Auditory Stream Segregation Using Cochlear Implant Simulations

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Yingjiu Nie

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Peggy Bull Nelson, Ph.D., Advisor

June 2010

Acknowledgements

I am eternally grateful to Dr. Peggy Nelson who inspired me with enlightening ideas and systematic directions, encouraged me with contagious passion for research and teaching, and guided me through this difficult journey. I'm exceedingly fortunate to have her as my advisor. Thank you, Peggy!.

I cannot thank more to my dissertation committee members—Drs. Bert Schlauch, Neal Viemeister, and Yang Zhang, for their constructive suggestions, thoughtful comments, and careful reading.

This work cannot exist without Dr. Christophe Micheyl's and Dr. Andrew Oxenham's inspiring unpublished research work and insightful suggestions. I am also deeply indebted to Dr. Magdalena Wojtczak for her detailed and valuable suggestions.

My great appreciation goes out to Dr. Edward Carney, who consistently offered unlimited technical support and taught me computer programming. I'm particularly thankful to Dr. Sanford Weisberg for his advice on the statistical analysis in this work.

I'm as well profoundly appreciative for the support from the Doctoral Dissertation Fellowship at the University of Minnesota and the Dissertation Research Activity Award from the College of Liberal Arts. These funds enabled me to stay focused on the dissertation work.

My biggest fortune of the seven and a half years in the University of Minnesota was to have been around with the great people in the Department. I thank Drs. Arlene Carney, Jane Carlstrom, Sarah Angerman, and Leslie Glaze for their mentorship in my clinical training and pursuit of my masters' degree; I thank Drs. Benjamin Munson and Aparna Rao for their support and friendship; I thank my colleagues in Dr. Peggy Nelson's and Dr. Yang Zhang's labs—Liz Anderson, Sharon Miller, Melanie Gregan, Michelle Lewis, and other students. Their friendship made the years of my study even more worthwhile.

I cannot complete my acknowledgement without thanking Dr. Yi-Sheng Chi who was my very first mentor in the field of hearing science. After initiating my interest in the magic of hearing with his insightful guidance in my masters' pursuit in China, he followed up with my Ph.D. study and continued to provide caring suggestions.

I truly appreciate my parents, my sister and her husband, and my husband's family for their continuous encouragement, love, and support. My heartfelt thanks go to my son—Haochen, and daughter—Angelina. During the seven years, Haochen has constantly been a great help ever since he was a toddler. His self-discipline and considerateness made my dual role as a mother and a doctoral student much easier than it would have been otherwise. My life cannot be more enjoyable without his dynamic hobbies in science, sports, and arts. The birth of baby Angelina has adorned my joyful life with precious fun and laughter.

Finally, my wonderful husband—Wen is the one that I am most grateful to. Thank you for sharing every joy and bitterness with me for every second. Thank you for being extremely patient with my long-lasting study. Thank you for your endless understanding and support.

Abstract

This project studies auditory stream segregation as an underlying factor for poor speech perception skills in cochlear implant (CI) users by testing normal-hearing adults who listen to CI simulated sounds. Segregation ability was evaluated by behavioral responses to stimulus sequences consisting of two interleaved sets of noise bursts (A and B bursts). The two sets differed in physical attributes of the noise bursts including spectrum, or amplitude modulation (AM) rate, or both. The amount of the difference between the two sets of noise bursts was varied. Speech perception in noise was measured as the AM rate of the noise varied and at different spectral separations between noise and speech. Speech understanding and segregation ability are correlated statistically.

Results show the following: 1. Stream segregation ability increased with greater spectral separation, with no segregation seen when A and B bursts had the same spectrum or when they involved the most overlapping spectra. 2. Larger AM-rate separations were associated with stronger segregation abilities in general. 3. When A and B bursts were different in both spectrum and AM rate, larger AM-rate separations were associated with stronger stream segregation only for the condition that A and B bursts were most overlapping in spectrum. 4. Speech perception in noise decreased as the spectral overlapping of speech and noise increased. 5. Nevertheless, speech perception was not different as the AM rate of the noise varied. 6. Speech perception in both steady-state and modulated noise was found to be correlated with stream segregation ability based on both spectral separation and AM-rate separation.

The findings suggest that spectral separation is a primary/stronger cue for CI listeners to perform stream segregation. In addition, AM-rate separation could be a secondary/weaker cue to facilitate stream segregation. The spectral proximity of noise and speech has a strong effect on CI simulation listeners' speech perception in noise. Although neither the presence of noise modulation nor the modulation rate affected CI simulation listeners' speech understanding, the ability to use the AM-rate cue for segregation is correlated with their speech understanding. The results suggest that CI users could segregate different auditory streams if the spectral and modulation rate differences are large enough; and that their ability to use these cues for stream segregation may be a predictor of their speech perception in noise.

Table of Contents

Results and Discussion

List of Tables

List of Figures

Introduction

A. Auditory stream segregation in listeners with normal hearing

Auditory stream segregation (also referred to as auditory streaming) is an auditory process that occurs naturally in daily life. When listening to a talker at a party or when following a melody played by an instrument in an orchestra, listeners with normal hearing interpret the mixture of sounds in such a way that sounds from different sources are allocated to individual sound generators. Listeners can attend to the ongoing sounds from individual sources (streams) are perceptually concurrent.

1. Cues for auditory stream segregation

Auditory stream segregation in humans has been studied primarily in laboratory settings with non-speech sounds. In early laboratory studies, Van Noorden (1975) reported that frequency and temporal cues were critical for the formation of auditory stream segregation .

Van Noorden (1975) presented subjects with long sequences of tonal triplets ABA, where A stands for one tone with a variable frequency and B stands for another tone with a fixed frequency. The tone repetition time (i.e., the onset to onset time between the two adjacent tones, also referred to as stimulus onset asynchrony—SOA) was varied across conditions. The listener perceived either a galloping rhythm (integrated perception); or two segregated melodies (segregated perception), one with a pitch corresponding to the frequency of B tones and the other one with a pitch corresponding to the frequency of A tones. The subject was instructed to try to hold either the integrated or the separated perception in different conditions. When the intended perception was no longer heard as the frequency separation between A and B tones varied, the subject would respond. At that point the frequency separation and the corresponding tone repetition time for this perceptual breaking point were recorded. He found that the the frequency separation needed to be increased as the tone repetition time decreased for the listener to segregate the two streams. Interestingly, a listener could hold either the integrated perception or the segregated perception depending on the instructions for a certain range of frequency separation. Van Noorden observed two boundaries for these ranges—the fission boundary (FB) and the temporal coherent boundary (TCB) (Figure 1). When the frequency separation and the tone repetition time were in the area under the FB, a listener would integrate the A and B tones, even though he/she was instructed to hold a segregated perception. On the other hand, when the frequency separation and the tone repetition time fell in the area above the TCB, the listener would perceive segregation even though he/she was instructed to hold an integreted perception. When the frequency separation and tone repetition time fell in between FB and TCB, the listeners could hold either the integrated or the segregated perception as instructed.

The effect of the interation between frequency and tone repetition time on auditory stream segregation revealed by Van Noorden (1975) showed that when the frequency separation was smaller than approximately 3 semitones, the tones with different frequencies tended to be perceived as integrated regardless of the tone repetition time. It also demonstrated that shorter tone repetition time facilitated segregation remarkably when the frequency separation was greater than about 4 semitones.

Darwin and Carlyon (1995) have shown that other acoustical features of tonal stimuli could affect auditory stream segregation as well, such as pitch, onset aynchrony, location and timbre. They noted that the strongest cues to segregation are pitch and onset asynchrony.

Research in the recent literature supports the notion that listeners can segregate streams based on envelope cues. Grimault and his colleagues (2002) examined stream segregation base on amplitude modulation (AM) rate. In the study, they used broad band noise carrying AM to minimize the cue of frequency information. The repeated ABA pattern in Van Noorden's (1975) was adopted, except that A and B represent bursts that were amplitude modulated by two different rates. To ensure that sufficient AM cycles carried by each burst was available to the listener, the authors set the duration of each burst at 100 ms which was longer than that was used for unmodulated signals (50 ms in Van Noorden, 1975; 60 ms in Roberts, et al, 2002). In addition, they only studied relatively fast

AM rates—from 100 to 800 Hz. Their results showed that, for a burst repetition time (comparable to tone repetition time) of 120 ms and 100% modulation depth, the AM rate separation of 0.9-1 octave or greater could elicit stream segregation. This finding is potentially important for cochlear implant users who don't have a strong representation of pitch.

In our preliminary research (Nie and Nelson, 2007), we found that stream segregation could be elicited with modulation rates lower than 100 Hz for the same burst repetition time (i.e., 120 ms) as in Grimault et al study (2002), and a burst duration of 80 ms which was 20 ms shorter than what Grimault et al have used. The lowest modulation rate in Nie and Nelson's study was 25 Hz. The findings that auditory stream segregation can be based on AM rate and, that slow modulation rate as 25 Hz can be used in the experiment, are the basis of our present project.

In this experiment, the Roberts approach (Roberts, et al, 2002) was used (Details shown on page 10-12). Figure 2 shows the results of Nie and Nelson (2007). Based on findings of Grimault et al. (2002) it was hypothesized that stimuli with no modulation (shown in blue) or with modulation rates that differ by less than one octave (5075, shown in green) elicit the best thresholds. That is, when listeners do not segregate the stimulus streams using AM, the accumulated delay is easily detected.

It was hypothesized for those conditions where AM rates differed by one octave or more (5025 shown in red, and 50300 shown in pink), thresholds would be poorer because listeners segregate the A and B streams based on AM. For most listeners shown in panels A through D, this hypothesis was supported. Their preliminary results suggested that in some conditions AM could be used to form separate streams.

2. Mechanisms underlying auditory stream segregation

Bregman (1990) proposed two mechanisms—primitive and schema – for auditory stream segregation. The primitive mechnism refers to a stimulus driven process which does not require focused attention. In Van Noorden's (1975) figure (Figure 1.) showing TCB and FB, this mechanism underlies the perception corresponds to the areas above TCB and below FB. When the frequency separation and tone repetition time of the stimulus sequences fall into these two areas, either two segregated streams or one integrated stream is perceived respectively, even though the listener's attention is directed to the opposite perception. With respect to the primitive mechanism, the physical attributes of the stimulus are definite to elicit a segregated or integrated perception without the involvement of focused attention.

On the other hand, the schema mechanism refers to a top-down process with focused attention. The focused-attention-based perception corresponding to the area between the TCB and FB (Van Noorden, 1975) manifests this mechnism. When the frequency separation and the tone repetition time fall in this area, the listeners can form either a segregated or a integrated perception depending on which perception his/her attention is directed to. For a stimulus with ambiguous physical attributes, the listener's intention (or focused attention) can facilitate different perceptions. Consequently, in this condition, auditory stream segregation involves focused attention.

Both behavioral and neurophysiological studies have examined the role of attention in the formation of auditory stream segregation (Carlyon et al, 2001; Bregman and Rudnicky, 1975; Susmann et al, 1998, 1999, and 2002). Although there is a debate over whether the attention is a necessary factor for the primitive mechnism, the bulk of research in the literature has demonstrated that focused attention can modulate auditory stream segregation.

a. Measuring the effect of attention on stream segregation using behavioral approach:

Brochard, et al (1999) investigated the attention effect on auditory stream segregation using a behavioral approach. They presented listeners sequences of complex tones, a mixture of four or fewer subsequences (or streams) that differed in frequency. The tone onset to onset time was different across subsequences , but constant within subsequences . In one observation interval, the onsets of the initial tones in every subsequence were synchronized, and so were the onsets of the final tones in every subsequence . Different numbers of tones were presented for individual subsequences in one observation interval. If the listener could segregate the subsequences , he/she would have perceived different rhythms for each subsequence. Presented with a prime of the subsequence at the lowest frequency alone, the listener was cued to direct attention on this subsequence (referred by the authors as ‘focused stream’, as opposed ‘nonfocused stream’ for other subsequences). One of the tones in the focused subsequence (stream) was either advanced or delayed, which generated an irregular rhythm for this subsequence. The attention effort was evaluated by the threshold of the advance or delay for a listener to detect the irregular rhythm. Two findings in the Brochard, et al (1999) study are of particular interest to our project: One is that less attention effort (lower threshold in temporal jitter) was needed for stream segregation when only the focused subsequence was present compared with when both focused and unfocused subsequences were present. The other finding is that less attention effort was exerted when the frequency separation between the subsequences was larger. Both findings suggest that more attention effort is needed for stream segregation when the physical properties of a stream in a mixture of various streams are obscure.

Botte et al (1997) recorded attenuation effect on nonfocused streams in a multistream sequence in five out of eight subjects. They used a similar stimulation paradigm as that Brochard et al (1999) used. The subjects were presented with sequences mixed with three subsequences (streams of tones) differing in frequency and tempo (tone onset-to-onset time). Their attention was directed to one of the streams by a cued single stream preceding each stimulus sequence. The temporal irregularities were set in both the focused stream and one of the nonfocused streams. The researchers found that, for five out of eight subjects, the intensity of the nonfocused stream needed to be increased by 15 dB so that the detection of its temporal irregularity was equivalent to that of the focused stream. For all subjects, the detection of the temporal irregularity in the focused stream was ‘slightly decreased’ as the intensity of the nonfocused stream increased.

b. Measuring effects of attention on stream segregation using event-related brain potentials (ERP):

Supporting the observation that attention can modulate stream segregation, neurophysiological studies measuring event-related brain potentials (ERPs) have looked into the effect of attention on either a focused stream or nonfocused streams in a multistream mixture (Sussman et al, 1998 and 2005). In a series of studies, Sussman and her colleagues investigated the mismatch negtivity (MMN) as an index of the formation of auditory stream segregation. MMN is elicited by an oddball paradigm, in which one stimulus sequence is repeatedly presented (frequent stimulus, also referred to as the standard), another sequence occasionally replaces the standard (infrequent stimulus, also referred as the deviant). When the brain detects a standard-deviant change, it generates a negative wave component called mismatch negativity (MMN). MMN has been shown to reflect automatic detection of changes prior to a conscious judgment.

Sussman et al (1998) presented subjects with standards of reiterated sequences of six alternating high and low tones, L1-H1-L2-H2-L3-H3 (L1, L2, and L3 stand for three tones rising in frequency but within a low frequency range; H1, H2, and H3 stand for three tones rising in frequency but within a high frequency range). The high and low tones could potentially be perceived as two streams in high and low frequency ranges with a pattern of rising-pitch within each stream (L1…L2…L3 versus H1…H2…H3). Two deviants with a pattern of falling-pitch either in the low-frequency stream (L3-H1-L2-H2-L1-H3) or in the high-frequency stream (L1-H3-L2-H2-L3-H1) were presented to the subjects infrequently. The authors set the frequency separation between the two potential streams ambiguous for stream segregation. This was verified by the finding that no MMN was recorded for the deviants in either stream, when the subjects were reading a book and ignoring the acoustical stimulation. In contrast, when the subjects were instructed to focus attention on the high-frequency stream to identify the deviants for this stream, MMN was elicited by the deviants for the both low- and high-frequency streams. This finding suggests that for doublestream sequences with ambiguous physical properties to be segregated, focused attention may facilitate the formation of segregation.

In the Sussman et al (2005) study, tones of three potentially-perceived streams were interleaved in a stimulus sequence. The frequency separations between the streams were set to be unambiguous for segregation and recordable MMNs were elicited by the deviants for all the three streams, when subjects were ignoring the acoustical stimulation and focusing attention on reading a book. When the subjects' attention was directed to one of the three streams by identifying the deviants in the corresponding stream, the MMN was only elicited by deviants in the focused stream and no MMN was elicited by the deviants in the nonfocused streams. Their finding shows that focused attention can suppress the formation of nonfocused streams while maintaining the focused stream perception.

Overall it appears that attention can modulate stream segregation and that the experimental paradigm can affect the outcome of segregation studies.

3. Research paradigms to study auditory stream segregation.

One of the limitations of the early behavioral approach in studying stream segregation is that it relies on a listener's report of an overt perception thereby presenting difficulties in controlling subjects, who may not apply the same perceptual criteria for segregation. To overcome this limitation, Roberts et al

(2002) developed a paradigm (Figure 3) to test the primitive process for stream segregation without measuring subjects' overt perception. They used 12 cycles of alternating AB tones which differed in frequency. The duration of A and B tones were both 60 ms. Listeners were asked to compare two tone sequences, one with a regular rhythm and one with an irregular rhythm. The difference between these two sequences resided in the stimulus onset asynchrony (SOA), which is equivalent to the duration between the onsets of an A tone and a B tone in the same cycle. The sequence with the regular rhythm included a constant SOA of 100 ms. The sequence with the irregular rhythm applied different SOAs for its three portions: the first portion consisted of 6 cycles of AB tones which carried an SOA of 100 ms as the constant one used in the regularly rhythmic sequence; in the second portion (the four middle cycles from the 7th to the 10th cycles), the SOA (i.e., the duration between the onset of a B tone and the onset of the A tone immediately preceding the B tone) was progressively delayed; the third portion was composed of the last two cycles of AB tones, in which the delay of a B tone relative to the preceding A tone was maintained from the second portion. In the second and third portions, despite the delay of B tones relative to the A tones in the same cycles (i.e., B tones were delayed relative to the preceding B tones), the gap between two successive A tones remained the same as that in the regularly rhythmic sequence. The entire sequence lasted 2.4 seconds, with the first 6 cycles

(1.2 seconds) designed for the build-up time of stream segregation (Bregman, 1978).

The listeners’ task was to determine which one of the two intervals contained the sequence with irregular rhythm. The delay of B tones was adaptively decreased to determine a threshold reflecting the ability to segregate streams. The listener could potentially detect the irregular rhythm when either integrating A and B tones into one stream or segregating them into two streams. With an integrated perception, the listener would have compared the gaps between the B tone and either the preceding or following A tone. With a segregated perception, the listener would have compared the gaps between the adjacent B tones. Since the gaps between A and B tones were shorter than that between adjacent B tones, the delay of B tones would be equavalent to a larger proportional change relative to the gap between adjacent A and B tones than relative to the gap between adjacent B tones.Therefore, it would have required less effort for the listener to detect the same delay of B tones when he/she integrated the tones. Thus this stimulus paradigm requires subjects to focus attention effort on an integrated perception to achieve a better threshold and a low threshold in the delay of B tones implies poor segregation abilities.

An alternative stimulus paradigm that favors segregation (segregation-driven) has been proposed by Micheyl (personal communication). He adopted the pattern of repeated A-B-A triplets in a sequence (e.g., A-B-A-A-B-A-A-B-A…), A and B being two tones differing in frequency. He jittered the temporal placement of A tones and kept B tones constant in their nominal temporal positions. If segregation was generated, the listener would perceive B stream with a constant gap between the two adjacent B tones and A stream with varied gaps between the two adjacent A tones. If integration was generated, the listener would perceive one stream fluctuating in a high-low-high pitch with segments unevenly spaced in tempo. In this paradigm, Micheyl either delayed or advanced the temporal position of the last B tone yet still between the two A tones of the last triplet and the listeners were supposed to determine which direction this last B tone was arranged. This task requires the listener to focuse attention on the B stream. With an integrated perception, it would be considerably challenging for the listener to detect the changing position of the last B tone relative to its adjacent A tones, due to the fact that the jitter of A tones generated various A to B or B to A durations within each of the previous triplets. The approach proposed in this study was inspired by Micheyl's design.

4. Auditory stream segregation and speech perception in acoustic hearing

Various acoustical cues could generate auditory stream segregation with non-speech stimuli. To our knowledge, only one group of researchers (Mackersie, Prida, and Stiles, 2001) has reported the correlation between speech perception skills and stream segregation ability with tonal stimuli in hearing-impaired listeners. The authors used a traditional stimulus paradigm which encompassed sequences of repeated triplets ABA as was reported by van Noordon (1975). The acoustical cue for stream segregation investigated in this study was frequency difference between A and B tones. The stimulus sequences started at a larger frequency separation presented to a listener. Two groups of listeners were involved, normal-hearing (NH) listeners and hearing impaired (HI) listeners. The listeners indicated hearing one or two streams by pressing two different keys on the keyboard of a computer. In the initial trials, the listener could hear two streams as the frequency separation was set sufficiently large to elicit the perception of segregation. The frequency separation between A and B tones gradually decreased until the listener reported hearing one stream (intergating A and B tones). The frequency separation at which the listener changed from the perception of two streams to that of one stream was named “fusion threshold” by the authors. This fusion threshold corresponds to the fission boundary (FB) in van Noorden (1975) and was expressed in semitones. A smaller fusion threshold is associated with a stronger segregation ability. The HI listeners showed a significant larger fusion threshold than that of the NH listeners, implying a degraded stream segregation ability based on frequency difference.

In the speech perception study, the listeners were simultanously presented pairs of sentences, one spoken by a female talker and the other one spoken by a male talker. The listeners repeated both sentences. The percent correct of words was recorded for each sentence. The words correct in the sentence first repeated was found decreasing as the fusion threshold increased in HI listeners. The Pearson product-moment correlation coeficient for this correlaiton was -0.85. This implies that stream segregation ability may be an appropriate predictor for speech perception skills in hearing impaired listeners.

B. Auditory stream segregation and cochlear implants

1. Implant listeners' difficulties in segregating speech from background noise

Cochlear implant listeners rely on the electrical signals encoding information in the acoustical signals to stimulate the auditory nerve to form an auditory perception. All the incoming sounds are processed through the processor according to some programmed rules.

Cochlear implant listeners' spectral resolution has been shown to be degraded (Fu, et al, 1998; Friesen, et al., 2001;). Fu and his colleagues (1998) studied vowel and consonant recognition in noise on four simulation listeners and three CI listeners with variable numbers of channels. They found that, for a given SNR, both simulation listeners and real users' performance deteriorated as the number of channels decreased. To reach subjects' maximum performance, more channels were needed in noise conditions than in quiet.

Friesen, et al (2001) quantified the effect of number of spectral channels on speech recognition in noise by investigating acoustical simulations and more cochlear implants users with more electrode/spectral band conditions than those used in Fu, et al (1998). They measured the recognition scores for vowels, consonants, CNC words, and HINT (Hearing In Noise Test) sentences in both quiet and noise conditions. Consistent with Fu et al (1998) findings, as the number of electrodes/channels increased (up to 7/8 electrodes for Nucleus 22 and 10 electrodes for Clarion), the subjects' speech recognition scores in noise improved and then reached the plateau. For simulation listeners, their speech recognition kept improving as the number of channels increased to 20. Shannon, et al. (1995) reported that in the quiet listening conditions, good vowel recognition was achieved with only four spectral bands of acoustic simulations. The increment of speech recognition in noise with the enhancing numbers of channels/electrodes for both implant users and acoustically simulated subjects (Friesen, et al., 2001) demonstrated that more spectral cues are required in comprehending speech in noise than in quiet situations.

In the quiet environment, good CI listeners' speech recognition performance can approach a perfect score (Friesen, et al., 2001). When a speech signal is presented in competing speech, CI users cannot benefit from the masking release as normal hearing listeners can (Stickney, et al., 2004). The authors studied the effect of different types of background noise on CI listeners' speech perception. They examined steady-state noise with a long-term speech spectrum, competing speech spoken by the opposite gender, and competing speech spoken by the same gender. The result showed that CI listeners' speech perception in these noise conditions decreased significantly; in contrast, the listeners with normal hearing demonstrated greatest difficulty in steady-state noise when listening to natural speech.

Qin and Oxenham (2003) reported consistent findings from CI simulation listeners. In this study, three simulated CI channel conditions (4, 8, and 24 channels) and natural speech were presented to subjects with normal hearing. Speech recognition was measured in four background conditions (speech-shaped noise, amplitude-modulated speech-shaped noise, single male talker, and single female talker). The results showed that, for simulation listeners, the background noise with real talkers had more detrimental effect on speech recognition than the speech-shaped noise; so did the amplitude-modulated noise in both 4- and 8-channel conditions. Nevertheless, with natural speech, the subjects achieved best performance in the modulated speech-shaped noise.

Nelson and Jin published their series of study on CI listeners' speech understanding in different background noises (Nelson and Jin, 2004; Jin and Nelson, 2006). In this series, they tested subjects with normal hearing and CI users in both steady-state noise and gated noise. The speech recognition of normal hearing listeners was also studied when they listened to CI simulations. As opposed to the expected masking release (that is, better performance in gated noise than in the steady-state noise for normal-hearing listeners), CI users and simulation listeners scored equally poorly in both gated and steady noise. Moreover, while the normal hearing listeners maintained a high sentence understanding score in the gated noise even when the duty cycle of the gated noise was lengthened or its rate was increased to a relatively high degree, the CI listeners' scores dropped remarkably to around 20% as long as the gated noise was applied regardless of the its duty cycle or rate.

From the stream segregation view, the above studies suggest that CI listeners have significant difficulty in integrating the multiple snapshots of a sentence spoken by the one talker into a holistic meaningful item, and integrating the snapshots of another sentence spoken by a different talker (or on-and-off noise) into another item. This implies that CI listeners have difficulties in segregating the ongoing sounds into different streams.

2. Auditory stream segregation in cochlear implant users and its potential correlation with users' speech understanding skills.

Cochlear implants extract the temporal envelopes in incoming sound waves and impose them upon the electrical pulses CIs generate themselves. In consequence, the temporal envelope is a crucial cue for CI users to perceive sounds. If CI users could segregate streams based on the temporal envelope as normal-hearing listeners do, their poor speech understanding performance in noise may be partially accounted for by their segregation ability.

Four studies (Chatterjee and Galvin, 2002; Hong and Turner, 2006; Chatterjee, et al, 2006; Cooper and Roberts, 2007) have been published demonstrating that, at least, some implant listeners can do stream segregation. However, the four studies revealed great discrepencies about whether CI users are able to segregate auditory streams: The studies of Chatterjee and Galvin (2002) and Hong and Turner (2006) supported that CI users can form stream segregation but with reduced capability; whereas, Cooper and Roberts (2007) argued for the opposite view. Even within the same study, CI users' stream segregation skills varied considerably: Half of Hong and Turner's subjects showed skills comparable to the normal hearing group while half of them showed extremely decreased skills; in Chatterjee et al (2006), only one out of five of subjects demonstrated clear stream segregation, although the results of the other four subjects were uncertain.

Methodology could be one of the issues responsible for the inconclusive results. In both studies by Chatterjee and her colleagues, as well as Cooper and Roberts (2007), the stream segregation was assessed by subjects' report of their subjective perception, which involves an uncontrollable definition or degree of segregation across subjects. Hong and Turner (2006) adopted an objective method (Roberts et al, 2002). However, this approach requires that the listener try to integrate rather than segregate the two tones in one sound sequence in order to differentiate the two sequences. As the dissimilarity of sounds through CIs is decreased, CI users supposedly tend not to segregate sounds. A task that requires them to apply mental effort toward integration rather than segregation cannot disclose their true ability to segregate streams. An objective approach favoring stream segregation is needed.

The enormous heterogeneoty in CI listeners, such as etiology, the duration of CI usage, rehabilitation history, etc, may have resulted in the inconsistency in the mentioned studies. A study on normal listeners listening to simulated CI processed stimuli may help lay basis for further investigations on real users.

Very recently, Hong and Turner (2009) have published results of a study using a stimulus paradigm in favor of stream segregation. In this study, a three-interval two-alternative procedure was used. The listeners were presented, in each of the last two intervals, a stimulus sequence consisting of broadband noise bursts carrying AM at two different modulation rates, denoted respectively A bursts corresponding to one AM rate and B bursts corresponding to the other. A target temporal pattern composed of four A bursts was embedded in one of the sequences. The listeners were primed with the target temporal pattern in the first interval and were required to choose the sequence containing the target pattern. The strength of stream segregation was measured in two ways, including the threshold of AM-rate separation between A and B bursts and of modulation depth for stream segregation. With respect to the former, the AM rate of A bursts was fixed at 80 Hz while the AM rate of B bursts was adaptive to track the threshold in AM-rate separation between A and B bursts for stream segregation. With the latter, the AM rate of A bursts were fixed at either 80 Hz, or 200 Hz, or 300 Hz, while B bursts were unmodulated. The modulation depth of A bursts was adaptive to track the threshold in AM depth for stream segregation at different AM rates. Their normal-hearing listeners and CI users showed comparable abilities to segregate the A and B streams based on AM-rate difference only when the AM-rate difference between A and B streams was sufficiently large and when the listeners focused attention was direction to stream segregation. Both listener groups also demonstrated marked variability in this segregation ability—four out of the twelve normal-hearing listeners and two out of the ten CI users either did much poorer than the other listeners in the same group or were unable to segregate the two streams. This suggests that even normal-hearing listeners may have a wide range of abilities to use the cue of AM-rate as a solo cue to segregate auditory streams. It posited a further question—with the addition of reduced spectral cue to the AM-rate difference for normal-hearing listeners—which is equivalent to the situation of CI simulations, how would the these two cues interact to affect the segregation ability?

Due to the lack of research on the correlation between speech perception skills and stream segregation ability for CI listeners, research in this area is needed. The correlation between stream segregation and speech perception skills in noise was only explored in one study (Hong and Turner) with inconclusive findings.

Hong and Turner (2006) correlated CI subjects' speech recognition threshold (SRT) in noise with their ability to segregate streams. They presented CI users with pure tone sequences. The ability of segregation for different frequency range was tested by selecting three base tones—200, 800, and 2000 Hz for A tone. The B tone was varied systematically by a fraction of an octave from the A tone of a particular base frequency. They found a statistically significant correlation between SRT and segregation ability for base tones of 800 and 2000 Hz, although no significance for the base tone of 200 Hz. Chatterjee et al (2006) stated that the only subject that showed definite stream segregation was an experienced user and demonstrated high speech perception skills in their previous study. A study of the correlation on CI simulation would shed a light in this area and lay ground for further research on CI users. If AM is an effective cue for stream segregation in CI and simulation listeners, future improvements implant signal-processing algorithms could incorporate AM to attempt to improve speech recognition in background noise.

The current study was conducted in order to address the following research questions:

1. When listening to simulated CI-processed sounds, can normal listeners two different streams of stimuli based on amplitude modulation rate and/or spectral difference?

2. Does segregation ability correlate with speech perception skills through CI simulations?

PARTICIPANTS

10 undergraduate and graduate students, 5 male and 5 female, participated in the study. They were 19 to 32 years of age and native American English speakers. Their hearing was no greater than 20 dB HL at audiometric frequencies of 250, 500, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz.

METHODS

I. Psychophysical study

Experiment 1 (auditory stream segregation):

Stimulus paradigm:

Twelve repeated pairs of A and B noise bursts, where A and B bursts were either broadband noise or vocoder bandpass noise carrying sinusoidal AM (100% modulation depth). They differed either in the center frequency of the noise band or the AM rate, or both.

The duration of an A or B burst was 80 milliseconds (ms) including 8-ms rise/fall ramps. The interval between the onsets of two consecutive bursts (i.e., the onsets of an A burst and its adjacent B bursts, or vice versa), namely burst repetition time (BRT), was 130 ms, while A bursts (excluding the initial one) were jittered ±40 ms from their nominated temporal locations.

Gaussian noise was used for the broad band noise (BBN) with a sampling rate of 22050 Hz and delivered through a TDH49 headphone. For the vocoder bandpass noises, the Gaussian noise with the same spectrum as that of the BBN was filtered into bandpass noises by the following means. The cutoff frequencies were adopted from Fu and Nogaki (2004). Table 1 shows the cutoff frequencies with a resolution of 16 bands. The bands were numbered from one to sixteen corresponding to bands with center frequencies from low to high. The lowest eight bands (bands 1through 8) were combined into one bandpass noise which was used for B bursts thus referred to as B band. The higher six bands (e.g., bands 11 to 16) were combined into another bandpass noise which was presented as A bursts thus referred to as A band. While the spectrum of the B band was constant (i.e., encompassing the lowest 8 vocoder bands), the spectra of the A bands covered three conditions, in terms of their relationship with the spectrum of B band:

1. no band overlapping, A11-16B1-8 (A band consists of vocoder bands 11 to 16; B band always consists of bands 1 through 8);

2. moderately band overlapping, A7-12B1-8 (A band consists of vocoder bands 7 to 12): 16.5% overlap in frequency—calculated as (high cutoff frequency of B band – low cutoff frequency of A band)/(high cutoff frequency of A band – low cutoff frequency of B band)—i.e. (1426-931)/(3205-200);

3. greater band overlapping, A5-10B1-8 (A band consists of vocoder bands 5 to 10): 42.8% overlap in frequency—calculated as (1426-591)/(2149-200).

4. completely band overlapping condition, AbbnBbbn (both A and B bands consist of broad band noise).

Four comparisons of AM rates were applied to A and B bands described as followed.

1. 0 Hz vs 0 Hz (AM0-0): no AM applied to either A band or B band;

2. No separation of modulation rate (AM25-25): 25 Hz vs 25 Hz—both A and B bands were modulated at a rate of 25 Hz;

3. modulation rates 2 octaves apart (AM25-100): 25 Hz vs 100 Hz —A and B bands were modulated at rates of 25 Hz and 100 Hz respectively;

4. modulation rates 3.58 octaves apart (AM25-300): 25 Hz vs 300 Hz—A and B bands were modulated at rates of 25 Hz and 300 Hz respectively.

These conditions were selected based on the results of Nie and Nelson (2007).

Procedure:

Measurement: d’s were measured through a single interval yes/no approach. In each interval, the last B burst of the stimulus sequence was either delayed from its nominated temporal position by 30 ms (signal sequence) or advanced by a varied period ranging from 0 to 10 ms (no-signal sequence). The total duration was 3.1 seconds for the delayed stimulus sequences and 3.06-3.07 seconds for the non-delayed sequences.

The stimulus sequences were presented monaurally to the right ear. The task was to determine which interval contained the delayed B burst and which interval contained the advanced B burst. The subjects were given the two response options in two graphic boxes on a computer screen, one showing “1 Longer” for the “delayed” option and the other one showing “2 Shorter” for the “non-delayed” option. The subjects pressed on the keyboard number 1 (for the “delayed” choice) and number 2 (for the “advanced” choice). Feedback was provided following each response by flashing the box corresponding to the correct answer on the screen. Subjects were allowed to take as much time as they needed to make the selection for each trial.

The rationale for the task is as follows: To detect the delayed last B bursts, listeners had to discriminate the prolonged gap between the last two B bursts as opposed to the constant gaps between the previous adjacent B bursts. Hence, Listeners had to follow B bursts and ignore A bursts in order to determine the gaps between B bursts. In other words, listeners needed to segregate A bursts from B bursts and make mental efforts to generate a perceptual stream of B bursts. The design of the stimulus paradigm with a steady tempo of B bursts and a jittered occurrence of A bursts was to facilitate listeners' segregation of a stimulus sequence interleaved with A and B bursts into two streams: The temporally predictable occurrence of B bursts provided a pattern for listeners to follow while the temporally unpredictable occurrence of A bursts introduced difficulties for integrating A and B bursts constantly. The better a listener could segregate A and B bursts, the stronger the perception of the stream of B bursts. As a consequence, the listener would perform better at identifying the stimulus sequences with the last B burst delayed.

On the other hand, one could potentially detect the delayed sequences without stream segregation by simply comparing the A-B gaps of the last a few pairs of A and B bursts. In this way, listeners didn't have to segregate streams of A and B bursts; instead, they could have integrated the last couples of A and B bursts. In addition, theoretically, a listener could compare the last A-B gap in one trial against that in another trial to identify the delayed sequence. In other words, the listener is performing a gap discrimination task in this situation.

Although the ability to identify the delayed sequences based on the integrated perception or gap discrimination is not expected to be different for sequences containing three pairs of A-B bursts or sequences containing 12 pairs of A-B bursts, performance based on stream segregation is presumably better with sequences containing 12 pairs as opposed to sequences containing 3 pairs. Bregman (1978, 1990) reported that auditory stream segregation needed approximately 1 second to build up. In the current experiment, a stimulus sequence with 3 pairs lasted 0.73-0.77 seconds which was supposedly not long enough to generate perceptual segregation. Therefore, listeners would perform worse with this short version of the stimulus sequences than the longer version, if they relied on stream segregation. Consequently, this short version (3 pairs of stimulus sequences) was run as a control condition on all subjects with the greater-band-overlapping condition (A5-10B1-8).

One other potential cue that listeners could use is to generate an overall range of A-B or B-A gaps and compare the last A-B gap against this overall range. If A bursts had not been jittered, the gaps between the adjacent A and B bursts would have been fixed at 50 ms. With A bursts being jittered by ±40 ms from their nominated positions, the gaps between the adjacent A and B bursts ranged from 10 ms to 90 ms effectively. While listening to a stimulus sequence, a listener could possibly detect this gap range by averaging across all the gaps throughout the sequence. When the last B burst was delayed by 30 ms, the gap between the last A burst and the last B burst ranged from 40 ms to 120 ms. For those delayed stimulus sequences including a gap larger than 90ms between the last A and B bursts, there was a chance that listeners detected that this gap was beyond the gap range.

If this were the case, listeners would perform better with 12-pair sequences than with 3-pair sequences when other experimental parameters were comparable between the two conditions. This possibility would contradict the hypothesis brought forward previously, that better performance with 12-pair-burst sequences is associated with stream segregation. To rule out this confounding factor, a control condition with 3-pair-burst sequences was run on five participants (randomly selected) in the completely-overlapping band condition (Abbn-Bbbn) as the A and B bursts were carrying the AM at the same rates (AM0-0 and AM25-25). In this control condition, the A and B bursts in a sequence were literally the same—they were both broadband noise and, either unmodulated, or amplitude modulated at the same rate of 25 Hz. It is presumed that no auditory stream segregation would be elicited by these sequences composed with same elements. If the listeners had contrasted the intra-pair gap with the averaged A-B gap, they would have performed worse for the 3-pair-burst sequences than for the 12-pair-burst sequences. Two AM rate conditions (AM25-100 and AM25-300) were also included when testing participants to compare with the performance from their corresponding conditions with 12-pair-burst sequences. Based on our preliminary data, we had hypothesized that AM rate separation alone would not be a sufficient cue to elicit segregation in our stimulus paradigm. Consequently, the ability to identify the delayed stimulus sequences would not show differences between the 12-pair-burst and 3-pair-burst sequences when both A and B burst were broad band noise.

Four blocks of 70 trials were run for each condition with 50% chance of occurrence for either the signal sequence or the non-signal sequence. The first 10 trials were designed to facilitate the subject to form and maintain stream segregation. The hit rate and false alarm rate were calculated from the last 60 trials. The individual hit rates and false alarm rates from the four blocks for the same condition were averaged and the d’ for that condition was computed from the averaged hit rate and false alarm rate.

It should be noted that the standard approach of obtaining four individual d’s from each of the four blocks in a certain condition and then computing a d’ average across all the four blocks for that condition, was not used. The reason of not deriving a d’ average was: Pilot data had shown a substantial variability across listeners regarding the hit and false alarm rates for the current task. With the same delay of the last B burst (30 ms), following 10-20 hours listening experience, some listeners had reached ceiling performance in some blocks of relatively easy conditions while some other listeners only performed close to chance level in some moderately difficult conditions. To avoid causing listeners with poor ability to perform at chance level in every condition, the selection of stimulus parameters had to be set to a level that resulted in ceiling performance for some listeners in a few runs. When the ceiling performance was reached (i.e., 100% for hit rate or 0% for false alarm rate), d’ was unable to be derived for that block. As a consequence, a d’ was derived from hit rate average and false alarm average across 4 blocks for each condition.

It is noteworthy that participant S1 had participated pilot study and had had approximately 40 hours experience of similar tasks prior to undertaking the current auditory stream segregation task. Her hit rates of all the 4 blocks in one condition (no-band-overlapping—A11-16B1-8 and no AM—0-0Hz) reached 100%. Data were recollected from another 4 blocks of the same condition on a different day, and an average was taken across all the 8 blocks for both hit rate and false alarm rate.

Summary of Conditions:

Experimental conditions—stimulus sequences with 12 pairs of A-B bursts:

Bands (4): 1) completely-band-overlapping (broadband noise): AbbnBbbn;
2) greater-band-overlapping : A5-10B1-8;
3) moderate-band-overlapping: A7-12B1-8;
4) no-band-overlapping: A11-16B1-8.

AM rates (4): 1)AM0-0: 0 vs 0 Hz (no AM); 2) AM25-25: 25 vs 25 Hz; 3) AM25-100: 25 vs 100 Hz; 4) AM25-300: 25 vs 300 Hz.

The total number of conditions was 16.

<u>Control conditions—stimulus sequences with 3 pairs of A-B bursts (short version of stimulus sequences):</u>

a) Undertaken by all participants:

   Bands: S-A5-10B1-8.

   AM rates: 1) AM0-0; 2) AM25-25; 3) AM25-100; 4) AM25-300.

The total number of conditions was four.

b) Undertaken by 5 participants:

   Band: S-AbbnBbbn

   AM rates: 1) AM0-0; 2) AM25-25.

The total number of conditions was two.

The experimental conditions and control conditions were all included in a randomization. Each of the six possible band conditions was presented to participants in a randomized order (note: an experimental condition and a control condition with the same band condition were considered two band conditions). Four AM rate conditions were nested and randomized under each band condition. All sessions were approximately 1.5 hours. Participants had one training session (see the section of Familiarization for detail). Experimental conditions were undertaken in eight sessions on different days, one or two sessions on each day. All participants completed half of an experimental band condition (i.e., two AM rate conditions nested under that band condition) in each session. In addition, all participants completed one of the control band conditions (S-A5-10B1-8) (with 4 AM rate conditions) in one other session; five participants completed another control band condition (S-AbbnBbbn) (with 4 AM rate conditions) in another additional session). Participants were encouraged to take a 5-minute break after 2 or 3 blocks.

Familiarization:

Training session: The first 1.5-hour session was designed for the purpose of training. The structure of stimulus sequences was described to the participants verbally and with a schematic demonstration. Participants performed the same task as in the experiment—determining whether the last B burst in a sequence was delayed or advanced. The experimental conditions were presented to the listeners in this training session. Blocks of 40 trials were presented.

Participants were initially presented with the presumably easiest condition—A11-16B1-8/AM25-300. They were asked to follow the low-pitched B bursts and ignore the high-pitched A bursts. All participants reported perceiving segregated streams in this block. Additional blocks of the same condition were undertaken until a participant’s d’ was larger than 2.

Then, the conditions presented to the participants were sequentially decreased in spectral separation. The AM rate separation was either AM25-300 or AM0-0 for band conditions of A11-16B1-8 or A5-10B1-8, as well as AM25-300 for the band condition of AbbnBbbn. Blocks of one band condition were run until the participant reported perceptual segregation; the next band condition with a smaller spectral separation was then presented.

All participants reported consistent segregation perception throughout at least one block in the conditions of no-band-overlapping, moderate-band-overlapping, and greater-band-overlapping. However, they reported difficulties in holding the segregation perception for the complete-band-overlapping condition (i.e., AbbnBbbn) with only AM separation (AM25-300). In consequence, participants needed 45-60 minutes for this condition to repeat at least 8 blocks; whereas, they took 30-45 minutes for other band conditions in each of which 1-5 blocks were needed to reach a consistent segregation perception.

Experimental sessions: Prior to data collection, participants practiced the task with two 40-trial blocks, one for the no-band-overlapping condition and one for the completely-band-overlapping condition with the AM rate condition of AM 25-300. All participants reported the capability of holding segregation perception throughout the block of no-band-overlapping condition. More blocks were presented if participants reported absolutely no perception of segregation for the completely-band-overlapping condition until they reported intermittent segregation perception.

In the experimental blocks, the first 10 trials were designed for participants to build and reinforce a segregation perception for that condition. The occurrence of the signal and non-signal sequences were both 50% chance in these 10 trials. The participants' responses to these 10 trials were not included in the computation of d's.

Experiment 2 (auditory gap discrimination)

Stimulus:

Stimulus sequences consisted 12 broad band noise bursts which were temporally separated by silent gaps. The duration of each burst was 80 ms including 8-ms sinusoidal rising and falling ramps. The SOA was 260 ms. These stimulus sequences were equivalent to the sub-sequence of B bursts of broad band noise, in the stimulus sequences in Experiment 1.

Two types of sequences were included in the study: 'no-delay' and 'delayed' based on the length of the last SOA: No-delay sequences refer to those with the last SOA same as others (260 ms); whereas the delayed sequences refer to those with the last SOA longer than 260 ms which rendered a delayed onset of the last noise burst. The gap was quantified as the amount of time that the last SOA was elongated from 260 ms.

Procedure:

A two-alternative forced-choice (2AFC) paradigm was employed. A delayed sequence and a no-delay sequence were randomly placed in either of the listening intervals in each trial. The inter-stimulus-interval (ISI) was 500 ms. Thresholds were estimated using a two-down, one-up adaptive track converging on the 70.7% correct point on the psychometric function. The starting gap duration was 30 ms which made the last SOA in the delayed stimulus sequences 290 ms. Initially, gap durations were adjusted by a factor of 1.414. This was reduced to 1.189 after the second reversal. The track continued until ten reversals had been obtained, and the geometric mean of the gap duration at the last eight reversals was taken as the threshold estimate.

The stimulus sequences were presented monaurally to the right ear. The presentation level was 70 dB A for each 80-ms noise burst. The total duration of a standard stimulus sequence comprised of twelve noise bursts is 3.12 seconds. The total duration when the noise bursts on was 0.96 second while the rest duration of the sequence is silence. This resulted in an overall level of 65 dB A of the stimulus sequence. The listening intervals were visually cued by lighting graphical boxes on a computer screen. The task was to determine which interval contained the delayed sequence and press either number 1 or 2 on the computer keyboard to indicate the response. Feedback of correct interval was provided following each response by flashing the box corresponding to the correct answer on the screen. Participants were allowed to take as much time as they needed to make the selection for each trial.

Six estimates of gap threshold were measured for each participant. The participant's final threshold was taken as the geometric mean of all the six estimates collected for that condition.

Experiment 3 (auditory amplitude modulation detection)

Stimuli:

Sinusoidal amplitude modulation (SAM) was imposed on an 80-ms independent white noise including 8-ms rise/fall ramps. The three SAM rates used in Experiment 1 were examined: 25 Hz, 100 Hz, and 300 Hz. The stimulus waveform was defined by

$$A(t) = [1 + m \sin(2\pi f_m t)]\, N(t), \qquad (2)$$

where t is time, m is the modulation index, f_m is the modulation rate, N(t) is the noise carrier.

Procedure:

A 2-AFC paradigm was adopted. A stimulus without SAM and a stimulus with SAM were randomly placed in either of the listening intervals in each trial. The inter-stimulus-interval (ISI) was 500 ms. Thresholds were estimated using a two-down, one-up adaptive track converging on the 70.7% correct point on the psychometric function. The starting modulation index was 1 corresponding to 0 dB modulation. Initially, modulation depths were adjusted by a factor of 4 dB.

This was reduced to 2 dB after the second reversal. The track continued until ten reversals had been obtained, and the mean of the modulation depth at the last eight reversals was taken as the threshold estimate.

The stimuli were presented monaurally to the right ear. The presentation level was 70 dB A for both unmodulated noise and modulation noise. The listening intervals were visually cued by lighting graphical boxes on a computer screen. The task was to determine which interval contained the modulated noise and press either numbers 1 or 2 on the computer keyboard to indicate responses. Correct interval feedback was provided following each response by flashing the box corresponding to the correct answer on the screen. Participants were allowed to take as much time as they needed to make the selection for each trial.

For each participant and each of the three modulation rates, at least four estimates of a modulation threshold were obtained. Two additional estimates were measured and averaged with the previous 4 estimates, if the standard deviation across the first 4 estimates was larger than 3 dB. In total, 30 modulation thresholds were obtained with three for each of the ten participants. Two out of the 30 thresholds had a standard deviation larger than 3 dB, but smaller than 4 dB.

II. Speech perception study

Stimuli:

IEEE sentences and white noise were filtered through 16 filter bands, whose cutoff frequencies were adopted from Fu and Nogaki (2004). Bandpass butterworth filters of 4th order were used. The temporal envelopes of speech were extracted from individual lower 10 bands by low-passing the bands of speech through a 4th order Butterworth filter at a cutoff frequency of 160 Hz. The speech envelopes from each band were half-wave rectified, and imposed on the white noise filtered through the same band. Each noise band carrying speech temporal envelopes was re-filtered through the same bandpass filter to eliminate the spectral spread resulting from imposing envelopes on noise. The 10 bands of noise carrying speech envelope were added to formulate the speech stimuli.

The background noise was the bandpass noise carrying sinusoid amplitude modulations. The spectra of the background noise were set to completely overlapping, partially overlapping, and not overlapping with that of the speech. For the completely overlapping condition, the individual noise bands 5 through 10 were combined into one noise band with the lower cutoff frequency same as that of band 5 and the upper cutoff frequency same as that of band 10 (N5-10 condition); for partially overlapping conditions, the individual noise bands 7 through 12, 9 through 14, and 11-16 were combined into single noise bands (N7-12, N9-14 and N11-16 conditions); for the not overlapping condition, the individual noise bands 13-16 were combined (N13-16 condition).

Five modulation rate (N-AM) conditions were studied—0 Hz (i.e., steady noise), 4 Hz, 16 Hz, 32 Hz, and 64 Hz.

<u>Summary of Conditions:</u>

Band pass noise conditions (5):

1) completely overlapping: Noise bands 5 to 10 and speech bands 1-10 (N5-10Spch1-10)

2) greater overlapping: Noise bands from 7 to12 and speech bands from 1 to 10 (N7-12Spch1-10);

3) moderately overlapping: Noise bands from 9 to 14 and speech bands from 1 to 10 (N9-14Spch1-10);

4) slightly overlapping: Noise bands from 11 to 16 and speech bands from 1 to 10 (N11-16-Spch1:10);

5) no overlapping: Noise bands from 13 to 16 and speech bands from 1 to 10 (N13-16Spch1-10).

Noise AM rate conditions (5): 1) 0 Hz (N-AM0), 2) 4 Hz (N-AM4), 3) 16 Hz (N-AM16), 4) 32 Hz (N-AM32), and 5) 64 Hz (N-AM64).

The total number of conditions was 26: 5 (bandpass noise conditions)* 5 (noise AM rate conditions) + 1 (quiet condition)..

<u>Procedure:</u>

The speech stimuli were presented at 70 dB A monaurally to the right ear when no background noise was present; when background noise was present, the speech stimuli were presented at 70 dB A and the noise was at a signal to noise ratio (SNR) of 10 dB.

In each condition, two randomly chosen lists of 10 sentences were presented to the participants. Participants wrote down what they heard after presenting each sentence and pressed the "Enter" key on the computer keyboard to listen to the next sentence. Each participant was scored for the percent of keywords correctly recorded on the response sheet. The total number of keywords from two IEEE lists was 100. If participants responded in homophones of keywords, they were scored correct for these corresponding keywords. No correct feedback was provided.

Conditions were randomized across the ten participants and the order of the ten sentences within each list was randomized. To complete the speech experiment, participants took four to six hours in total, which was broken into one to two hour listening sessions. Participants undertook the sessions on different days.

Practice listening was given at the beginning of the first session of the speech experiment. Participants listened to speech stimuli from two randomly chosen IEEE lists. One of the two practice lists was presented without background noise (Quiet condition) and the other one was presented with background noise of bands 13 through 16 carrying SAM with a modulation rate of 64 Hz.

RESULTS AND DISCUSSION

I. Auditory Stream Segregation

Result 1. Auditory stream segregation based on spectral and AM-rate cues:

1) Overall analysis (two factors—spectral separation and AM-rate separation, and their interaction):

Fig. 4 shows d's averaged across subjects for the 12-pair sequences in Experiment I (stream segregation) for each of the spectral overlap and AM-rate conditions. A two-way (4x4) repeated measure Analysis of Variance (ANOVA) was conducted. The two factors were spectral (band) separation and AM-rate separation of A and B bursts. There were four levels for each factor, i.e. AbbnBbbn, A5-10B1-8, A7-12B1-8, and A11-16B1-8 for the spectral separation, and AM0-0, AM25-25, AM25-100, and AM 25-300 for the AM-rate separation.

There was a significant difference within spectral (band) separations ($F(3,27)=12.136$, $p<0.001$) where the statistical sphericity assumption for the dependent variables was met; a significant difference was shown within AM-rate separations ($F(3,27)=5.108$, $p=0.031$ where Greenhouse-Geisser method was adopted to compensate for the violation of sphericity assumption. No significance was found for the interaction of spectral (band) separation and AM-rate separation (F(9, 81)=1.213, p=0.299 where the sphericity assumption was met). These findings suggest that when either the spectral separation or the AM-rate separation increases, CI simulation listeners could perform better in segregating ongoing interleaved acoustic elements into different perceptual acoustic streams. Additionally, each of the main factors held the same trend across all the four levels of the other factor.

2) Pairwise comparisons between spectral separations:

With respect to the spectral (band) separation, pairwise comparisons were undertaken with Bonferroni adjustment for multiple comparisons. Significantly better performance was found for spectrally separated conditions of A11-16B1-8 and A7-12B1-8 than for the broadband condition of AbbnBbbn (p=0.001 and p=0.007 respectively). No other significant difference in performance was found between paired spectral (band) conditions. These results suggest that, the stream segregation performance in the greater-band-overlapping condition (i.e., A5-10B1-8) was neither better than that in the completely overlapping condition (i.e., AbbnBbbn) nor poorer than that in the either of the two conditions of less spectral (band) overlapping (i.e., A7-12B1-8 and A11-16B1-8). This indicated that, with the spectral (band) separation of A5-10B1-8, listeners were able to use the differences between A and B bursts to facilitate stream segregation to some degree; however, these differences were not sufficient enough for the listeners to generate a strong perception of stream segregation to perform considerably better than in the situation when there was no spectral (band) difference.

3) Pairwise comparisons between AM-rate separations:

Pairwise comparisons were also conducted to investigate the AM-rate separation effect with Bonferroni adjustment for multiple comparisons. Significant differences were revealed between AM25-300 and AM0-0 ($p=0.002$), as well as between AM25-100 and AM0-0 ($p=0.016$); but no difference was shown between other comparisons. These results suggest that when the AM-rate difference is 2 octaves or larger, it could be a cue for listeners to segregate the running A and B noise bursts in the current study into two auditory streams.

4) The effect of AM-rate separation within each spectral separation:

Further ANOVA with repeated measurement was conducted between AM separations within each band condition. Significant differences were found for the A5-10B1-8 spectral overlapping condition ($F(3,27)=9.082$, $p<0.001$ when sphericity assumption was met). Further pairwise comparisons between AM separations showed a larger average d’ in the AM25-300 condition than that in the AM25-25 condition ($p=0.001$) and that of the AM0-0 condition ($p=0.036$); and larger d’ in the AM25-100 condition than that in the AM25-25 condition (p=0.024). This reflects that, for this specific spectral (band) separation which was the greater-band-overlapping, the AM-rate difference of two octaves or larger resulted in stronger auditory stream segregation than what spectral (band) difference cue alone produced. This difference was not found in other spectral (band) separation conditions. (Statistical analyses will be described in the following paragraph.) The non-significant difference between AM25-25 and AM0-0 (p=1.000) is consistent with the notion that the differences of A and B bursts were essentially in spectrum only for both conditions: The AM rates on A and B bursts were the same in the either condition; thus, there was no AM-rate separation in either condition.

No significant difference in performance was found between AM-rate separations in any of the other three spectral (band) separations. The results from statistical analysis in the spectral (band) separation conditions were as follows: For AbbnBbbn condition, F(3,7)=0.430, and p= 0.734 with sphericity satisfied; for A7-12B1-8 condition, F(3,7)=0.413, and p=0.745 with sphericity satisfied; and for A11-16B1-8, F(3, 7)=3.681p=0.054 using Greenhouse-Geisser method to adjust for the violation of sphericity assumption.

It is noteworthy that, overall statistical findings did not show significant interactions between the two main factors (spectral or band separation and AM-rate separation), which indicated that the same trend of one of the factors held across all the levels of the other factor. However, the contrast that significant difference of d’s in the AM-rate separation occurred only in the greater-band-overlapping condition (i.e., A5-10B1-8) but not in the other spectral (band) separations implies that: a) when the spectral (band) separation cue is not strong, such as the greater-band-overlapping condition in the current study, the AM-rate separation cue could be an additional cue to facilitate the formation of stream segregation; b) when the spectral (band) separation cue does not exist, such as in the complete-band-overlapping condition (i.e., AbbnBbbn), the AM-rate separation alone is not sufficient for a clear stream segregation perception; c) when the spectral (band) separation is sufficiently large to generate stream segregation, such as in the moderate-band-overlapping condition (i.e., A7-12B1-8) and the no-band-overlapping condition (i.e., A11-16B1-8) in this study, the AM-rate separation does not contribute to stronger stream segregation.

5) Difference in stream segregation between two spectral separations—AbbnBbbn and A5-10B1-8:

Independent repeated measures of ANOVA were conducted between these two spectral (band) separations within each of the four AM-rate separations. The significant difference was found between the two spectral separations when A and B bursts were modulated at rates of 25 Hz and 300 Hz respectively (F(1,9)=10.432, p=0.01). No difference was found in other AM-rate separations, and the statistical estimates are as follows: a) in the condition of AM0-0, F(1,9)=1.862 and p=0.206; b) in the condition of AM25-25, F(1,9)=1.935 and p=0.198; c) in the condition of AM25-100, F(1,9)=3.905 and p=0.080) . This finding showed first that the spectral separation of A5-10B1-8 alone together with a smaller AM-rate separation (i.e., AM25-100) did not elicit stronger stream segregation than that in the conditions without spectral separation (i.e, AbbnBbbn). Second, with an additional larger AM-rate separation, a stronger segregation was elicited by the condition with the least spectral separation (i.e., A5-10B1-8) as opposed to the condition with no spectral separation (i.e., AbbnBbbn).

Result 2. Stream segregation versus gap discrimination

The current stimulus paradigm may not require only stream segregation to perform the task. Listeners could possibly focus on the last a few A-B pairs (or even the last pair) to identity the sequences with the last B bursts delayed. The control condition consisting three A-B pairs with the greater-band-overlapping (i.e., S-A5-10B1-8) was undertaken across all the four AM-rate separations. The hypothesis was that it would be easier to identify the delayed sequences with the 12-pair sequences than with the 3-pair sequences if a listener was based on stream segregation to undertake the task; whereas, the task would be equally easy or difficult regardless the length of the stimulus sequences, if the integrated perception or simply gap discrimination was the grounds for the task. It was expected that a significantly higher performance for the 12-pair sequences comparing the 3-pair sequences.

The selection of the greater-band-overlapping condition (i.e., A5-10B1-8) to control the length of the stimulus sequences was based on the following considerations drawn from pilot data:

a) The AM-rate separation in the completely-band-overlapping condition (AbbnBbbn) may not elicit stream segregation for some listeners. Therefore, the mechanism to perform the task may be same for the 12-pair and 3-pair conditions (i.e., catching the local gap-difference cue at the end of the stimulus sequences). Consequently, it was speculated that no difference between the two number-of-pair conditions would be seen, if the 3-pair condition had been conducted for the completely-band-overlapping condition. In this case, the goal of the control condition to examine stream segregation as the underlying mechanism for the experimental task would not be met.

b) In our pilot study, participants reported that the 3-pair stimulus sequences generated a strong perception of stream segregation when the spectral (band) separation between A and B bursts was large, such as that in the condition of A11-16B1-8. In addition, they also reported that they involuntarily followed the B bursts to determine whether the last B burst was delayed. Although listeners would supposedly show better performance in the 12-pair condition than in the 3-pair condition based on the build-up mechanism in stream segregation

(Bregman, 1990), the stream segregation perception may be sufficiently strong that it didn’t require as much build-up time. With this confounding factor, the contrast between the two conditions with different numbers of pairs would not be straightforward.

c) It was predicted that the difference between the two number-of-pairs conditions would be larger in the greater-band-overlapping condition (i.e., A5-10B1-8). The spectral (band) separation would create a cue for stream segregation but still weak enough so that listeners need to use focused attention to maintain the segregation perception. Due to these points, we expect to see the effect of stream segregation in the 12-pair condition but not in the 3-pair condition.

Figure 5. shows the comparison of performance with 12-pair sequences and with 3-pair sequences (control condition 1) in the greater-band-overlapping condition. All the ten listeners participated in the control condition 1. Repeated measures of ANOVA (2 factors, one with 2 levels in the number-of-pairs condition and one with 4 levels in the AM-rate condition) shows that d’s in the 12-pair conditions were larger than that in the 3-pair conditions (F(1,9)=8.021, p=0.02); that d’s were somewhat different in AM-rate conditions (F(3,27)=3.023, p=0.047 when sphericity assumption was satisfied), but no significant difference was found in the pairwise comparisons with Bonferroni adjustment method; and

that the interaction of number-of-pairs and AM-rate conditions was significant (F(3, 27)=3.902, p=0.019 with sphericity assumption satisfied).

The better performance in the 12-pair condition suggested our prediction that listeners were following the B bursts along the stimulus sequence instead of focusing on the end of the sequence, which was consistent with the notion that the listeners were performing stream segregation in the paradigm in Experiment 1. In addition, the effect of AM-rate on listener's performance showed that AM induced stream segregation when both 12-pair and 3-pair conditions were combined. This finding is consistent with the AM-rate effect in the 12-pair condition alone.

However, the pairwise comparisons showed significant differences when the 12-pair condition alone was considered (see Result 1), whereas they didn't show significance when the 12-pair and 3-pair conditions were both considered. This suggested that the addition of 3-pair condition decreased the effect of AM-rate difference, which is consistent with the significant interaction between the AM-rate and number-of-pairs conditions. The independent repeated measure between the 4 AM conditions for 3-pair conditions (no difference (F(3,27)=0.803, p=0.503)) confirmed this speculation.

The interaction of the conditions of number-of-pair and AM-rate separation was further investigated for the difference between 12-pair conditions and 3-pair conditions for each of the four AM conditions. Independent repeated measures of ANOVA were performed individually. Significant differences were found for AM300-25 and AM100-25 conditions (F(1,9)=11.964, p=0.007 and F(1, 9)=9.233, p=0.014, respectively), but not for AM25-25 and AM0-0 conditions (F(1,9)=2.545, p=0.145, and F(1,9)=1.089, p=0.324, respectively). These results indicated that, in the current spectral (band) condition (i.e., A5-10B1-8), listeners' ability to perform the task was similar between the 12-pair and 3-pair sequences when there was no AM-rate difference. This suggested that, listeners would perform limited stream segregation without AM-rate difference in this spectral separation, because their performance with strong stream segregation possibility (12-pair sequences) was not better than that with limited stream segregation possibility (3-pair sequences). On the other hand, the results showed that the performance for 12-pair sequences was better than that for 3-pair sequences when 2-octave or larger. This suggested that the addition of AM-rate difference to the spectral (band) difference generated stronger stream segregation. In other words, the interaction of number-of-pair condition and AM-rate condition in the greater-band-overlapping condition is consistent with the notion that AM-rate difference could be an additional cue for stream segregation when the spectral cue alone is insufficient.

Overall, Result 2 shows that, listeners were following the B bursts throughout of the sequences and performing stream segregation with the 12-pair sequences when the spectral (band) difference and/or AM-rate difference between

A and B bursts were sufficiently large. It is unlikely that they performed the task by capturing the local cue with focusing on the end of the sequence alone.

Result 3. Stream segregation versus detection of overall averaged gap range

Figure 7. shows the comparison of performance with 12-pair sequences and that with 3-pair sequences (control condition 2) in the complete-band-overlapping (AbbnBbbn) plus no AM-rate difference conditions (i.e., S-AM25-25 and S-AM0-0). Six randomly selected listeners participated in the control condition 2. This control experiment was undertaken to rule out that listeners calculated an overall range of A-B or B-A gaps with A bursts jittered when 12-pair sequences were presented, and compared the last A-B gap against this perception of overall range of gaps (see pages 27-30 for details). The hypothesis was, if listeners had been using the overall range of gaps as a baseline, they would have performed better with 12-pair sequences than with 3-pair sequences when A and B bursts were the same in spectrum and AM rate, which corresponded to the complete-band-overlapping and no AM-rate difference conditions. Repeated measures of ANOVA revealed difference between the performance with 12-pair sequences and 3-pair sequences (F(1, 5)=8.575, p=0.033). Although this finding did not completely rule out the possibility that listeners averaged the range of the A-B or B-A gaps throughout the stimulus sequence, the d's in the moderate-band-overlapping conditions (A7-12B1-8) and the no-band-overlapping conditions

(A11-16B1-8) for the 12-pair sequences were greater than their counterparts in the AbbnBbbn conditions. This improved performance indicates that listeners have segregated A and B streams in addition to acquiring the overall gap-range differences to perform the task.

Result 4. Gap delay discrimination thresholds:

The ten participants' gap delay thresholds averaging across 6 runs for each subject are shown in Figure 8, ranging from 11.70 ms to 25.6 ms. This means that the 30-ms gap delay in the stream segregation experiment was detectable to all the listeners. The standard deviations above the means ranged from 2.28 to 8.44, while the standard deviations below the means ranged from 1.91 to 5.80. The geometric mean of the ten thresholds was 16.6 ms with a standard error of 4.87 above the mean and a standard error of 4.06 below the mean.

A linear regression was performed with gap delay detection thresholds as an independent variable and the d' of a certain condition as a dependent variable. The regressions were conducted for all the individual conditions in the stream segregation experiment. No significant correlation was found between the two variables in most of the experimental conditions, except two AM-rate conditions (AM300-25 and AM100-25) in the moderate-band-overlapping condition (i.e., A7-12B1-8). The Pearson coefficients were -0.734 and -0.881, respectively, while the corresponding R squares were 0.481 and 0.749.

The absence of the correlation indicates that the measured d’ for stream segregation is independent of the gap detection ability. In other words, the d’ value reflects the strength of a perception different from the temporal gap detection, even though the stimulus paradigm itself requires the involvement of a detection of temporal gaps to perform the task. That is, the d’ reflects the strength of stream segregation.

Result 5. AM detection:

The averages of AM detection thresholds at three modulation rates of 25 Hz, 100 Hz, and 300 Hz were computed across all participants and are depicted in Figure 8. The thresholds of individual participants’ are shown in Table 2. Overall, the individual thresholds covering all the three modulation rates ranged from -23.00 dB to -7.92 dB, which verified that the AM in the experiment was detectable to all participants at all modulation rates.

The plot of AM detection threshold as a function of modulation rates showed a band pass feature. The thresholds at modulation rates of 100 Hz and 300 Hz were consistent with the ranges reported in Viemeister’s study (1979) using gated wideband-noise carriers with three durations which were 250 ms, 500 ms, and 1500 ms.

On the other hand, the thresholds at 25 Hz in the current study were considerably higher (worse) than those of Viemeister’s (1979) study. The average threshold in the current study was -11.57 dB, while the interpolation from Viemeister study with gated carriers was approximately -20 dB. This difference could be attributed to Viemeister's hypothesis of multiple looks which states that, for a gated carrier with a short duration, such as the one with an 80-ms duration in the current study, only a limited number of cycles of modulation is available for the listener; whereas, for a longer carrier, such as the 500-ms gated carrier in Viemeister (1979), the number of cycles available increases. Consequently, a listener has more "looks" with a longer carrier. In the case of an 80-ms carrier versus a 500-ms carrier, a 6.5-fold difference in the number of cycles of modulation is yielded. The increase in the number of modulation cycles allows more observational opportunities, i.e., "looks", which results in a better threshold with a longer carrier. This effect is more salient at low modulation rates. In the Viemeister's study, the carrier duration effect was not obvious when the modulation rate was 100 Hz or higher. The results in the current study are consistent with Viemeister's report.

Result 6. Speech perception:

Figure 10 shows listeners' understanding of sentence keywords through CI simulations with a resolution of 16 bands (see cutoff frequencies in Methods). Table 3 shows scores obtained from two lists for each condition in individual participants. The speech spectrum was available to listeners only in the lower ten bands in the current experimental design.

The scores varied greatly from one listener to another for the same conditions. For instance, listeners correctly identified 53-80% of the keywords with an average of 64.7% in the quiet condition and 10-54% when the noise spectrum was completely overlapping with the speech spectrum (i.e., the N5-10Spch1-10 condition). A univariate ANOVA was conducted to evaluate the between subject difference with subject as an independent variable and the speech understanding score as a dependant variable and showed a significant variability across subjects ($F(8, 216)=8.962$, $p<0.001$).

Post Hoc multiple comparisons using Tukey HSD method were further undertaken. The analysis showed significantly poorer performance from S4 and S7 when compared to the others. The sentence understanding scores were significantly lower for S4 than for eight out of the other nine listeners (p values <0.01 for all the eight comparisons). No difference was found between S4 and S7. S7scored significantly lower than three out of the other nine listeners, which were coded as S1, S2, and S11 (p values ranging from smaller than 0.001 to smaller than 0.05). The overall performance of eight out of the ten listeners, apart from S4 and S7, was not different from one to another.

Repeated measurement ANOVA on conditions with background noise revealed significant differences in spectral (band) condition; that is, as the spectrum of the background noise approached or overlapped more with the spectrum of speech, listeners identified fewer keywords (F(4, 144)=100.851, p <0.001 with Greenhouse-Geisser adjustment). However, different AM rates of the modulated bandpass noise did not render significant improvement in speech understanding. Speech perception with modulated noise did not improve from that with steady-state noise. No interaction between the spectral (band) separation and AM-rate separation was found either.

Further pairwise comparisons with Bonferroni adjustment were performed on the factor of spectral (band) separation between noise and speech. The keyword identification score of each noise-speech spectral (band) separation was significantly higher than those scores obtained in the smaller noise-speech spectral (band) separations..That is, with the five noise-speech spectral (band) separations, from large to small, N13-16Spch1-10, N11-16Spch1-10, N11-14Spch1-10, N7-12Spch1-10, N5-10Spch1-10, the speech understanding score was significantly higher in a certain separation than that in each of the separations with a smaller noise-speech spectral distance. The p values of the comparisons ranging from smaller than 0.001 to smaller than 0.05.

It is noteworthy that the score of the quiet condition was not included in the above repeated measurements. This was based on the consideration of the unbalanced data due to the fact that there was merely one data point for each listener in the quiet condition, while five data points were available for each listener for each level of either noise-speech spectral (band) separation or AM-rate separation.

As no difference in keyword identification was found between AM-rate separations, the scores from different AM-rate separations in the same noise-speech spectral (band) separation were averaged to yield a single score for each noise-speech spectral (band) separation. Pairwise comparisons from the repeated measurement of ANOVA conducted based on these averaged scores and the scores in the Quiet condition showed that the performance in the Quiet condition was significantly better than the conditions where the noise was present, except the condition where the spectra of noise and speech were completely separate—N13-16Spch1-5.

Result 7. Correlation between speech perception and auditory stream segregation:

The correlation of sentence understanding score and auditory stream segregation ability was evaluated with the linear regression method. The speech understanding scores were designed as a dependent variable. The d’s from the stream segregation experiment were set as the independent variable. Due to the insignificant effect of the presence of AM on noise or/and the AM rate of noise on speech perception, for each listener, an average score of speech perception in a certain spectral separation between speech and noise was calculated across the conditions with unmodulated background noise and with modulated background noise at all the four AM rates. Subsequently, five average scores of speech understanding in noise were obtained for each listener, each for one of the five spectral separations between speech and noise. The coefficients of a linear equation between average speech understanding score in a certain speech-noise spectral separation and the d's from all the conditions in the stream segregation experiment was estimated using the stepwise entry method. In the regression analysis, d's from conditions with unmodulated A and B bursts (i.e., AM0-0) were not included. Linear regressions were further analyzed with speech understanding scores from each of the AM rates of modulated noise in all the speech-noise spectral separations as the dependent variable, and d's in the stream segregation experiment as the independent variable, using the stepwise entry method. Again, in the regression analysis, d's from conditions with unmoduated A and B bursts (i.e., AM0-0) were not included.

Table 4. shows the all the linear equations of regression between the speech understanding score and the d' reflecting stream segregation ability. It had been hypothesized that, if speech understanding in noise could be explained to some degree by listeners' capability for stream segregation, correlations between them would be found. The following findings were predicted:

a) Scores in the speech perception experiment conditions with more spectral overlap between speech and noise bands would positively correlate with conditions in the stream segregation experiment with more overlap between A and B bursts.

b) Scores in the speech perception experiment conditions with the AM rate on noise closer to the syllabic rate in speech(typically around 4 Hz) would correlate positively with conditions in the stream segregation experiment with smaller separation between the AM-rates on A and B bursts.

The analysis shows:

a) The first five rows in Table 4 show stream segregation measured in some spectral and AM-rate separations could account for speech understanding in noise when only the spectral separation between speech and noise was taken into consideration (i.e., when the speech understanding scores were collapsed into an average score across the AM rates of the noise in one spectral separation). The stream segregation ability in the condition with the moderate-band-overlapping and largest AM-rate separation (i.e., A7-12B1-8/AM25-300) accounted for 77.8% and 56.8% of the variability in the speech understanding for the two largest spectral separations between speech and noise, N13-16Spch1-10 and N11-16Spch1-10 respectively. When the spectral separation between speech and noise decreased to the moderate-band-overlapping and greater-band-overlapping conditions—N9-14Spch1-10 and N7-12Spch1-10—the stream segregation ability in the condition with the moderate-band-overlapping and largest AM-rate separation (i.e., A7-12B1-10/AM25-300) could also account for 80.6% and 57.3%

of the variability in speech understanding scores, respectively. In the condition with complete-band-overlapping of speech and noise—N5-10Spch1-10—the summation of two conditions in the stream segregation experiment (i.e., A7-12B1-8/AM25-100 and A5-10B1-8/AM25-300) subtracting the effect of the condition with no spectral separation (i.e., AbbnBbbn/AM25-300) could account for 83.2% variability of speech understanding scores.

b) In the presence of unmodulated noise (i.e., steady-state noise), four out of the total of five speech-noise spectral separations were correlated with stream segregation obtained with various spectral and AM-rate separations between A and B bursts (rows 6 through 9). Speech perception in the conditions of N13-16Spch1-10/N-AM0 and N9-14pch1-10/N-AM0 both correlated with stream segregation in the condition of A11-16B1-8/AM25-300; speech perception in the condition of N11-16Spch1-10/N-AM0 correlated with stream segregation in the condition of A7-12B1-8/AM25-300; speech perception in the condition of N7-12Spch1-10/N-AM0 correlated with stream segregation in the condition of A7-12B1-8/AM25-100. No correlation was found between speech perception in steady-state noise and stream segregation in the conditions wherein AM rates of A and B bursts were both 25 Hz. Neither was speech perception in steady-state noise with the complete-band-overlapping between speech and noise (i.e., N5-10Spch1-10) found correlated with stream segregation.

c) In the presence of amplitude-modulated noise, regardless of the modulation rate, speech perception was correlated with d’s of two spectral (band) separations (regardless of AM-rate separations) between A and B bursts—A11-16B1-8 and A7-12B1-8. In five out the six spectral-overlapping conditions wherein speech perception was found correlated with stream segregation, speech understanding was correlated with stream segregation in the greater-band-overlapping condition—A11-12B1-8 (shown from rows 10 to 15); whereas, in two out of the five spectral-separated conditions (shown from rows 5 to 9), speech perception correlated with d’s in the no-band-overlapping (i.e., A11-16B1-8) and the speech understanding in the remaining three spectral-separate conditions still correlated with d’s in the moderate-band-overlapping condition (i.e, A7-12B1-8). This implies a relationship between speech perception skill in the presence of amplitude-modulated noise and stream segregation ability. It had been hypothesized that correlations between speech understanding and stream segregation would relate in a direct one-to-one manner, that is, speech understanding obtained in a condition with a certain spectral separation between speech and AM noise would correlate with stream segregation obtained in a condition with similar degree of spectral separation between A and B bursts. However, the findings from this study revealed that the stream segregation ability in the condition of moderate-band-overlapping (i.e., A7-12B1-8) was a more common predictor for most speech understanding in amplitude-modulated noise.

d) The analysis didn’t reveal one-to-one correlation between speech understanding in AM noise and stream segregation ability in terms of amplitude-modulation rate separation. This lack of one-to-one relationship could be partly accounted for by the insignificant effect of AM-rate of modulated noise on speech perception. On the other hand, the majority of linear regressions included a predictor of stream segregation in a condition with AM-rate separation between A and B bursts, which indicates speech perception in AM noise was still generally correlated with stream segregation based on AM-rate difference.

e) The analysis didn’t show that the stream segregation in the complete-band-overlapping condition (i.e., AbbnBbbn) could be a predictor for speech understanding skill in modulated noise regardless of the degree of spectral separation between speech and noise. This is probably because no significant stream segregation was obtained in the complete-band-overlapping condition.

Result 8. Correlation between speech perception and AM detection:

Correlations were examined between AM detection thresholds and speech perception scores in conditions when noise was either present or absent. With the Quiet condition (i.e., absence of noise) of the speech perception experiment, one score of correct keyword identification for each listener was included in the analysis to correspond to one average AM detection threshold at a certain modulation rate. With the conditions wherein noise was present, five speech perception scores—each from one of the five AM rates (i.e., steady-state noise, 4 Hz, 16 Hz, 32 Hz, and 64 Hz)—for each listener were included for each spectral (band) separation between noise and speech.

Correlations were evaluated using linear regression. Each set of AM detection thresholds obtained at a certain AM rate was assigned as an independent variable, while the speech perception scores obtained in a certain spectral (band) separation condition were entered as an dependent variable. An independent regression test was undertaken between one of the three AM detection conditions (i.e., three AM rates) and one of the spectral (band) separation conditions in the speech perception experiment.

No correlation was found between speech perception in quiet and AM detection threshold. In the presence of noise, a consistent correlation between AM detection thresholds at the AM rate of 25 Hz and speech perception scores was revealed as shown in Figure 11, except that when the spectra of noise and speech were completely overlapping (i.e., the condition N5-10Spch1-10). On the other hand, speech perception score in the condition of N5-10Spch1-10 was found to correlate with AM detection threshold at the rate of 300 Hz. No correlation between speech perception and AM detection threshold at the modulation rate of 100 Hz was found, which is different from the correlation between speech perception and AM detection threshold at the rate of 100 Hz reported in CI users by Fu and Nogaki (2004).

Following an examination of the figure 11, the correlation of between AM detection threshold at the modulation rate of 25 Hz and speech understanding in each of the spectral (band) separation conditions could be the result of just one of the listeners (S4) whose AM detection threshold was the poorest and whose speech understanding was at the lower end as well. The removal of this participant’s data made the correlation between speech perception and AM detection threshold disappear. However, the previous analysis in Result 5 (Speech Perception Results) showed poorer speech perception scores in two participants (S4 and S7). S7 also showed a poor AM detection threshold at the rate of 25 Hz. Therefore, the disappearance of the correlation between speech perception and AM detection threshold due to the removal of S4’s data may have been resulted from the reduced power of the analysis rather than the removal of an outlier. Consequently, it is plausible to include all the subjects in the analysis.

In summary, the results have shown that CI simulation listeners were able to segregate auditory streams based on spectral and AM-rate cues. Their speech perception in noise was better when the spectra of speech and noise were farther separate but was not affected by the AM rate of noise. The listeners’ capability of stream segregation based on both spectral and AM-rate cues, as well as their AM-depth detection ability, can be a predictor for their speech perception in noise.

General Discussion

I. Stream segregation

Overall the findings indicate the following:

- Stream segregation can occur for amplitude-modulated noise band stimuli that are similar to CI simulations.
- Frequency separation of the noise bands was a stronger cue to segregation, with AM rate a secondary cue. In a limited number of conditions, AM and spectral separation interacted to improve streaming, but in most conditions spectral separation alone accounted for the stream segregation.
- Findings are consistent with previous research on stream segregation using tones and noises.

1. Cues of stream segregation in the current study

1) Spectral separation was the primary/strongest cue for stream segregation

Results from the current study show that normal hearing listeners could segregate two interleaved subsequences of noise elements into two auditory streams when the noise subsequences are different in spectrum. There were four levels of spectral separation of interleaved A and B noise bursts involved: no-band-overlapping (A11-16B1-8), moderate-band-overlapping (A7-12B1-8), greater-band-overlapping (A5-10B1-8), and complete-band-overlapping (AbbnBbbn). As the band overlapping between A and B noise bursts increases, the stream segregation decreases.

This can be accounted for by the ‘‘peripheral channeling’’ model proposed by Hartmann and Johnson (1991), which explains that the formation of auditory stream segregation relies largely on the difference in the spectral excitation pattern of the two interleaved “auditory streams”. The development of this hypothesis arose from a series of experiments that they continued based on Dowling (1973, Experiment 1). In the experiment, two familiar melodies were presented to the listeners in a way such that the notes temporally alternated between melodies, with the notes of one melody occupying the odd temporal locations and the notes of another melody occupying the even locations. The task for the listeners was to identify both melodies. Listeners’ accuracy of identifying the melodies (i.e., measure of stream segregation ability) was recorded with differences between the two melodies in various acoustic attributes. Hartmann and Johnson drew the conclusion that “peripheral channeling is almost an adequate explanation of stream segregation effect” (page 175) based on their data.

To examine the peripheral channeling hypothesis, the authors used the difference in various acoustic attributes of the notes in the two melodies to obtain various levels of spectral overlapping of the two melodies. This was similar to the concept of increasing the amount of band overlap in the current study. Four levels of band overlap were included in Hartmann and Johnson's study, from essentially no overlap to completely overlapping.

Hartmann and Johnson (1991) found that listeners obtained significantly higher accuracy in segregating the two melodies in the three conditions with spectral differences than in the completely-overlapping conditions. In the current study, we found better stream segregation performance in the two conditions with larger spectral separations (A11-16B1-8 and A7-12B1-8) than in the complete-band-overlapping condition (AbbnBbbn). When the spectra of the A and B bursts were more overlapping (i.e., A5-10B1-10), the small spectral difference did not render better performance than the complete-band-overlapping condition.

Two factors could potentially explain the insignificance of the effect of small spectral difference on stream segregation in the current study in contrast to the significance in Hartmann and Johnson (1991).

First, Hartmann and Johnson used sinusoids and harmonics as stimuli, whereas bandpass noises were used in the current study. The harmonics in Hartmann and Johnson could be resolvable in the peripheral auditory filters (especially in the low frequency region for the first four harmonics) whose bandwidth can be calculated from the formula (Glasberg and Moore, 1990):

$$ERB_N=24.7(4.37F+1); \qquad (2)$$

where ERB_N stands for equivalent rectangular bandwidth for normal-hearing listeners, F for the center frequency of the peripheral auditory filter band. The mean tones for both melodies were 440 Hz, which resulted in the first two odd harmonics of 440 Hz and 1280 Hz and the first even harmonics of 880 Hz and 1760 Hz. The calculated equivalent rectangular bandwidths (ERB) are 72 Hz and 163 Hz for the peripheral auditory filters centered at the first two odd harmonics, as well as 120 Hz and 215 Hz for the first two even harmonics. These bandwidths show that the first two odd and the first two even harmonics fall into four separate auditory filters. Consequently, the two melodies—one composed with odd harmonics and the other one composed with even harmonics—could be segregated into different streams based on the spectral difference from the resolved harmonics.

On the other hand, interleaved bandpass noises may elicit weaker stream segregation than interleaved sinusoids at the center frequencies of the noises, even when the individual noise bands are within the range of different single auditory filters. The noises may somehow diminish the perception of stream segregation. This is supported by Dannenbring and Bregman (1976). The authors used sinusoids and bandpass noises to investigate the strength of stream segregation. They presented pairs of sinusoids or bandpass noises to the listeners repeatedly. The listeners reported whether they had perceived one or two streams after each trial. The researchers paired sinusoid with sinusoid, sinusoid with bandpass noise, and bandpass noise with bandpass noise. One of the conditions with a fast presentation rate incorporated an onset-to-onset time similar to the BRT (i.e., burst repetition time—the onset-to-onset time between two adjacent noise bursts) in the current study. Dannenbring and Bregman used 135 ms including a 120-ms sinusoid (or noise burst) and a 15-ms silent gap, which was close to the 130-ms BRT in the current study. Thus, their results have a logical application to the current study. Dannenbring and Bregman compared the strength of stream segregation generated by pairing bandpass noises and that by pairing sinusoids, while the center frequencies of the paired bandpass noises were set the same as the frequencies of the paired sinusoids. Both small and large spectral separations between the two elements in the pair were studied, with 1000 Hz versus 1200 Hz for a small separation and 1000 Hz versus 3000 Hz for a large separation. Although the researchers found that listeners could distinguish the bandpass noise with a center frequency of 1000 Hz from the noise with a center frequency of 1200 Hz, their results show that pairing noises induced weaker stream segregation than pairing sinusoids when the spectral separation was small. However, pairing sinusoids and pairing bandpass noises—with frequencies of the sinusoids same as the center frequencies of the bandpass noises—elicited the same strength of stream segregation. Therefore, the use of bandpass noise rather than sinusoids or harmonic complex may result in different strength in stream segregation, hence may be an account for the lack of segregation when the center frequencies of the two noise passbands are only slightly separate.

Second, the lack of stream segregation for the greater-band-overlapping condition may be attributed to the large amount of spectral overlap between the two noise bands in the current study. In the Dannenbring and Bregman (1976) study described previously, stream segregation was still elicited by bandpass noises even when their center frequencies were very close—200 Hz apart. Dannenbring and Bregman used 1/8-octave bandwidths which led the two narrow bands of noise to fall in two 1/3-octave peripheral auditory filters at the center frequencies of the corresponding noise bands. When the center frequencies of the two noise bands were 1000 Hz and 1200 Hz respectively, the two peripheral auditory filters activated by the two noise bands were slightly overlapping and Glasberg and Moore's (1990) formula estimates equivalent rectangular bandwidths of 130 Hz and 133 Hz for them respectively. The frequency ranges of the two filters are, therefore, 870-1130 Hz and 1067-1333 Hz. The overlap of these two auditory filters in terms of percentage of the total frequency range from both filters can be calculated. Using the equation (1130-1067)/(1333-870), a spectral overlap of 13.6% was obtained.

In the current study, due to the wide bandwidths, a noise passband was exciting multiple auditory filters at the same time. In the greater-band-overlapping conditions (i.e., A5-10B1-8), although the spectra of the A and B bursts did not overlap in the frequency ranges between 200 and 591 Hz and between 1426 and

2149 Hz, the large amount of overlapping frequency range between 591 and 1426 Hz resulted in a 42.8% overlap in frequency. This frequency overlap is larger than that from the peripheral auditory filters of the most closely placed narrow-band noises in Dannenbring and Bregman (1976). This large spectral overlap may have rendered a sufficient similarity of the two noise bursts that outweighed the spectral non-overlap and that led the listeners unable to segregate A and bursts into different streams

Furthermore, a shallow spectral slope of -12 dB per octave for the filters used in the current study was chosen to simulate real CI users' frequency turning curve (Anderson, et al, 2007), in contrast to the slope of -48 dB per octave in Dannenbring and Bregman (1976). This shallow slope was likely to have resulted in more spectral overlap of the A and B bursts, which additionally contributed to the diminished stream segregation. This postulation is supported by Bregman, Liao, and Levitan's work (1990). They used vowel-like sounds to investigate the effects of *F*0 frequency and spectral shape on stream segregation. The stimuli were complex tones with a single formant peak. The authors varied the sharpness of the formants while keeping the distances in center frequency the same, and found that sounds with sharper formant peaks elicited stronger stream segregation. This could be attributed to the more spectral overlap resulting from the broader formant bandwidths. Consequently, the large band overlapping of the A and B bursts as well as the shallow spectral slopes of the A and B bands may have resulted in the spectral similarity to a degree that no stream segregation could be shown when A and B bursts were only different in spectrum by the overlapped conditions of A5-10B1-8.

2) AM-rate separation appeared to be a secondary cue for stream segregation

We found that AM could aid stream segregation in some conditions. Control conditions were run to determine the audibility of the AM and to confirm that AM differences (not merely the presence of AM) contributed to stream segregation.

To examine the effect of AM-rate separation on stream segregation in CI simulations, four AM-rate separation conditions were studied. The conditions of AM0-0 served as a baseline showing the strength of stream segregation based on spectral separation alone. The performance in the segregation task with AM-rate separations was contrasted against the corresponding baseline performance. Better performance in the conditions with an AM-rate separation would show AM-rate as a cue for stream segregation. The conditions of AM25-25 were carried out to examine whether any segregation was indeed due to the difference in AM rate not merely due to the presence of AM. The finding of insignificant differences between conditions of AM25-25 and AM0-0 (Result 1) indicates that the application of AM of the same rate to A and B bursts did not result in improvement in performance from that in the conditions when no AM was applied to either burst.

A supplemental experiment measuring listeners' thresholds of AM depth detection was conducted to evaluate the audibility of the amplitude modulations carried by the A and B bursts. The thresholds at the three modulation rates—25, 100, and 300 Hz ranged from -7.91 dB to -23 dB across all the listeners (Table 2). The modulation depth employed at all modulation rates in the stream segregation experiment was uniformly 0 dB which was at least approximately 8 dB above the thresholds. Consequently, all the amplitude modulations were supposedly audible to all the listeners.

To reduce AM interference introduced by the intrinsic fluctuations from narrow-band-noise carriers (Dau, et al, 1999), wide-band noises were used in the experiment (see Parameters in Experiment 1). This was realized by combining the several individual vocoder bands with a 16-band resolution in Fu and Nogaki (2004) into a single wide band. The smallest bandwidth of the all the bandpass noises was 1226 Hz of the B bursts. This bandwidth was sufficient (Zwicker and Fastl, 1990) to smooth the noise waveform hence to reduce the interference from intrinsic noise fluctuations on AM detection and assure the audibility of the AM.

Results showed better performance associated with AM-rate separations of 2 octaves or larger (i.e., 25 vs 100 Hz and 25 vs 300 Hz) when compared to no AM on the A and B bursts. This finding indicates that AM-rate difference could be a sole cue for stream segregation, which is consistent with the report by Grimault, et al (2002). However, they reported a lower (better) AM-rate difference—1 octave—for eliciting stream segregation. This may be due to several factors:

First, the current study did not include smaller AM-rate differences than 2 octaves. Although our preliminary data using the Roberts et al (2002) method did not show that 1-octave AM-rate difference was sufficient to elicit stream segregation (Nie and Nelson, 2007), it is uncertain whether listeners in the current study could stream-segregate when the AM-rate difference is between 1 and 2 octaves using the current method where intention focused on segregation.

Second, the longer duration and higher modulation rates in Grimault, et al study may have led to segregation using lower AM-rate differences. The stimulus sequences in the Grimault, et al study were composed of repeated ABA triplets. The duration of the bursts was 100 ms and the AM rates were between 100 and 800 Hz—as opposed to 80-ms burst duration and modulation rates of 25, 100 and 300 Hz in the current study. The combination of the longer noise bursts and the faster AM rates in Grimault et al study enabled more modulation cycles in each burst duration. This may have provided the listeners with a stronger cue than that in the current study and Nie and Nelson (2007).

Third, the different repetition times in the current study and Grimault et al study may also account for the difference. Grimault et al used a repetition time of 120ms which is slightly shorter than the 135 ms used in the current study. According to Van Noorden’s (1975) finding, when the repetition time is greater than 90 ms, the TCB (i.e., Temporal Coherence Boundary), which reflects the frequency difference to elicit stream segregation as a function of repetition time, changes abruptly. For instance, for repetition times between 60 and 90 ms, the TCB is stable around 4 and 5 semitones. In contrast, the TCB increases from approximately 8 semitones at a repetition time of 120 ms to 10 semitones at a repetition time of 130 ms and to 11.5 semitones at a repetition time of 140, which corresponds to a slope of approximately 2 semitones per 10 ms. Consequently, the extending the repetition time from 120 ms as in Grimault et al to 130 ms as in the current study may presumably result in, at least, an increase of 2-3 semitones in AM-rate difference needed to elicit stream segregation. If the broader AM filters than the peripheral frequency filters (Ewart and Dau, 2000) were taken into consideration, this increase in AM-rate difference would be larger.

Finally, Grimault, et al measured stream segregation thresholds in terms of AM-rate difference; whereas a super-threshold AM-rate difference may be needed to perform the task in the current study. However, it is uncertain regarding the minimum AM-rate difference required for the current task to measure stream segregation; as it was not the intent of the current study, no pertinent data were collected.

3) Spectral separation and AM-rate separation seemed to interact to improve segregation for greater-band-overlapping conditions.

Overall, the current study didn’t find a significant interaction between band separation and AM-rate separation for stream segregation. This suggests that the same trend—the larger AM-rate separation the stronger stream segregation–holds within each of the band (spectral) separations. The significant effect of AM-rate separation on stream segregation in overlapping band conditions A5-10B1-8 indicates that the AM-rate separation provides an additional cue for stream segregation when spectral differences are pushed to a very limited degree but not completely diminished. When the band (spectral) separation between A and B bursts becomes either completely overlapping (i.e., the conditions of AbbnBbbn) or overlapping very little as in the conditions A7-12B1-8 or A11-16B1-8, the AM-rate cue did not significantly improve stream segregation. The interpretation of this finding in addition to the insignificance of the interaction between the AM-rate cue and band (spectral) cue is that the effect of AM-rate separation potentially exists in the other three band (spectral) conditions but is not sufficiently strong to be revealed. In the conditions of AbbnBbbn wherein no spectral separation is available, the AM-rate separation may be too weak to elicit stream segregation on itself. On the other hand, in the conditions of A7-12B1-8 or A11-16B1-8 wherein the spectral separation is larger, the amount of spectral separation is sufficiently large to elicit stream segregation—which can be inferred from the high d’ values in these conditions and the cue of AM-rate separation is too weak to make any statistical difference.

Interestingly, in the only spectral separation (i.e., A5-10B1-8), wherein AM-rate separation resulted in improved segregation, the performance based on spectral cues was neither significantly better than that in the conditions of AbbnBbbn nor significantly worse than that in the conditions with larger separations. This suggests that the AM-rate cue may have its greatest effect when the stream segregation elicited by spectral cues is borderline. This idea is supported by analyzing the data from a different angle—although the difference in spectral separation between A and B bursts did not result in an overall difference in stream segregation between the greater-band-overlapping condition (i.e., A5-10B1-8) and the complete-band-overlapping condition (i.e., AbbnBbbn), further analyses showed that, with the addition of a large AM-rate separation (i.e., modulation rate of 25 Hz versus 300 Hz), a significant difference is seen between the two different spectral separations. It is noted that this difference is attributed to the combination of AM-rate separation and spectral separation rather than the AM-rate separation alone. The same AM-rate separation (i.e., 25 Hz vs 300 Hz) was applied to both spectral conditions; had there been no difference in the two spectral conditions, the same AM-rate separation would not have resulted in difference in stream segregation. On the other hand, the application of AM with no rate difference or the smaller rate difference (i.e., 25 Hz versus 100 Hz) to both spectral conditions did not render a significant difference in stream segregation between the two spectral conditions. (Note that the statistical measure ($p=0.080$)

with the addition of AM25-100 is approaching significance.) These suggest that the AM-rate separation has to reach a sufficient level to facilitate the limited cues of spectral separation for stream segregation. The findings in the current study are not sufficient to derive an equation to show the amount of AM-rate separation changes as a function of the amount of spectral separation that would yield a significant increase in stream segregation based on the combination of the two cues. It was not the intent of this research. Nevertheless, the current findings provide evidence that, when the spectral cues are extremely limited, the addition of AM-rate difference may bring in substantial contribution to increase the strength of stream segregation in CI simulation listeners.

4) Loudness difference

Previous research has shown that introducing a difference of 6-8 dB in level between two interleaved melodies resulted in listeners' improved performance for identifying the individual melodies (Hartmann and Johnson, 1991). Stainsby, et al (2004) documented that, with a 5-dB or larger difference in level between the two sets of complex tones (A set and B set) presented in a pattern of iterated triplets (ABA) following van Noorden (1975), listeners perceived stronger stream segregation than what they did without a level difference. They also found that an obligatory stream segregation—a perception of stream segregation even when the listener's mental effort is on integration—

could also be formed with at least a 5-dB difference in level between the A and B sets of unresolvable complex tones when they were presented using the Roberts et al (2002) paradigm. These findings are consistent with the notation that a loudness difference may lead stream segregation. However, no level difference smaller than 5 dB has been reported to induce significant stronger stream segregation.

To examine whether the loudness could be a potential cue for stream segregation in the current study, the loudness level of each wide-band noise was calculated following the method described by Glasberg and Moore (2006). This method was based on the loudness model proposed by Moore et al (1997). The predicted loudness levels of the four unmodulated noise passbands are: 64.6 phons for the passband including vocoder bands 1 through number 8, 66.1 phons for the passband including vocoder bands 5 through 10, 65.1 phons for the passband including vocoder bands of 7 through 12, and 63.5 phons for the passband including vocoder bands 11 through number 16. As B bursts were fixed at the passband of vocoder bands of number 1 through number 8 (B1-8) and A bursts were varying across the other three passbands, the individual differences in level between the passband consisting of the first four vocoder bands and other passbands can be used as a factor to predict the potential existence of loudness cue for stream segregation. These three differences are -1.5, -0.5, and 1.1, all of which are smaller than 5 dB. In consequence, it is unlikely that loudness difference between the A and B bursts created a cue for stream segregation in our study.

5) Effect of focused attention on stream segregation

The current study attempted to use an objective approach to quantify stream segregation when listeners' attention effort was focused on holding a perception of stream segregation. This attention effort is consistent with the schema-driven mechanism described by Bregman (1990). The schema-driven mechanism and the primitive mechanism are referred to as opposite in terms of the need for focused attention to form stream segregation. Bregman proposed that a primitive mechanism forms definitive stream segregation due to strong difference in the physical properties between the interleaved subsequences even when the attention effort is directed to forming an integral perception; whereas schema-driven mechanism forms stream segregation only with the facilitation of attention effort directed to it because the physical properties of the interleaved subsequences are ambiguous and the direction of attention effort will determine the perception of either stream segregation or integration.

As has been noted in the Introduction, both behavioral and neurophysiological studies have documented the involvement of attention on stream segregation. Brochard, et al (1999), in their behavioral study, reported that less attention effort was needed for stream segregation when only the focused subsequence was present compared with when both focused and unfocused subsequences were present; and that, less attention effort was exerted when the frequency separation between the subsequences was larger. Both findings suggest that with the modulation of more attention effort, perception of stream segregation can be preserved when the physical properties of a stream in a mixture of various streams are obscure.

Botte et al (1997) found that, with equivalent performance in identification of a temporal irregularity in either the nonfocused stream or focused stream when elements of both streams were interleaved, the presentation level needed for the focused stream was 15 dB lower than that for the nonfocused stream. This indicates that focused attention could induce stronger stream segregation.

Neurophysiological studies (Sussman, et al, 1998; 2005) have shown that, while the mismatch negativity (MMN) reflecting automatic detection of of stream segregation prior to a conscious judgment was absent due to the ambiguous differences between the two subsequences, focused attention directed to segregation could elicit it. These findings provide physiological support to the notion that focused attention may reverse the absence of stream segregation to presence.

Overall, the findings in the literature suggest that the focused selective attention may compensate for the insufficiency of the reduced differences in physical properties between interleaved subsequences for auditory stream segregation. The auditory signals available for CI users are degraded relative to those available for people with normal hearing; thus, CI users may need to execute intensively focused attention to segregate auditory streams with the limited cues.

Although it is inconclusive, research has been emerging to show that CI users might not be capable of forming obligatory stream segregation based on primitive mechanisms (Cooper and Roberts, 2007; Gaudrain, et al, 2008). On the other hand, Chatterjee, et al (2006) used a traditional method with which subjective report of perception of number of streams was recorded to investigate stream segregation in CI users. Their results show that, when the listeners' attention was free instead of being directed to segregation or integration, one out of five CI listeners participated in the experiments showed definitive stream segregation. In the current study, a stimulus paradigm requiring focused attention directed to stream segregation was used, which may provide some insights on the debate whether CI users could perform stream segregation.

The finding in our current study that normal-hearing listeners can generate stream segregation in CI simulations suggests that, the one successful CI participant in Chatterjee et al study may have, at least partly, used the schema-driven mechanism. That participant showed stream segregation when the stimulus subsequences carrying different AM rates were exciting the same electrode pair. That is, the participant could have used the AM-rate difference cue alone to form stream segregation when the spectral cues were completely eliminated. This finding is in contrast to the implication from the current study that AM-rate difference alone cannot be a sufficient cue for CI users to segregate auditory streams. The explanation to this discrepancy is as follows:

Chatterjee, et al used a faster presentation rate which would have made it less demanding on the strength of cues for stream segregation. Their burst repetition time (BRT) was 100 ms, including a 50-ms burst and a 50-ms silent period followed the burst, as opposed to 130 ms in the current study. Investigation of the plot showing TCB and FB in van Noorden, one would find that TCB increases abruptly when the tone repetition time (TRT) is beyond 90 ms. A frequency difference falling in the region between TCB and FB would be much closer to the TCB—the boundary beyond which obligatory stream segregation would form based on the primitive mechanism—at a faster presentation rate, for instance, 100 ms than it would be at a slower rate, such as, 130 ms. This suggests that, in Chatterjee et al study, the frequency difference is closer to generating obligatory stream segregation based on the primitive mechanism, which is consistent with a less demanding requirement for the difference between the two streams for segregation.

Consistent with our findings, Hong and Turner (2009) developed a stimulus paradigm in favor of stream segregation (see page 20 for detailed description of their methodology) and found that both normal-hearing listeners and CI users may form stream segregation based on temporal envelope cues alone using the schema-driven mechanism.

However, there are two discrepancies between Hong and Turner (2009) and our findings with respect to effectiveness of AM-rate separation on stream segregation with a carrier of broadband noise and the maximum modulation rate that could elicit stream segregation.

First, Hong and Turner found that a one-octave AM-rate separation in the absence of spectral cues could elicit stream segregation; whereas, in the current study, no statistical significance was found based on the AM-rate separation (even when it is two octaves or larger) with broadband noise carriers. The following explanation may account for this discrepancy.

The duration of the noise bursts in Hong and Turner was longer than that in the current study, with 125 ms in Hong and Turner and 80 ms in our study. In addition, Hong and Turner used a faster AM rate (80 Hz) as reference, whereas the current study used an AM rate of 25 Hz. Both longer duration and faster rate would have resulted in more "looks" of the AM, thus probably allowed listeners to segregate the two modulation rates more easily. The findings in both Grimault et al (2002) and Hong and Turner that weaker stream segregation was associated with shallower AM depth may suggest that, when the perception of AM decreases, the strength of stream segregation based on the AM-rate cue decreases as well. The presumably decreased stream segregation resulting from the fewer chances to observe the reference modulation the current study may account for the discrepancy between the current study and Hong and Turner (2009). Furthermore, with the presumably more favorable design for stream segregation based on AM-rate separation in Hong and Turner, still three out of twelve listeners with normal hearing could not perform stream segregation based on AM-rate separation alone. It is anticipated that stream segregation is more difficult to form with the setup used in the current study.

Second, it was found in the current study that the AM rate of 300 Hz when paired with the AM rate of 25 Hz was a cue for stream segregation, whereas, Hong and Turner (2009) found that AM rate of 300 Hz was not a cue for stream segregation. By investigating the AM depth detection thresholds at 300 Hz (Figure 4 in Hong and Turner), it is noted that the average threshold of normal-hearing listeners was approximately 60% which is equivalent to -4.44 dB. This is poorer than the average of -12.49 dB obtained in the current study. With such a poor AM depth detection threshold as -4.44 dB, even the 100% AM may not be sufficiently salient for the listeners to follow. In particular, it might even not be audible to some of the listeners with thresholds poorer than average. Grimault, et al (2002) studied threshold in AM-rate separation for stream segregation as a function of AM depth and plotted in their Figure 3. The interpolation of 60% AM depth finds a threshold of approximately 1.5 octaves in AM-rate separation. The reference AM rate in Grimault et al was 100 Hz which implies to that the other

AM rate has to be 283 Hz or above for the listeners to perform stream segregation. The lack of stream segregation at an AM rate of 300 Hz in Hong and Turner (2009) may reflect a weak perception of AM strength that may or may not result in stream segregation.

In addition, Hong and Turner tested the strength of the cue as a function of AM rate by interleaving the modulated noise bursts with unmodulated noise bursts. With the poor AM depth detection threshold at the rate of 300 Hz, the AM noise may not be perceived as much different from unmodulated noise. Therefore, the A and B noise bursts may have sounded similar to the listeners. In contrast, we always interleaved modulated noise bursts at an AM rate of 25 Hz with those at the to-be-examined AM rate which was separated from 25 Hz in various distances. The contrast between A and B bursts in the AM domain may be more salient than that in Hong and Turner. Begman (1990) suggested that the more similar the subsequences, the more difficult they are for stream segregation. The difference in the level of similarity of the two subsequences between the two studies may also account for the discrepant findings.

6) Alternative explanations for performance:

It might be argued that stream segregation per se is not necessary for explaining these results. It is possible that listeners focused on the last pair of A and B bursts (instead of focusing on the ongoing sequence) still were able to detect the signal sequence with the last B burst delayed. This is actually an alternative explanation of the hypothesis that listeners could have used the cue in overall averaged gap range (stated on pages 28 and 29). With regard to signal detection theory, a calculation was conducted to simulate how well an ideal observer could detection the delayed B burst if only the last pair of A and B bursts were presented. The signal distribution (S) with the delayed B bursts and the noise distribution (N) with jittered-advanced B bursts were created. N is a rectangular distribution ranging from 0 to 90 ms in the silent gap between the offset of the A burst and the onset of the B burst with a mean of 45 ms. This range was derived in the following way: In the noise (non-signal) pairs of A and B bursts, the normal gap between A and B bursts is 50 ms; the jitter of ±40 ms on A bursts results in a range of 50±40 ms—range from 10 to 90 ms—in A-B gaps; the random advancing of B bursts in a range from 0 to 10 ms made the effective A-B gaps ranging from 0 to 90 ms with a mean of 45 ms. S is a rectangular distribution with ranging from 40 to 120 ms with a mean of 80 ms. This is the range of A-B gaps which was derived from adding the 30-ms delay of B bursts on the A-B gaps ranging from 10 to 90 ms resulted from the jitter of ±40 ms on A bursts. Five hundred thousand trials were applied to estimate the possibility of a value falls in the range of 40 to 120 ms is indeed from the S distribution and another five hundred thousand trials were adopted to estimate the possibility of a value falls in the range of 0 to 90 ms is indeed from the N distribution. The summation of the two probabilities renders the percent correct out of one million trials. The d’ for an ideal observer was derived from the percent correct and was approximately 1.07. The standard error of the d’ from the simulation was estimated based on 240 observational trials (120 trials each for signal sequences and non-signal sequences) in the behavioral experiment using the equations proposed by Gourevitch and Galanter (1967) (Equation 1).

$$var(d') = \frac{H(1-H)}{Ns[\phi(H)]^2} + \frac{F(1-F)}{Nn[\phi(F)]^2} \tag{3}$$

where var stands for variance, H for hit rate, F for false alarm rate, Nn for the number of trials with non-signal sequences, Ns for the number of trials with signal sequences, $\phi(H)$ and $\phi(F)$ for the heights of the normal density functions at the Z scores of H and F. The function ϕ can be calculated by Equation 2.

$$\emptyset(p) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z(p)^2} \quad , \tag{4}$$

where $z(p)$ is the Z score of a probability.

The estimated standard error (σ) is the square root of the variance, which was approximately 0.1705. The 95% confidence interval with the center d’ at 1.07 is 1.07±1.96 σ, that is, 1.07±0.33, and the resultant range is 0.74~1.40.

This range is consistent with the average d's obtained from the control conditions encompassing three pairs of A and B bursts with the same spectrum and AM rate; that is the conditions of AbbnBbbn/AM0-0 and AbbnBbbn/AM25-25 whose average d's were 1.09 and 1.07 respectively. However, in the 12-pair conditions, the listeners' performance was better than the predicted performance based on the last pair only when the A and B bursts were same in both spectrum and AM rate; the average d's were 1.52 and 1.56 for the conditions of AbbnBbbn/AM0-0 and AbbnBbbn/AM25-25 respectively.

One possible explanation to this better performance for the 12-pair conditions may be attributed to the hypothesis that some listeners were able to follow their memorized rhythm of the evenly spaced B bursts to a certain degree that assists the identification of the delayed B burst. This rhythm-based detection method might be used when the stimulus sequences were either 12 pairs or three pairs of A and B bursts. However the method may work better for the longer sequences (i.e., 12 pairs) as the listeners could have a longer time window to count the rhythm.

Nonetheless, this better performance associated with the potential of using rhythm cues did not eliminate the explanation that listeners were segregating streams to perform the task, because in those conditions (with spectral and AM-rate separations) that stream segregation was seen, the d's were greater than their counterparts in the AbbnBbbn/AM25-25 and AbbnBbbn/AM0-0 conditions. This improved performance indicates that listeners segregated the A and B streams in addition to following the rhythm to perform the task.

It appears that some form of stream segregation aided listeners for their use of spectral separation and AM-rate separation in the current study. This suggests that CI users can potentially use of spectral and AM-rate cues to segregate streams of stimulus sequences when their focused attention is directed to segregation.

II. Speech perception

Overall, results from the speech perception studies indicated large variability between listeners similar to that found in previous studies ((Friesen, et al, 2001; Fu and Shannon, 1998; Nelson, et al, 2003). In addition, findings suggested that the amount of spectral difference between background noise and speech affect speech understanding while the AM-rate of noise may have no effect on speech understanding.

1) With a resolution of 16 bands, the performance related to retaining the lower 10 bands in quiet

The performance of speech understanding in quiet ranging from 53-80% correct keyword identification provided an appropriate baseline to observe speech perception in various noise conditions. This medium range avoided ceiling effects when no noise was applied, as well as floor effects when noise was added in the further conditions. This was realized by adopting a 16-band resolution in CI simulations and maintaining the spectral information of speech of the ten bands at the low frequency end while eliminating that of the six bands at the high frequency end.

The application of 16-band spectral resolution in CI simulations was based on the previous findings on speech understanding through CI simulations that listeners with normal hearing asymptote their performance to approximately 80% correct or better at a resolution of twelve to twenty bands when different testing materials were used (Friesen, et al, 2001; Fu and Shannon, 1998). To study the effect of spectral overlapping of speech and noise, the speech in the six bands at the high frequency end were eliminated to retain a spectral region for noise alone. The processed IEEE sentences by this means resulted in an average intelligibility of 64.7% in quiet which set up a ideal baseline for further application of noise in the current study.

Furthermore, this wide range and the significant difference in understanding CI-simulated speech between listeners are consistent with findings from CI users (Nelson, et al, 2003), except that Nelson and colleagues reported a better performance range approximately between 65% and 90% correct in quiet conditions. This suggests that the range of speech understanding in the current study may simulate that of CI users in quiet listening conditions. The poorer performance from CI simulations in the current study has supposedly resulted from the elimination of spectral information from the six vocoder bands at the high frequency end.

2) Effect of spectral separation of noise and speech on speech understanding performance

The use of the lower 10 of 16 bands of the speech material in the current study is equivalent to low pass filtering the vocoded speech with a resolution of 16 bands at a cutoff frequency of 2149 Hz. Although, with the full spectrum available, CI simulation listeners have been reported to be able to achieve speech understanding scores in the quiet conditions close to that obtained from unprocessed speech (usually a perfect score), their understanding of the low passed vocoded speech is considerably impaired comparing to their understanding of low-pass unprocessed speech. The unpublished data from Nelson research lab testing normal-hearing listeners’ understanding of low pass filtered IEEE sentences at a cutoff frequency of 2000 Hz—similar to that was used in the current study—showed nearly perfect scores which were not different from the scores obtained using IEEE sentences with full spectrum. The low-pass filtering did not seem affect the intelligibility of speech. This notion is supported by Bashford and Warren's (1987) finding that normal-hearing listeners participated in the study were scored 98% or higher when listening to PB words and CID sentences both of which were low passed filtered at a cutoff frequency of 1100 Hz. The average score was 64.7% correct in the current study when the low pass filter with a cutoff frequency at 2149 Hz was applied. The performance is much poorer than the perfect score, which indicates that CI users' speech understanding is very vulnerable to spectral detriments—the spectral loss that does not affect speech understanding in normal-hearing listeners could substantially corrupt CI users' ability to understand speech.

The vulnerability of speech through CI simulations is further demonstrated by the reduced scores for speech understanding as the background noise increased spectral overlap with speech. The masking effect of background noise started when the vocoder bands carrying noise were still separate from those carrying speech, with vocoder bands 11 through 16 carrying noise while vocoder bands 1 through 10 carrying speech. As the bands of noise moved down to the lower frequency end while maintaining the same overall intensity across conditions, every condition with further band overlapping of speech and noise resulted in a significant decrease in speech understanding from the previous less-overlapping condition. This suggests that information from each vocoder band contributes so substantially to the overall speech intelligibility that, each time, masking one more band would take the listeners' understanding down significantly. This further suggests the importance of spectral information in CI users.

An interesting finding in the current study is that when background noise was carried by vocoder bands two bands separate from those carrying speech, listeners' performance did not differ from that in quiet condition. This may indicate that, despite of the importance of spectral cues in speech understanding through CI's, background noise through channels in a distance from channels of speech may not necessarily cast significant interference to speech understanding, although this distance is not necessarily one channel apart as found in the current study.

3) Effect of AM-rate of noise on speech understanding:

Temporal envelope is a crucial cue to CI users. Nevertheless the current study did not find an effect of modulation rate of noise on the interference of noise with speech understanding. It is noted, however, that the CI simulation listeners did not score higher when the background noise was modulated than when the background noise was unmodulated (i.e., steady-state noise). This finding is consistent with Nelson et al (2003) and Qin and Oxenham (2003) in that no masking release was recorded in either CI simulation listeners or CI users. The CI coding strategies lack transmission of temporal fine structure, thus pushing CI users to rely more on the temporal envelope cue. They have limited access to the redundant cues in other aspects typically available for normal-hearing listeners. Thus, the AM of noise may introduce an envelope interference that may lead to remarkable disruption in CI users' speech understanding while not affecting speech perception in normal-hearing listeners. This warrants further study which is beyond the scope of this paper.

However, the AM-rate difference—one of the temporal envelope cues—did not seem to facilitate CI simulation listeners to segregation the auditory stream of speech from that of noise. The listeners did not perform better when the AM-rate of noise was remote from a typical syllabic rate which is around 4 Hz. In fact, they performed similarly across all the AM rates of noise including 0 Hz (literally steady-state noise), 4 Hz, 8 Hz, 16 Hz, and 32 Hz. Nonetheless, the findings in Nelson et al (2003) suggest a potential effect of AM rate of noise—they reported a better speech understanding performance in both CI simulation listeners and CI users when the AM rates (1 Hz, 16 Hz, and 32 Hz) of background noise were remote from 4 Hz in contrast to that when the noise was unmodulated, although the difference did not reach a statistical significance. However, this trend was not revealed in the data of the current study. The different types of AM used in the two studies may be an account for the discrepancy. Nelson and colleagues used square AM, while the current study used sinusoidal AM. It is possible that the square AM at modulation rates around 4 Hz generates a perception similar to speech and hence creates more interference; whereas, the SAM noise does not generate as much of the similarity between the perception of noise and speech. The above postulated explanation was drawn from the description in Nelson et al about their listeners' perception of the similarity of square amplitude modulated noise and speech (page 967). However, this explanation needs to be further examined.

Overall, the current study implies that spectral separation is crucial for CI users to segregation interested speech from interference and that separation of the sinusoidal AM rate of noise from the typical syllabic rate did not facilitate CI users to segregate noise from speech.

III. Correlation between stream segregation and speech perception

Two cues—spectral separation and AM-rate separation—have been systematically investigated for their effect on stream segregation and speech understanding through CI simulations. The ability to segregate non-speech streams of noise based on either cue was also examined for its correlation with the skill to segregate the streams of speech and noise based on the same cue. The results revealed the following.

1) Speech understanding in both unmodulated and AM noise: When speech is presented in noise, regardless of the speech-noise spectral separation and the AM rate of noise, the ability to segregate non-speech streams based on

moderate spectral separation between A and B bursts (i.e., A7-12B1-8) may be used to predict speech understanding, provided both A and B bursts are amplitude modulated and the AM-rate separation of A and B bursts is two octaves or greater. This indicates that stream segregation based on the combination of spectral and AM-rate cues may be a predictor for speech understanding in noise. However, speech understanding was not correlated with conditions in the stream segregation experiment with band difference but without AM-rate difference (i.e., conditions of AM25-25) between A and B bursts. This suggests that stream segregation based on spectral cues alone didn't seem to be a predictor of speech understanding in noise through CI simulations in general.

However, no other inter-subsequence spectral separations than A7-12B1-8 were found to be a predictor for speech understanding in noise; neither was found a one-on-one correlation between speech perception in noise and stream segregation at a similar level of signal-interference spectral separation. There may be two reasons, both of which are related to the fact that speech-noise spectral separations didn't correspond to the A-B-burst spectral separations exactly.

a) In the speech experiment, the ten vocoder bands at the low-frequency end were remained for speech, while in the stream segregation experiment, eight vocoder bands at the low-frequency end were remained for the target subsequence—subsequence B. This small difference in the number of vocoder bands carrying the target stream may have resulted in that the spectral separations of target stream and interfering stream did not match exactly between the speech experiment and the stream segregation experiment.

b) There were five speech-noise spectral separations (i.e., N13-16Spch1-10, N11-16Spch1-10, N9-14Spch1-10, N7-12Spch1-10, and N5-10Spch1-10) in the speech perception experiment while only three A-B-burst spectral separations (i.e., A678B1-4, A456B1-4, and A345B1-4) were involved in the stream segregation experiment. This mismatched number of target-interference spectral separations may have missed the A-B-burst spectral separations that could be potential predictors for speech perception with certain speech-noise spectral separations.

2) Speech understanding in steady-state noise: CI simulations listeners' speech understanding in unmodulated noise may be predicted by the stream segregation ability based on both spectral cue and AM-rate cue, provided that the noise spectrum is not completely overlapping with the speech spectrum—the conditions of smallest speech-noise spectral separation—N5-10Spch1-10 (see rows 6 through 9 in Table 3). However, again, no consistent finding was seen that speech understanding with more spectral overlapping between speech and noise correlated with the stream segregation ability with larger band overlapping between A and B bursts.

It is noteworthy that the stream segregation based on spectral cues alone (i.e., when both A and B bursts carrying the same AM rate—25 Hz) was not a predictor for speech understanding in unmodulated noise through CI simulations. This indicates that CI simulation listeners may have been using AM-rate cue to separate speech and noise even when noise is unmodulated. This AM-rate difference between speech and unmodulated noise may arise from the comparison of the gross rate of temporal fluctuations in speech and the flattened temporal fluctuations on the steady-state noise.

Furthermore, when the noise spectrum is completely overlapping with the speech spectrum, the speech understanding in steady-state noise cannot be explained by stream segregation based on the either spectral cue, or AM-rate cue, or both. This may be because the spectral separation of speech and noise is so limited that the stream segregation ability based on larger spectral separations cannot be applied to predict speech understanding in these situations.

3) Speech understanding in AM noise: When speech was presented in AM noise, in most cases, speech perception correlated with stream segregation ability when the band separation of A and B bursts was at the medium level (i.e., A7-12B1-8). Speech perception with larger spectral separations between speech and noise (i.e., N13-16Spch1-10 and N11-16Spch1-10) also correlated with stream segregation ability with a larger band separation between A and B bursts (i.e., A11-16B1-8) as shown in rows ten and thirteen in Table 3, although speech

perception with smaller spectral separations between noise and speech (i.e., N567Spch1-5, N456Spch1-5, and N345Spch1-5) did not show correlation with stream segregation ability with a smaller band separation of A and B bursts (i.e., A5-10B1-8). It is noteworthy that the stream segregation ability in the spectral condition of A5-10B1-8 did not correlate with speech perception obtained with any spectral separation between speech and noise. This may be due to the poor stream segregation ability shown in the condition of A5-10B1-8 which was not better than that in the complete-band-overlapping condition (i.e., AbbnBbbn). The poor performance may not reflect individual listeners' stream segregation ability sufficiently to show the correlation between speech perception and stream segregation.

IV. Correlation between AM detection and speech perception

As envelope cues are critical in CI coding strategies, the ability to detect AM may affect speech understanding in CI simulation listeners. The current study found that the AM detection threshold at the rate of 25 Hz correlates with listeners' speech understanding in the presence of background noise, although this correlation disappeared when the noise spectrum was completely covered by the speech spectrum (i.e., N5-10Spch1-10).

CI users' speech perception has been reported to be correlated with AM detection (Cazals, Pelizzone, Saudan, et al, 1994; Fu, 2002). Cazals and colleagues measured TMTF (temporal modulation transfer function) in CI listeners and estimated the rejection strength of the low-pass filter of TMTF by subtracting the threshold at the AM rate of 400 Hz from that at the AM rate of 71 Hz. The rejection strength when transformed into the percentage of CI users' dynamic range showed a moderate correlation with CI users' identification of consonants and vowels, although no correlation was shown when the rejection strength was calculated in dB—a conventional approach for acoustic hearing. However, in contrast to the findings in the current study, Cazals, et al did not find that AM detection thresholds correlated with the identification of vowels and consonants.

On the other hand, Fu's research supported that AM detection thresholds correlate with speech understanding in CI users. Fu attributed the significance in his study as opposed to the insignificance in Cazals, et al (1994) to using the average thresholds of five to seven thresholds obtained at carrier levels covering a wide range from 10% to 90% of the CI user's dynamic range. Although the current study did not measure AM detection thresholds at carrier levels over a wide range, a correlation between speech understanding and AM detection thresholds was also found. This may be attributed to the fact that CI simulation listeners were tested. If there is a correlation, it might tend to manifest more consistently in CI simulation listeners; thus a carrier presented at an easily detectable level may be sufficient to demonstrate this relationship.

Additionally, the current study shows a correlation between speech understanding and the AM detection thresholds at the rate of 25 Hz in CI simulation listeners, although Fu's (2002) finding shows that the correlation was between the threshold at AM rate of 100 Hz and speech understanding. This may be due to the difference in the duration of the carrier from the two studies. The current study used 80-ms bursts to carry the AM while Fu used 300-ms pulse trains to carry the AM. With a carrier of 80 ms in length, an AM at 25 Hz only involves two modulation cycles, which makes the detection task more challenging compared with an AM at 100 Hz with an 80-ms carrier (which renders an eight modulation cycles). The detection in this difficult listening situation may be more reflective of AM cues listeners use in speech understanding, as a duration of 80 ms is close to the duration of formant transitions (Schwab, et al, 1981). Furthermore, Fu did not measure the AM detection thresholds at 25 Hz in the CI listeners. Therefore, whether the AM detection thresholds at 25 Hz would have been correlated with speech understanding in his study is inconclusive.

Generally, the current study shows that AM detection thresholds correlate with speech understanding in CI simulation listeners. This is supported by previous studies, regardless what AM detection feature (detection threshold or AM rejection strength) or what AM rate at which the detection threshold was measured is involved in the correlation. This provides support to the notion that AM cues are important to CI users.

Conclusions and Implications

Conclusions of the current study:

Overall, the current study demonstrates that stream segregation is measurable in CI simulation listeners when focused attention is directed toward segregation. In this schema-driven stream segregation, spectral separation between signals and noise provides an essential cue and AM-rate separation provides a facilitating cue on top of the spectral separation. When spectral separation is insufficient but not completely absent, the addition of AM-rate differences may improve the ability to segregate the stream of signals from that of noise; while the addition of AM-rate difference does not show a significant effect when the spectral separation is either sufficient to elicit strong stream segregation or completely absent.

The spectral separation between noise and speech affects speech perception significantly with better speech recognition resulted from larger spectral separations. In addition, whether the noise is amplitude modulated and whichever the AM rate is carried by the noise do not affect speech perception in noise, regardless of the spectral separation between speech and noise. However, speech perception in either modulated or unmodulated noise was found to be correlated with the ability to segregate streams of signal and noise based on both spectral separation and AM-rate separation between speech and noise. Furthermore, speech perception in noise was also found correlated with AM detection threshold at the AM rate of 25 Hz. In other words, AM differences between speech and noise did not aid speech understanding. Further studies could focus on whether listeners could learn to use AM differences for understanding speech in noise.

Implications for CI's:

In acoustic hearing, spectral and temporal cues have been concluded to provide redundancy in speech recognition (Schroeder, 1966). Auditory signals through CI's are much degraded in spectral differences (Dorman, et al, 1998; Friensen, et al, 2001). The current available coding strategies extracting temporal envelope cues do not transmit the temporal fine structure information to CI users. On the other hand, near-normal AM detection ability has been documented in CI users (Shannon, 1992). The design of the experiment on stream segregation the current study simulated CI in a way such that the spectral cues are smeared but the AM is intact. The findings from normal-hearing listeners listening to the signals through CI simulations may imply that CI users could: a) possibly formulate stream segregation when the two excited channels—each comprising of a number of consecutive electrodes—are widely separated or moderately overlapping even if the temporal envelope cues are not available; b) possibly not to formulate stream segregation when the excited channels are largely overlapping to a certain degree when temporal envelope cues are not available; c) possibly formulate stream segregation when the signals from the two channels are amplitude modulated at two largely-different rates; d) unlikely to formulate stream segregation when the excited channels are completely overlapping, even when the signals from the two channels are amplitude modulated at two largely separate rates, like, with a 2-octave separation.

The above summarized implications suggest that CI users could potentially use the AM-rate cue to segregate streams of speech and noise, although this cue is weak and the spectral difference is the dominant cue for speech recognition. The AM-rate difference will facilitate CI users to segregation streams of signals and noise when the spectral difference between the two is limited. CI users' speech recognition also correlates with their AM detection ability. Consequently, it is implied that, although spectral cues are crucial for speech recognition in CI users, AM cues may be as important to CI users since their spectral resolution is extremely restricted due to the limitation from processing strategies. This casts an implication to processing strategies that enhancement of coding AM may potentially improve CI users' ability to segregate interleaved ongoing perceptual streams with the limited spectral channels currently available to them.

Table 1. The cutoff frequencies for the bandpass noises associated with that of the 16 vocoder bands from Fu and Nogaki (2004).

Greater-band-overlapping condition					**A bursts**: Bands5-10											
Moderate-band-overlapping condition							**A bursts**: Bands 7-12									
No-band-overlapping condition											**A bursts**: Bands 11-16					
	B bursts: Bands1-8															
Vocoder band (16-band resolution)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Low cutoff frequency (Hz)	200	272	359	464	591	745	931	1155	1426	1753	2149	2627	3205	3904	4748	5768
High cutoff frequency (Hz)	272	359	464	591	745	931	1155	1426	1753	2149	2627	3205	3904	4748	5768	7000

Table 2. Amplitude modulation detection thresholds of individual participants at three different modulation rates. Results are shown in average across the runs plus/minus one standard deviation. Averages and standard deviations across participants are also shown in the table

	Amplitude Modulation Rate		
Participant	**25 Hz (Mean ± SD)**	**100 Hz (Mean ± SD)**	**300 Hz (Mean ± SD)**
S1	-11.11±1.29	-14.50±1.95	-14.88±1.68
S2	-13.4225±2.59	-17.55±0.70	-17.44±2.10
S3	-12.11±2.13	-23.00±2.91	-12.89±1.53
S4	-8.04±0.87	-13.44±3.78	-9.81±1.53
S6	-14.07±1.45	-18.54±2.77	-12.16±2.52
S7	-9.03±1.58	-16.86±3.15	-13.61±4.81
S8	-10.61±2.43	-17.11±1.12	-11.77±1.44
S9	-12.44±1.34	-17.39±1.88	-11.33±0.48
S10	-12.58±2.42	-15.48±1.65	-7.91±0.78
S11	-12.29±0.80	-15.05±1.08	-13.05±1.85
Average ± SE	**-11.57±1.89**	**-16.89±2.66**	**-12.49±2.62**

Table 3. Individual speech understanding scores (% keywords correctly identified in the Speech Perception Experiment

Spectral (Band) Condition	**AM-rate of Noise (Hz)**	**Participant (% keywords correctly identified)**										**Average (%)**	**Standard Error (%)**
		S1	S2	S3	S4	S6	S7	S8	S9	S10	S11		
Quiet		80	72	62	55	58	63	74	67	63	53	64.7	8.6
N78	0	68	68	57	37	62	54	49	59	64	65	58.3	9.7
	4	68	58	65	38	63	65	62	61	53	52	58.5	8.9
	16	75	74	73	43	67	53	50	72	67	64	63.8	11.3
	32	79	71	65	35	63	41	64	57	64	61	60.0	13.1
	64	85	68	63	44	62	46	64	55	56	72	61.5	12.2
Average (SD)		75.0 (7.3)	67.8 (6.0)	64.6 (5.7)	39.4 (3.9)	63.4 (2.1)	51.8 (9.1)	57.8 (7.6)	60.8 (6.6)	60.8 (6.0)	62.8 (7.3)		
N678	0	70	59	56	22	64	60	41	49	53	62	53.6	13.8
	4	67	76	39	24	43	43	54	65	56	60	52.7	15.5
	16	59	67	52	24	58	29	61	64	29	57	50.0	16.2
	32	55	48	60	38	66	46	49	42	49	45	49.8	8.4
	64	67	53	58	36	69	42	41	49	45	65	52.5	11.8
Average (SD)		63.6 (6.3)	60.6 (11.1)	53.0 (8.4)	28.8 (7.6)	60.0 (10.3)	44.0 (11.1)	49.2 (8.6)	53.8 (10.2)	46.4 (10.6)	57.8 (7.7)	63.6 (6.3)	
N567	0	63	53	56	26	37	37	36	53	65	45	47.1	12.9
	4	48	42	59	16	44	19	30	32	50	53	39.3	14.5
	16	46	53	52	29	46	37	49	41	37	75	46.5	12.5
	32	43	54	44	25	57	54	41	43	34	53	44.8	10.1
	64	45	30	53	27	43	24	54	27	42	54	39.9	12.0
Average (SD)		49.0 (8.0)	46.4 (10.4)	52.8 (5.6)	24.6 (5.0)	45.4 (7.3)	34.2 (13.6)	42.0 (9.7)	39.2 (10.1)	45.6 (12.4)	56.0 (11.2)		

Table 3. Individual speech understanding scores (% keywords correctly identified in the Speech Perception Experiment (continued).

Spectral (Band) Condition	**AM-rate of Noise (Hz)**	**Participant (% keywords correctly identified)**										**Average (%)**	**Standard Error (%)**
		S1	S2	S3	S4	S6	S7	S8	S9	S10	S11		
N456	0	43	30	44	17	45	30	50	35	38	49	38.1	10.3
	4	44	36	40	21	42	29	43	24	39	39	35.7	8.2
	16	31	44	41	26	39	28	40	37	22	37	34.5	7.3
	32	26	50	38	26	35	28	31	31	38	36	33.9	7.3
	64	34	40	40	18	36	34	49	26	42	41	36.0	8.8
Average (SD)		35.6 (7.8)	40.0 (7.6)	40.6 (2.2)	21.6 (4.3)	39.4 (4.2)	29.8 (2.5)	42.6 (7.7)	30.6 (5.6)	35.8 (7.9)	40.4 (5.2)		
N345	0	24	35	34	15	48	24	33	17	27	54	31.1	12.5
	4	26	30	51	10	17	23	24	20	17	26	24.4	11.0
	16	46	51	36	14	37	14	36	32	31	24	32.1	12.1
	32	38	23	29	22	28	21	40	28	14	37	28.0	8.4
	64	43	45	32	12	18	35	30	11	17	36	27.9	12.5
Average (SD)		35.4 (9.9)	36.8 (11.3)	36.4 (8.6)	14.6 (4.6)	29.6 (13.1)	23.4 (7.6)	32.6 (6.1)	21.6 (8.4)	21.2 (7.4)	35.4 (11.9)		

Table 4. Statistics of regression estimates with linear equation between speech perception scores and stream segregation ability.

Row	Condition of Speech Perception Experiment (Dependent variable denoted as Y)	Condition in Stream Segregation Experiment (Independent variable denoted as X)	Linear Equation (standardized coefficients)	Adjusted R square	Significance level (p value)
1	N13-16	A7-12B1-8/AM25-300	Y=0.901X	0.788	<0.001
2	N11-16	A7-12B1-8/AM25-300	Y=0.785X	0.568	0.007
3	N9-14	A7-12B1-8/AM25-100	Y=0.91X	0.806	<0.001
4	N7-12	A7-12B1-8AM25-100	Y=0.788X	0.573	0.007
5	N5-10	A7-12B1-8/AM25-100 (X1) AbbnBbbn/AM25-300 (X2): *negative correlation* A345AM25-300 (X3)	Y=1.119X1-0.948X2+0.463X3	0.832	0.003
6	N13-16Spch1-10/N-AM0	A11-16B1-8/AM25-300	Y=0.774X	.0.55	.009
7	N11-16Spch1-10/N-AM0	A7-12B1-8/AM25-300	Y=0.701X	.428	.024
8	N9-14Spch1-10/N-AM0	A11-16B1-8/AM25-300	Y=0.934X	.856	<.001
9	N7-12Spch1-10/N-AM0	A7-12B1-8/AM25-100	Y=0.786X	.570	.007
10	N13-16Spch1-10/N-AM16	A11-16B1-8/AM25-25	Y=0.851X	.690	.002
11	N13-16Spch1-10/N-AM32	A7-12B1-8/AM25-100	Y=.880X	.746	.001
12	N13-16Spch1-10/N-AM64	A7-12B1-8/AM25-100	Y=0.829X	.648	.003
13	N11-16Spch1-10/N-AM4	A11-16B1-8/AM25-25	Y=0.171*X	.443	.021

Table 4. Statistics of regression estimates with linear equation between speech perception scores and stream segregation ability (Continued).

Row	Condition of Speech Perception Experiment (Dependent variable denoted as Y)	Condition in Stream Segregation Experiment (Independent variable denoted as X)	Linear Equation (standardized coefficients)	Adjusted R square	Significance level (p value)
14	N11-16Spch1-10/N-A64	A7-12B108/AM25-100	Y=0.767X	.537	.010
15	N9-14Spch1-10/N-AM4	A7-12B1-8/AM25-100 (X1) A11-16B1-8/AM25-100 (X2)	Y=0.728X1+0318X2	.909	.000
16	N9-14Spch1-10/N-AM16	A7-12B1-8/AM25-100	Y=0.636X	0.330	.048
17	N9-14Spch1-10/N-AM64	A7-12B1-8/AM25-100	Y=.811X	.614	.004
18	N7-12Spch1-10/N-AM4	A7-12B1-8/AM25-100 (X1) A11-16B1-8/AM25-25 (X2): *negative correlation*	Y=1.174X1-.504X2	.782	.002
19	N5-10Spch1-10/N-AM4	A7-12B1-8/AM25-300	Y=0.709X	.441	.022
20	N5-10B1-10/N-AM16	A7-12B1-8/AM25-100	Y=0.693X	.416	.026

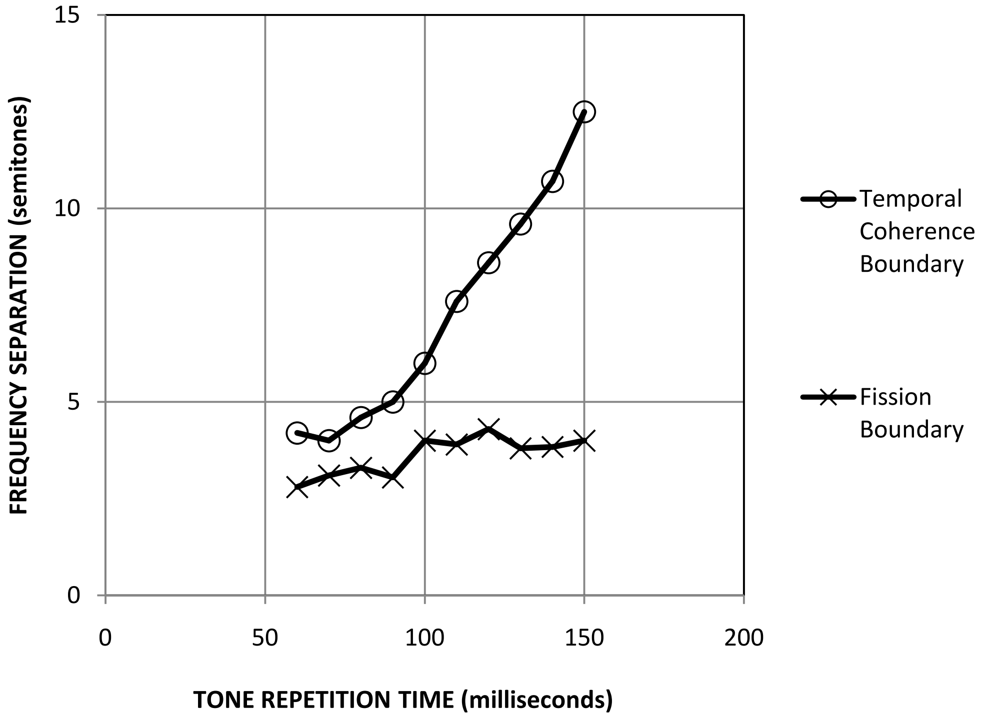

Figure 1. The Fission Boundary (FB) and the Temporal Coherence Boundary (TCB) for auditory stream segregation in van Noorden's unpublished dissertation "Temporal coherence in the perception of tone sequences". (Adapted from van Noorden, 1975)

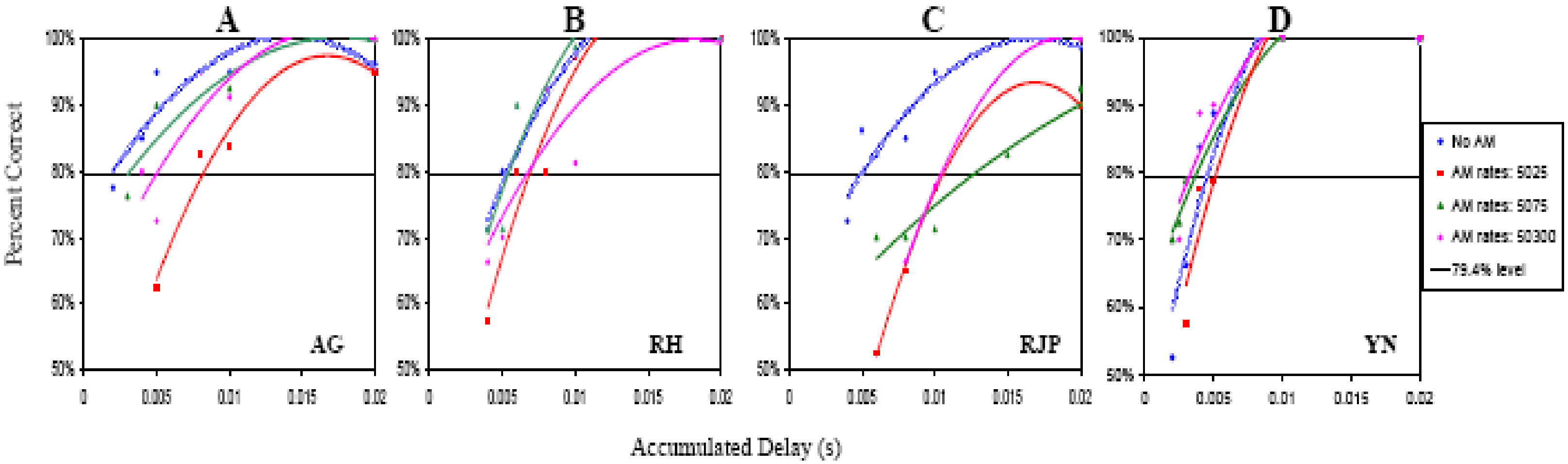

Figure 2. Findings supporting that slow AM rate as 25 Hz can be used for stream segregation experiments. (Adopted from Nie and Nelson, 2007).

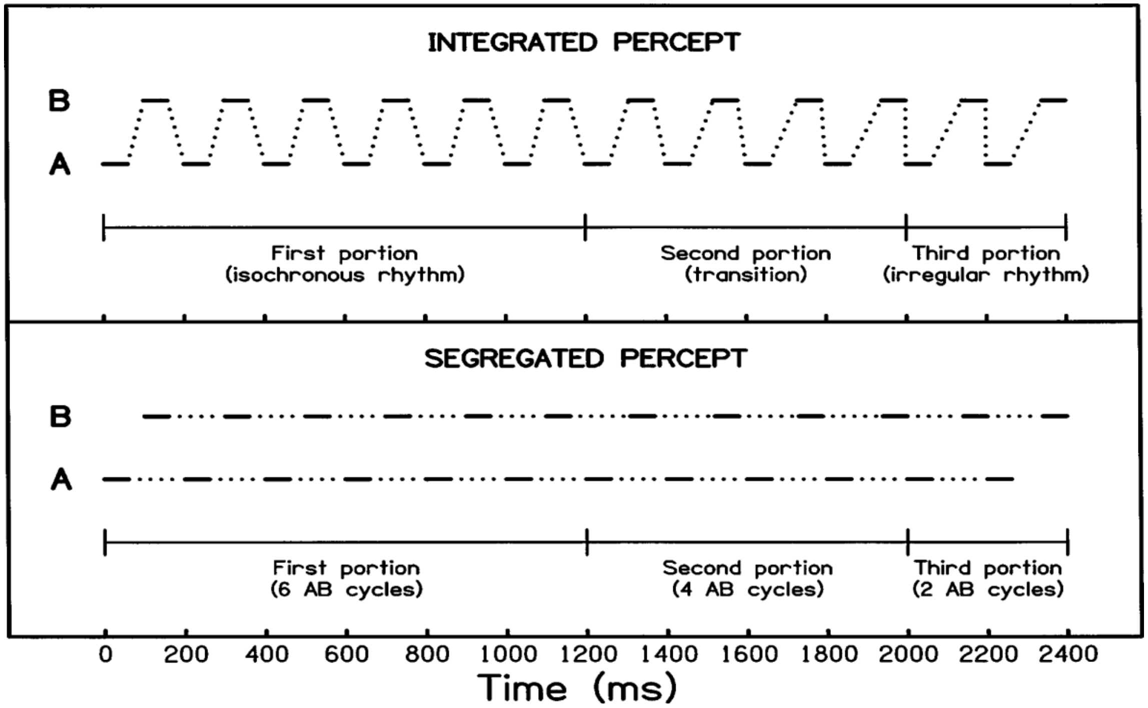

Figure 3. The schematic plot showing the design of Roberts et al study.

Reprinted with permission from ROBERTS, B., GLASBERG, B. R., MOORE, B. C. J., PRIMITIVE STREAM SEGREGATION OF TONE SEQUENCES WITHOUT DIFFERENCES IN FUNDAMENTAL FREQUENCY OR PASSBAND, 112, 2074, (2002). Copyright 2002, Acoustical Society of America.

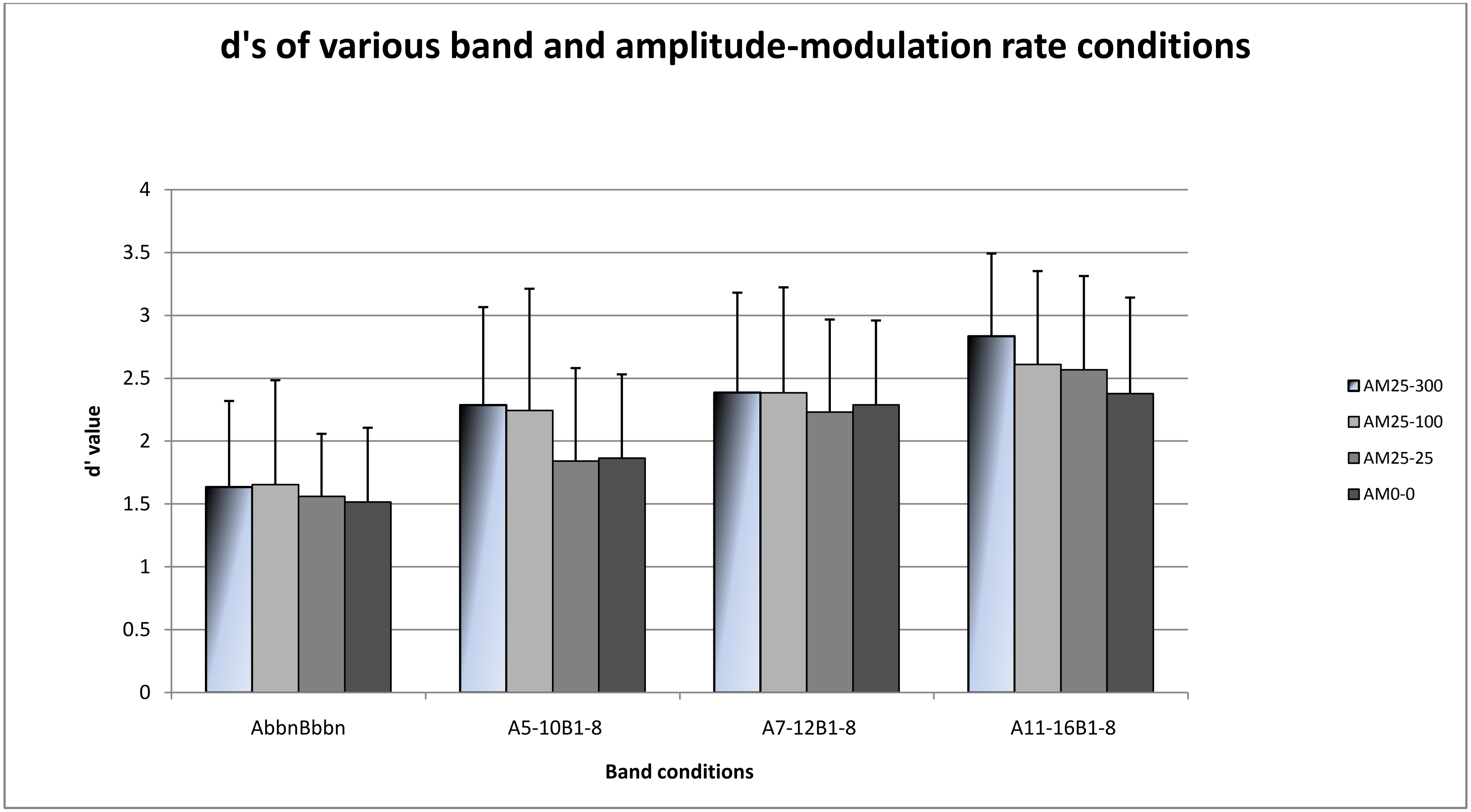

Figure 4. d’s in various band and AM rate conditions of the 12-pair stimuli sequences in the auditory stream segregation portion in the psychophysical experiment (error bars depict one standard deviations above the mean).

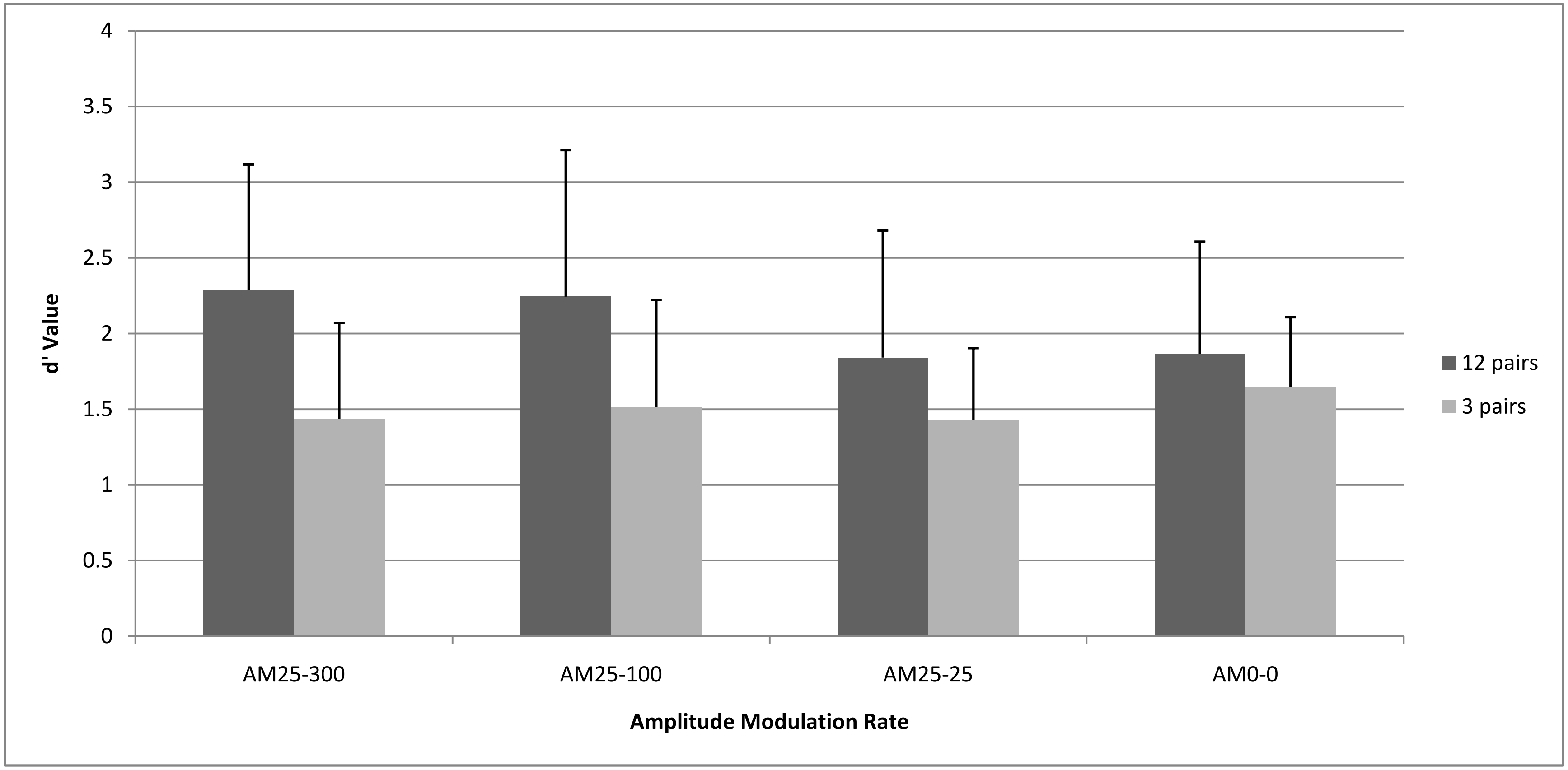

Figure 5. Comparison of d’s of 12-pair and 3-pair stimulus sequences in the greater overlapping band condition (A5-10B1-8) in the auditory stream segregation portion in the psychophysical experiment (error bars depict one standard deviations above the mean).

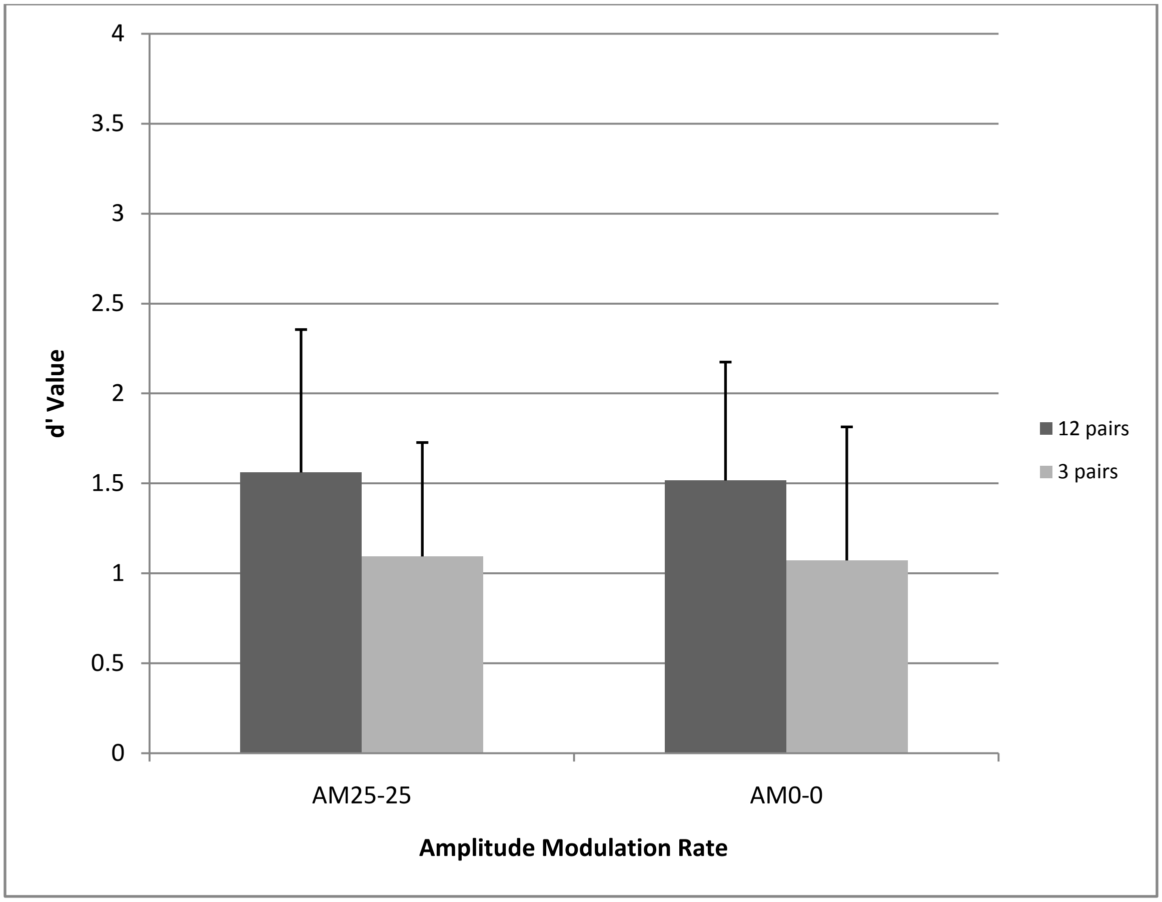

Figure 6. Comparison of d’s of 12-pair and 3-pair stimulus sequences in the completely overlapping band condition (AbbnBbbn) in the auditory stream segregation portion of the psychophysical experiment (error bars depict one standard deviations above the mean).

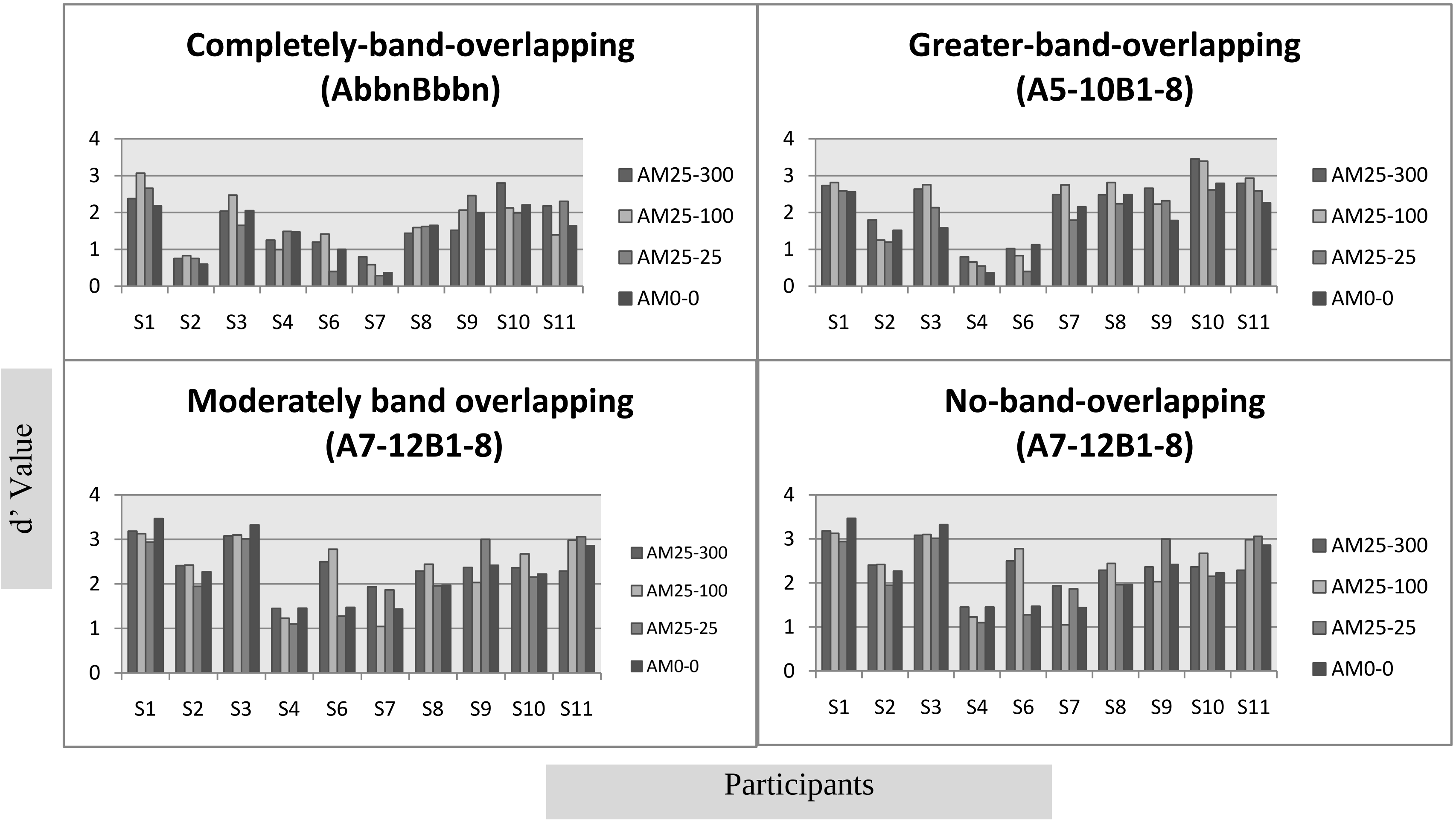

Figure 7. Individual d's of various band conditions and AM rate conditions for 12-pair stimulus sequences in the auditory stream segregation portion of psychophysical experiment.

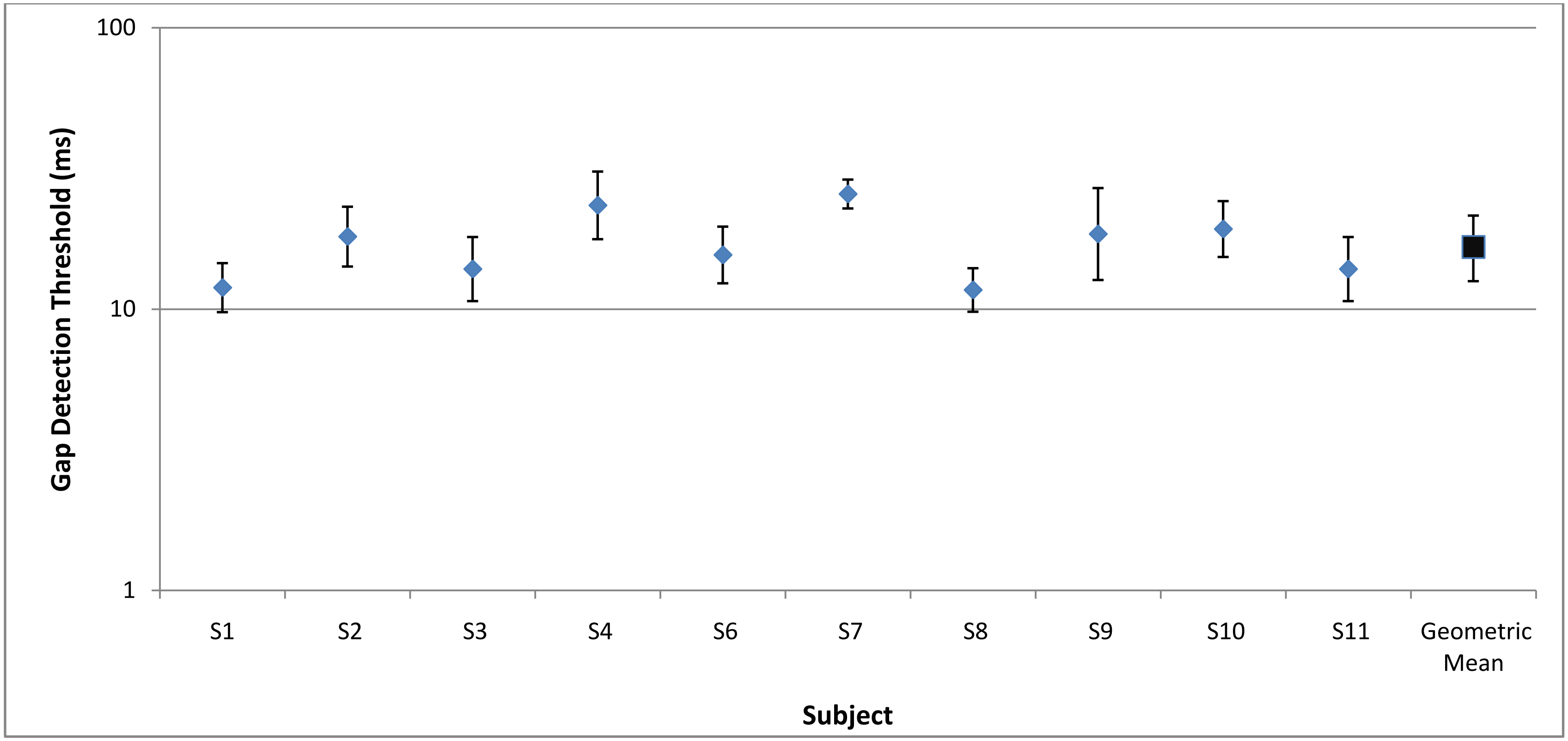

Figure 8. Gap-delay detection threshold: thresholds and plus/minus one standard deviations are shown for each of the ten participants. The filled square shows the geometric mean of the ten individual thresholds; the y-error bars show one standard deviations around the geometric means.

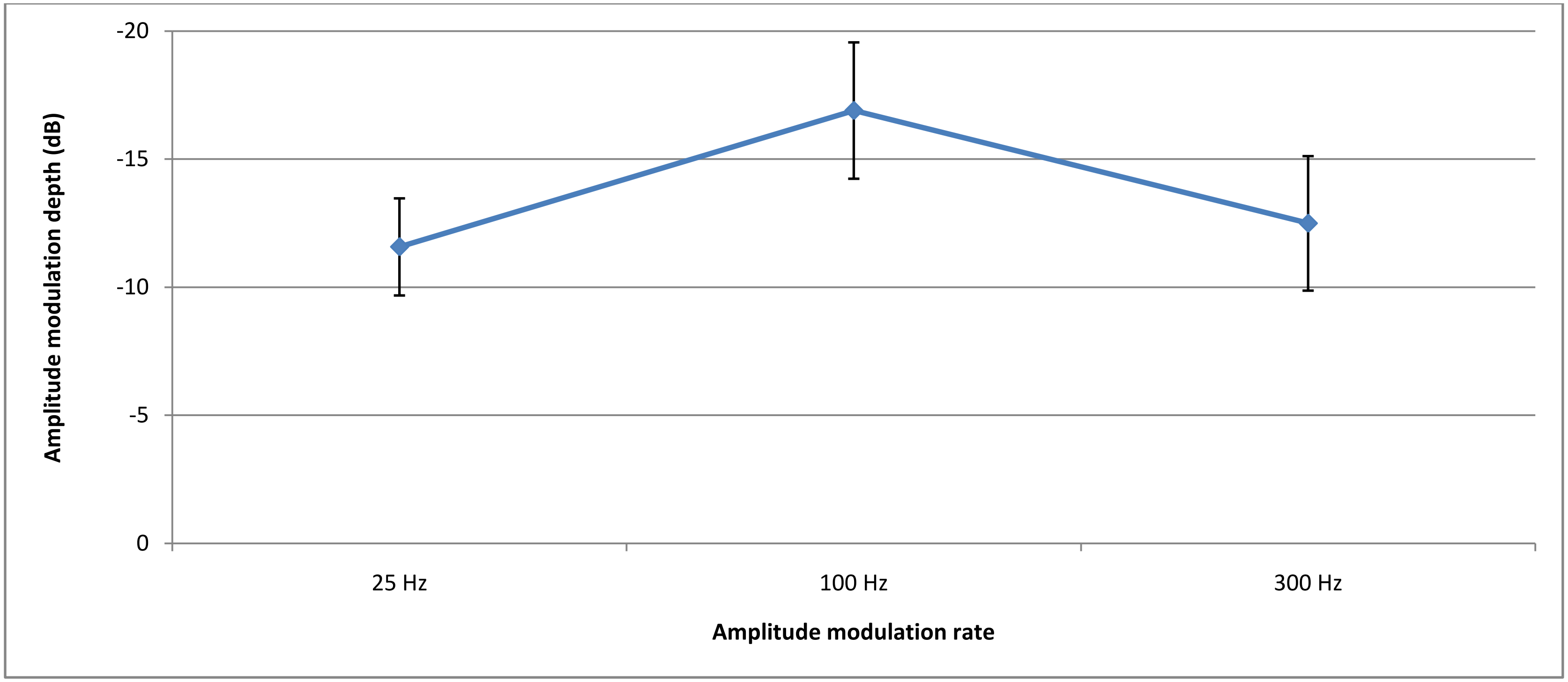

Figure 9. Amplitude modulation detection thresholds. Error bars show one standard deviations above and below the means.

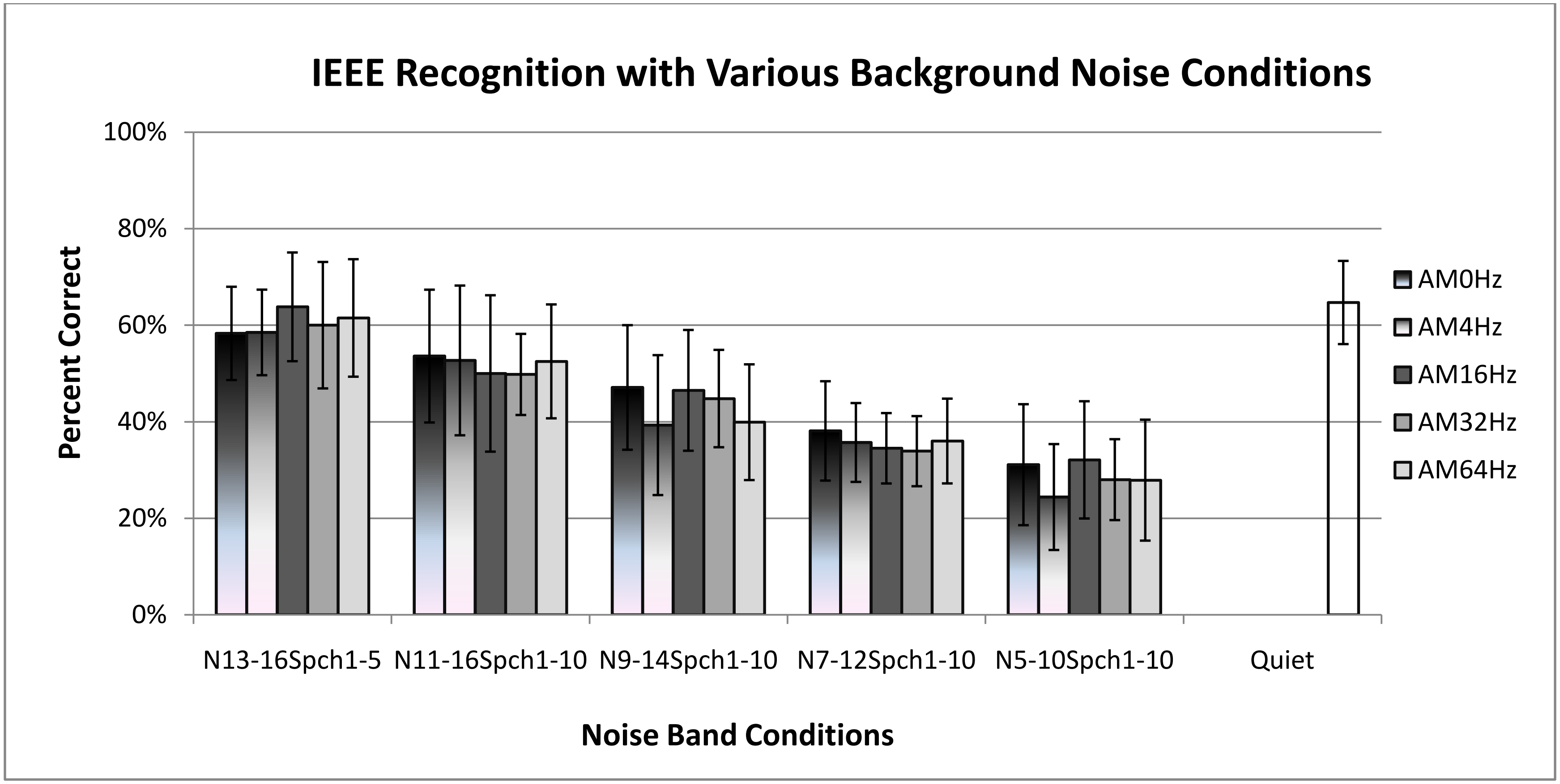

Figure 10. Understanding of sentence keywords through cochlear implant simulations. Y-error bars show one standard deviation above and below the mean.

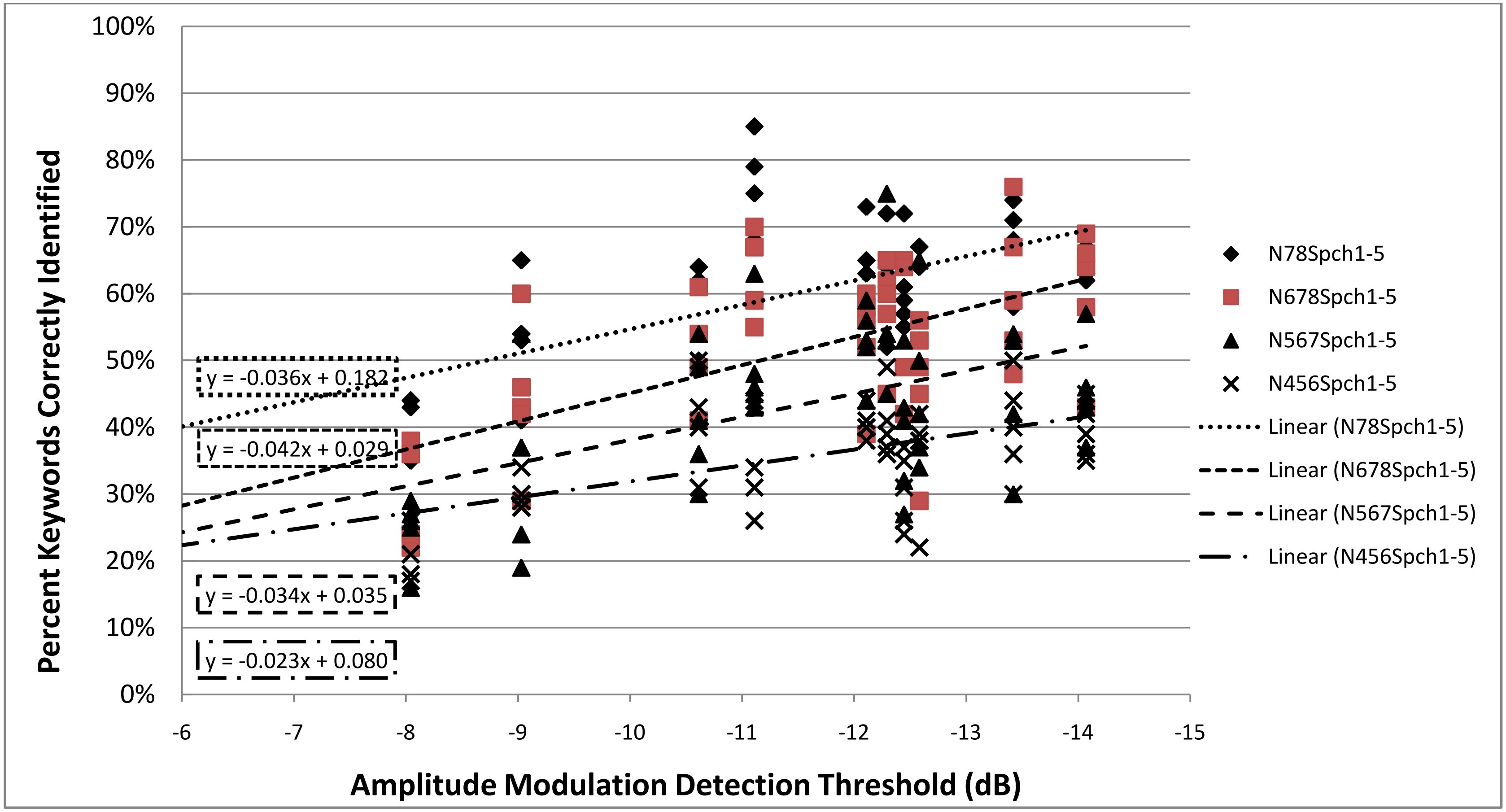

Figure 11. Percent keywords of speech in noise correctly identified as a function of amplitude modulation detection threshold at the modulation rate of 25 Hz. The regression linear equations are surrounded by lines in the styles corresponding to the speech perception experimental conditions denoted in the legend.

REFERENCES

Anderson, E.S., Nelson, P.B., Kreft, H.A., & Nelson, D.A. (Submitted).Spatial tuning curves and speech perception in cochlear implant users.

Bashford, J.A. and Warren, R.M. (1987). Effects of spectral alternation on the intelligibility of words and sentences. Perception & Psychophysics. 42 (5), 431-438.

Botte, M-C, Drake, C., and Brochard, R. (1997). Perceptual attention of nonfocused auditory streams. Perception & Psychophysics. 59 (3), 419-425.

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.

Bregman, A.S. (1978). Auditory streaming is cumulative. Journal of Experimental Psychology: Human Perception and Performance. 4, 380-387.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244-249.

Bregman, A.S., Liao, C, and Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. Canadian Journal of Psychology/Revue canadienne de psychologie. 44, 400-413.

Bregman, A.S. and Rudnicky, A. (1975). Auditory segregation: Stream or streams? Journal of Experimental Psychology: Human Perception and Performance, 1, 263-267.

Brochard, R., Drake, C., Botte, M-C, and McAdams, S.(1990). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. Journal of Experimental Psychology: Human Perception and Performance. 25 (6), 1742-1759.

Carlyon, Robert P., Cusack, Rhodri, and Foxton, Jessica M. (2001). Effects of attention and unilateral neglect on auditory stream segregation. 27 (1), 115-127.

Cazals, Y., Pelizzone, and M., Saudan, O., et al. (1994). Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants. J. Acoust. Soc. Am. 96, 2048-2054.

Chatterjee, M., Galvin, J.J., 2002. Auditory streaming in cochlear implant listeners. Journal of the Acoustical Society of America. 111, 2429.

Chatterjee, Monita, Sarampalis, Anastasios, and Oba, Sandra I. (2006). Auditory stream segregation with cochlear implants: A preliminary report. Hearing Research, 222, 100-107.

Cooper, H. R. and Roberts, B. (2007). Auditory stream segregation of tone sequences in cochlear implant listeners. Hearing Research, 225, 11-24.

Dannenbring, G.L. and Bregman, A.S. (1976). Stream Segregation and the Illusion of Overlap. *Journal ol Experimental Psychology: Human Perception and Performance.* 2, 544-555

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), Hearing (Vol. 6, pp. 387-424). Orlando, FL: Academic.

Dau, T., Verhey, J., and Kohlrausch, A. (1999). “Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers,” Journal of the Acoustical Society of America. 106, 2752-2760.

Dorman, M.T., Loizou, P.C, Fitzke, J., Tu, Z (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6 – 20 channels. Journal of the Acoustical Society of America. 104, 3583-3585.

Dowling, W.J. (1973). The perception of interleaved melodies. Cognitive Psychology, 5. 322-337.

Ewert, S.D. and Dau, T. (2000). “Characterizing frequency selectivity for envelope fluctuations,” Journal of the Acoustical Society of America. 108, 1181-1196.

Friesen, L.M., Shannon, R.V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. Journal of the Acoustic Society of America. *110*, 1150-1163.

Fu, Q-J. (2002). Temporal processing and speech recognition in cochlear implant users. Neuroreport, 13. 1635-1639.

Fu, QJ and Nogaki, G. (2004). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. Journal of the Association for Research in Otolaryngology. 6, 19-27.

Fu, Q.J., Shannon, R.V., and Wang, X. (1998). Effects of noise and spectral resolution an vowel and consonant recognition: Acoustic and electric hearing. *Journal of the Acoustic Society of America. 104*, 3586-3596

Gaudrain, E., Grimault, N., Healy, E., et al. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. Journal of the Acoustical Society of Americas. 124. 3076-3087.

Glasberg, B. R. and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data, Hearing Research. 47, 103-138.

Glasberg, B. R., and Moore, B. C. J. (2006). "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model," J. Acoust. Soc. Am. 120, 585-588.

Grimault, Nicolas, Bacon, Sid P., and Micheyl, Christophe. (2002). Auditory stream segregation on the basis of amplitude modulation rate. Journal of the Acoustical Society of America. 111 (3), 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. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. Journal of the Acoustical Society of America. 120 (1), 360-374.

Hong, R.S. and Turner, C. (2009). Sequential stream segregation using temporal periodicity cues in cochlear implant recipients. Journal of the Acoustical Society of America. 126, 291-299.

Jin, SH and Nelson, PB (2006). Speech perception in gated noise: The effects of temporal resolution. Journal of the Acoustical Society of America. 119(5), 3097-3108.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). A model for the prediction of thresholds, loudness and partial loudness. J. Audio Eng. Soc., 45, 224-240.

Nelson, PB. and Jin, SH. (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. Journal of the Acoustical Society of America.115 (5), 2286-2294.

Nelson, P.B., Jin, S-H, Carney, A.E., et al., (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. Journal of the Acoustical Society of America, 113. 961-968.

Nie, Y. and Nelson, PB (2007). Auditory stream segregation in amplitude modulated bandpass noise. Conference on Implantable Auditory Prosthetics. Lake Tahoe, CA. July 16-20.

Qin, MK and Oxenham, AJ. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. Journal of the Acoustical Society of America. 114 (1), 446-454.

Roberts, B., Glasberg, B., and Moore, B. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. Journal of the Acoustical Society of America. 112 (5). 2074-2085.

Schroeder, M.R. 1966. Vocoders: Analysis and synthesis of speech. Proceedings of the IEEE. 54, 5 (May), 720- 734.

Schwab, E. C., Sawusch, J. R.., and Nusbaum, H.C. (1981). The role of second formant transitions in the stop-semivowel distinction. Perception & Psychophysics. 29, 121-128.

Shannon, R.V. (1992). Temporal modulation transfer functions in patients with cochlear implants. Journal of the Acoustical Society of America. 91, 2156-2164.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science. 270, 303-304.

Stainsby, T.H., Moore, B.C.J., Medland, P.J., et al. (2004).Sequential streaming and effective level differences due to phase-spectrum manipulations. Journal of the Acoustical Society of America. 115. 1665–1673

Stickney, G.S., Zeng, F.G., Litovsky, R., and Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America. 116*, 1081-1091.

Sussman, E.S., Bregman, A.S., Wang, W.J., and Khan, F.J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cognitive, Affective, and Behavioral Neuroscience. 5 (1), 93-110.

Sussman, E., Ritter, W. and Vaughan, HG. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. Brain Research. 789, 130-138.

Sussman, E. [S.], Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology, 36, 22-34.

Sussman, E. [S.], Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. Cognitive Brain Research, 13, 393-405.

van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. Unpublished doctoral dissertation, Eindhoven University of Technology.

Viemeister, N.F. (1979). Temporal modulation transfer functions based upon modulation thresholds. Journal of the Acoustical Society of America. 66. 1364-1380

Zwicker, E. and Fastl, H.(1990). (E.ds.), Psychoacoustics. Berlin, Heidelberg, New York: Springer-Verlag,