation and frequency discrimination in normal hearing and hearing-impaired subjects," *Hear. Res.* **204**, 16–28.

Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man: I. Basic Psychophysics," *Hear. Res.* **11**, 157–189.

Shannon, R. V. (1989). "Detection of gaps in sinusoids and biphasic pulse trains by patients with cochlear implants," *J. Acoust. Soc. Am.* **85**, 2587–2592.

Shannon, R. V. (1992). "Temporal modulation transfer functions in patients with cochlear implants," *J. Acoust. Soc. Am.* **91**, 2156–2164.

Snyder, J. S., and Alain, C. (2006). "Sequential auditory scene analysis is preserved in normal aging adults," *Cereb. Cortex* **17**, 501–512.

Stainsby, T. H., Moore, B. C. J., Medland, P. J., and Glasberg, B. R. (2004). "Sequential streaming and effective level differences due to phase-spectrum manipulations," *J. Acoust. Soc. Am.* **115**(4), 1665–1673.

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.

Takahashi, G. A., and Bacon, S. P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech Hear. Res.* **35**, 1410–1421.

Trainor, L. J., and Trehub, S. E. (1989). "Aging and auditory temporal sequencing—Ordering the elements of repeating tone patterns," *Percept. Psychophys.* **45**, 417–426.

van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology, Eindhoven, Netherlands.

Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (1999). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," *J. Acoust. Soc. Am.* **106**, 938–945.

Wilson, B. S. (1997). "The future of cochlear implants," *Br. J. Audiol.* **31**, 205–225.

Zeng, F. G. (2002). "Temporal pitch in electric hearing," *Hear. Res.* **174**, 101–106.

Zhang, C., and Zeng, F.-G. (1997). "Loudness of dynamic stimuli in acoustic and electric hearing," *J. Acoust. Soc. Am.* **102**, 2925–2934.