Amplitude fluctuations in a masker influence lexical segmentation in cochlear implant users^{a)}

Trevor T. Perry^{b)} and Bomjun J. Kwon

Department of Hearing, Speech, and Language Sciences, Gallaudet University, Washington, DC 20002

(Received 25 September 2014; revised 12 March 2015; accepted 18 March 2015)

Normal-hearing listeners show masking release, or better speech understanding in a fluctuating-amplitude masker than in a steady-amplitude masker, but most cochlear implant (CI) users consistently show little or no masking release even in artificial conditions where masking release is highly anticipated. The current study examined the hypothesis that the reduced or absent masking release in CI users is due to disruption of linguistic segmentation cues. Eleven CI subjects completed a sentence keyword identification task in a steady masker and a fluctuating masker with dips timed to increase speech availability. Lexical boundary errors in their responses were categorized as consistent or inconsistent with the use of the metrical segmentation strategy (MSS). Subjects who demonstrated masking release showed greater adherence to the MSS in the fluctuating masker compared to subjects who showed little or no masking release, while both groups used metrical segmentation cues similarly in the steady masker. Based on the characteristics of the segmentation cues, the results are interpreted as evidence that CI listeners showing little or no masking release are not reliably segregating speech from competing sounds, further suggesting that one challenge faced by CI users listening in noisy environments is a reduction of reliable segmentation cues. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4916698>]

[TB]

Pages: 2070–2079

I. INTRODUCTION

Cochlear implants (CI) provide access to auditory sensations for individuals who have severe to profound sensorineural hearing loss and who show little or no benefit from hearing aids. Modern CI systems convert incoming sound signals into patterns of electrical pulses which are delivered to the auditory nerve via an electrode array implanted in the cochlear canal. When speech is presented in a quiet environment, many CI recipients can achieve high levels of speech understanding by interpreting the patterns of electrical current delivered by the device. CI users are highly susceptible to the presence of competing sounds, however, and their ability to understand speech degrades sharply with increasing levels of background noise (e.g., [Fu et al., 1998](#); [Friesen et al., 2001](#); [Nelson et al., 2003](#); [Fu and Nogaki, 2005](#)).

Normal-hearing (NH) listeners show more robust performance when listening to speech in noisy environments. In particular, for speech and noise presented at a given signal-to-noise ratio (SNR), NH listeners show greater speech understanding when the amplitude of the noise is fluctuating than when its amplitude is steady (e.g., [Wilson and Carhart, 1969](#); [Festen and Plomp, 1990](#)). This phenomenon of improved performance when the masker's amplitude is fluctuating is sometimes referred to as masking release, fluctuating masker benefit, or "listening in the dips." Compared to

the performance of NH listeners, hearing impaired (HI) listeners show less benefit from masker fluctuations ([Wilson and Carhart, 1969](#); [Festen and Plomp, 1990](#); [Takahashi and Bacon, 1992](#); [Eisenberg et al., 1995](#); [Bacon et al., 1998](#); [Bernstein and Grant, 2009](#)), and CI listeners demonstrate little to no masking release ([Nelson et al., 2003](#); [Stickney et al., 2004](#); [Loizou et al., 2009](#)) and may even demonstrate poorer performance in a fluctuating condition than in a steady condition, a pattern which has been called modulation interference ([Kwon and Turner, 2001](#)).

Several proposed explanations for the reduced or absent masking release in CI listeners have been examined in the literature (for a review, see [Kwon et al., 2012](#)). For example, one proposed explanation is that excessive forward masking in HI and CI listeners ([Glasberg et al., 1987](#)) decreases the intelligibility of speech information in the dips of a masker. However, data from HI and CI listeners suggest that the influence of forward masking cannot explain a majority of the reduction in masking release ([Nelson et al., 2003](#); [Jin and Nelson, 2006, 2010](#)). Another proposed explanation is that the degree of masking release depends on the SNR used for testing, and the difference in masking release between NH and HI listeners might be due to differences in the SNRs at which stimuli were presented ([Bernstein and Grant, 2009](#); [Oxenham and Simonson, 2009](#)). However, this explanation might not account for all of the observed reduction in masking release, and might not be applicable for certain masker configurations ([Christiansen and Dau, 2012](#); [Kwon et al., 2012](#)). Several investigations have found that CI listeners experience greater negative effects when masker fluctuations occur at rates ~4 Hz, similar to the syllabic rate of speech ([Nelson et al., 2003](#); [Jin et al., 2013](#)). These data, showing

^{a)}A portion of this study was presented at the 16th Conference on Implantable Auditory Prostheses in Lake Tahoe, California, July 2013.

^{b)}Author to whom correspondence should be addressed. Current address: Department of Speech-Language-Hearing Sciences, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455. Electronic mail: trevortperry@gmail.com

more negative effects when masker modulations are speech-like, can be interpreted roughly as perceptual masking (e.g., Carhart *et al.*, 1969) or informational masking (e.g., Durlach *et al.*, 2003). In this interpretation, speech and masker are fused rather than segregated, which reduces speech understanding due to difficulty in ignoring the masker while attending to the speech. Such fusion might be primarily due to the signal processing and delivery method of the CI system which gives CI users fewer salient cues for segregation. Another possible explanation is a deficit in CI users' ability to integrate speech information across time periods fragmented by the amplitude fluctuations of a masker. This explanation is supported by reports that CI listeners and NH listeners presented with CI simulations show substantial difficulty understanding temporally interrupted speech, in which the speech signal is periodically replaced with silence (Nelson and Jin, 2004; Chatterjee *et al.*, 2010; see also Gaudrain and Carlyon, 2013). Finally, Oxenham and Kreft (2014) theorized that possible differences between CI and NH listeners in the perception of the amplitude fluctuations inherent in notionally "steady" noise could explain the lack of masking release in CI listeners. In summary, there is still on-going debate about the mechanism(s) underlying the reduced or absent masking release in HI and CI listeners.

Investigations of masking release in the literature typically have generated fluctuating-amplitude maskers by using speech from one or more competing talkers or modifying a noise signal using one of the following general techniques: gating the amplitude of noise on and off, modulating the amplitude of noise according to a sinusoid, or imposing the temporal envelope of a speech signal onto noise. When the resulting masker is combined with a speech target, the speech is masked incidentally; in other words, the timings of dips in the masker are unrelated to the timings of amplitude fluctuations in the speech target and temporal interactions between the two are not deliberately controlled. In a previous study, we investigated the influence of masker fluctuations on speech understanding when the temporal overlap between speech and masker is carefully controlled rather than incidental (Kwon *et al.*, 2012). In this paradigm, we conceived of two temporal arrangements of speech and masker. In one, the temporal overlap between speech and noise energy was purposefully made small in order to promote masking release. We called this approach "+MR"

("plus MR"). In the other arrangement, noise energy is made to occur primarily during moments when speech energy is present. We called this "-MR" ("minus MR") as it should suppress masking release. The +MR and -MR labels refer to the effect that the temporal arrangement of speech and masker is anticipated to have on masking release: Any arrangement could be called +MR if it is designed to promote masking release or -MR if masking release is likely to be suppressed. We should note that our choice of +MR and -MR masker constructions are intended to promote or suppress masking release to investigate the mechanisms of masking release, not necessarily to produce absolute maximum effects for practical applications.

An example of the +MR configuration of speech and noise, based on the +MR condition in Experiment 1 of Kwon *et al.* (2012), is illustrated in Fig. 1. The illustration shows that when speech energy is momentarily high, noise energy is made to be low, and when speech energy is momentarily low, noise energy is made to be high. The +MR condition affords the listener an unusually high number of opportunities to extract speech information from the dips in the masker which should facilitate speech understanding. In the study by Kwon *et al.* (2012), NH subjects showed large amounts of masking release when tested in the +MR condition. Some CI subjects also demonstrated masking release in the +MR condition, although to a lesser degree than the NH listeners. Most CI subjects, however, did not show masking release, indicating that they did not take advantage of the extreme opportunities for listening in the dips in the +MR condition. A few CI subjects even showed modulation interference.

The lack of masking release in CI subjects, even when speech is highly exposed in the dips of the noise, suggests that the CI subjects might have had difficulty making sense of the speech even when it was audible. Considering that the noise in the +MR condition was designed to occur primarily during the quietest moments of speech, information about the boundaries between words and syllables could be weakened or obscured, particularly if noise and speech are fused rather than segregated. The diminished salience of word or syllable boundaries could impair speech segmentation.

Segmentation is the process whereby listeners arrive at the meaning of an utterance by breaking down the incoming stream of speech into linguistic units. Failures in this process

FIG. 1. Illustration of target speech (in black) and noise maskers (in gray). (A) The original sentence (*His father painted fruit*). (B) Noise with amplitude fluctuations in inverse proportion to the speech envelope. (C) Noise with steady envelope. +MR ("plus MR"): speech-noise mixture with the least temporal overlap [(A) + (B)]. Steady: speech mixed with steady noise [(A) + (C)]. The long-term RMS (root-mean-square) level of each noise displayed in this illustration matches that of the speech, i.e., the nominal SNR of each mixture is 0 dB.

hinder speech understanding. Research on segmentation has focused on two primary categories of cues: those that are signal-derived, such as prosodic, segmental, and acoustic-phonetic cues, and those that are knowledge-derived, such as lexical and syntactic cues (for a review, see [Davis et al., 2002](#)). Listeners' relative weighting of these cues has been found to depend on listening conditions; recent reports indicate that under ideal listening conditions, lexical and syntactic cues tend to dominate the process of segmentation, but when speech is severely degraded or subject to energetic masking, listeners give prosodic cues more relative weight ([Mattys et al., 2005](#); [Mattys et al., 2009](#)). Since the matter of masking release in CI subjects involves both signal degradation (due to coarse spectral representation in current CI systems) and energetic masking, prosodic cues should be particularly relevant.

One proposed segmentation strategy based on prosodic cues is called the metrical segmentation strategy (MSS; [Cutler, 1990](#)) and stems from the patterns of stress, or syllabic strength, in utterances. The MSS is a heuristic segmentation strategy in which strong syllables are assumed to be the onset of a new content word while weak syllables are assumed to be either the continuation of a multisyllabic word or the onset of a function word ([Cutler and Norris, 1988](#)). Strong syllables contain full vowels and typically receive either primary or secondary stress and weak syllables contain central, reduced vowels and are typically unstressed. Because a majority of content words in conversational English begin with strong syllables, the MSS should be highly successfully in segmenting English utterances ([Cutler and Carter, 1987](#)). In some cases, such as for content words that begin with weak syllables (e.g., “amused”), the MSS will produce an incorrect segmentation hypothesis. Therefore, listeners using the MSS are predicted to produce certain classes of lexical boundary errors more than other classes ([Cutler and Butterfield, 1992](#)). A listener commits a lexical boundary error by incorrectly inserting or deleting a word boundary. Boundary errors can be described by their type (insertion or deletion), as well as the strength of the syllable (strong or weak) immediately after the error. Examples of lexical boundary errors and their classifications are given in Table I. The left-hand column contains examples of stimuli presented to listeners in the current study, while the middle column contains examples of subjects’ verbal responses to the stimuli. The underlined syllables indicate the location of the boundary errors classified in the right-hand column. Listeners using the MSS are expected to make more incorrect insertions before strong syllables (IS) and deletions

before weak syllables (DW) than insertions before weak syllables (IW) or deletions before strong syllables (DS). Errors made by NH, HI, and CI subjects have been found to follow the patterns predicted for listeners using the MSS ([Cutler and Butterfield, 1992](#); [Liss et al., 1998](#); [Liss et al., 2000](#); [Liss et al., 2002](#); [Spitzer et al., 2009](#); [Woodfield and Akeroyd, 2010](#)), indicating that listeners use the MSS to segment speech.

Traditional measures of speech intelligibility, such as the proportion of words or sentences correctly identified by a listener, quantify the outcome of speech understanding, but might fail to lend insight into details of the process of speech understanding. Analyzing the segmentation errors made by listeners lends insight into how listeners organize and make sense of the signals they receive. From this, it may be possible to better understand why some CI listeners are unable to understand speech in an amplitude-fluctuating masker even when the speech is largely audible. Thus, one goal of the present study is to examine the role of segmentation in reduced or absent masking release demonstrated by CI users, particularly when listening to highly favorable configurations of speech and masker. This study tests the hypothesis that CI users who show little or no masking release also show less robust adherence to a useful segmentation strategy compared to those users who do show masking release. The amount of masking release in CI users was measured in a sentence keyword identification task with a steady masker and a +MR masker (i.e., a fluctuating masker promoting masking release). Their verbal responses were analyzed for boundary errors to determine their adherence to the MSS in each condition. If the use of segmentation cues underlies the diminished masking release in CI users, subjects who do not demonstrate masking release should show similar or weaker adherence to the MSS in the +MR condition compared to in the steady condition, and subjects who show masking release should produce segmentation errors that are more consistent with the MSS in the +MR masker than in the steady masker. The subjects’ use of metrical segmentation cues in the +MR masker should shed light on possible explanations, such as perceptual fusion of speech and masker, for the reduced or absent masking release in CI listeners.

II. METHODS AND PROCEDURE

A. Subjects

Eleven post-lingually deafened CI subjects implanted with Nucleus 24, Freedom, or Nucleus 5 devices completed the sentence keyword identification task. Subject

TABLE I. Examples of lexical boundary errors and their classifications based on the type (insertion or deletion) and location (before a strong or weak syllable) of the error. The syllables surrounding the boundary errors in each response and corresponding stimulus are underlined. Listeners following the MSS are expected to make more erroneous insertions before strong (IS) syllables and deletions before weak (DW) syllables than the other two classes of errors.

Stimulus presented	Subject response	Boundary error class
Platoons deserve respect.	The <u>tunes</u> deserve respect.	Insertion before strong (IS)
Clever women wrote this.	Clever windows <u>open</u> .	Deletion before weak (DW)
He <u>tapped</u> the black device.	<u>Attack</u> the black device.	Deletion before strong (DS)
The farmer <u>waters</u> crops.	The farmer <u>bought his</u> crops.	Insertion before weak (IW)

characteristics and their everyday clinical map settings are shown in Table II. Eight of the subjects (from CI1 to CI9) are the same subjects as in the previous study (Kwon *et al.*, 2012), so the same subject identifiers have been assigned for consistency, e.g., CI9 in this study indicates the same person as CI9 in the previous study. The use of human subjects was approved by the Institutional Review Boards of The Ohio State University and Gallaudet University.

B. Stimuli and apparatus

One hundred twenty sentences were constructed based on the stimuli of Cutler and Butterfield (1992). Each sentence had three keywords that were scored for identification performance. Each sentence contained six syllables arranged in an alternating pattern of strong (S) and weak (W) syllables. Half of the sentences followed a SWSWSW pattern (e.g., “Lemon juice is bitter.”) and half followed a WSWWSW pattern (e.g., “The sailors fired guns.”). To roughly balance the number of opportunities for listeners to make each class of boundary error, a third of the sentences contained only words that began with strong syllables (e.g., “Pirates plunder treasure.”), another third contained only words that began with weak syllables (e.g., “Release complete reports.”), and a third contained both strong-initial and weak-initial syllable words (e.g., “The king refused the gift.”). Excluding the first syllable of each sentence, as the first syllable of a sentence must necessarily be the first syllable of a word, each listener had 126, 163, 137, and 174 opportunities to make IS, DW, IW, and DS errors, respectively. The sentences were recorded in a double-walled, sound-attenuated booth as spoken by an adult male native speaker of Standard American English.

Sentences were presented with speech-shaped noise that had either a steady amplitude or a fluctuating amplitude designed to promote masking release, i.e., +MR. The design and generation of these maskers was identical to the steady and +MR noise maskers used in Experiment 1 of Kwon *et al.* (2012), and example illustrations of the masker waveforms are shown in Fig. 1. In brief, noise was generated each trial, on the fly, by analyzing the sentence using a bank of 1/3 octave band pass filters centered between 120 and 7680 Hz and filtering white noise to match the analyzed spectrum. For the +MR condition, the short-term amplitude

of the speech-shaped noise was adjusted in 50-ms windows in inverse proportion to the amplitude of speech, so that the level of noise was high when the level of speech was low, and the level of noise was low when the level of speech was high (for details, see Kwon *et al.*, 2012).

The experimental procedure was managed using custom software for sentence testing, Token (version 1.33), which read the “.wav” files, generated, filtered, and modulated the noise, and combined the signals at the appropriate SNR as specified in the AUditory syntaX (Kwon, 2012) module included in the Token software. Stimuli were then processed with the ACE speech coding strategy using the subject’s everyday clinical settings, shown in Table II, via custom software, ACEPlayer (version 1.91), based on the Nucleus MATLAB Toolbox (NMT; version 4.2) and the Nucleus Implant Communicator 2 library, from Cochlear Ltd. (Lane Cove, Australia). The output from ACEPlayer was delivered directly to the implant using an L34 speech processor and POD interface, also from Cochlear Ltd. The custom software packages Token and ACEPlayer were developed by the second author and are available for download at the author’s website (<http://auditorypro.com/aux/>). The presentation level of the sentence stimuli was equivalent to an acoustic input of 50 dB SPL (RMS) using the default sensitivity setting of the clinical processor.

C. Procedure

During the procedure, subjects were seated in a quiet room, and subjects who used a hearing aid or CI in the non-test ear were asked to remove their hearing aid or CI processor from the non-test ear. Subjects were presented with a sentence in noise and asked to repeat it back. The experimenter typed their verbal responses into the Token software and marked the keywords that were correctly identified. To avoid floor or ceiling effects, the SNR for each subject was selected following a short pilot experiment using other speech materials; during this pilot session, SNRs were chosen with the goal that each subject’s speech understanding performance in steady noise would be approximately half of the subject’s performance in quiet. SNRs used during testing are displayed on the right side of Fig. 2 and ranged from −2 to 7 dB. The same SNR was used in both masker conditions for a given subject.

In order to approximately balance the opportunities for the 4 classes of boundary errors between the two masker

TABLE II. Subject demographic and device information.

	Age	CI experience (yr)	Gender	Etiology	Device	Stimulation rate (Hz)
CI1	37	6.8	M	Unknown; possible noise-induced	CI24RE	900
CI2	71	4.3	F	Auto-immune	CI24RE	1200
CI3	65	2.7	M	Unknown; possible genetic component	CI24RE	900
CI4	65	3.4	F	Otosclerosis	CI24RE	500
CI6	61	6.7	F	Genetic	CI24RE	900
CI7	56	5.9	M	Ototoxicity	CI24RE	900
CI8	62	1.9	M	Meniere’s disease	CI512	500
CI9	69	3.1	M	Ototoxicity	CI24RE	900
CI15	46	3.0	M	Unknown; possible genetic component	CI24RE	900
CI16	31	3.6	M	Unknown; possible genetic component	CI24RE	900
CI17	76	7.0	M	Unknown; possible noise-induced	CI24R	900

FIG. 2. Sentence keyword identification for each subject in the +MR and steady conditions are represented by the gray and black bars, respectively. The SNR at which stimuli were presented for each subject is indicated on the right side of the figure. Subjects are ordered according to observed masking release, descending from top to bottom.

conditions, sentences were presented in blocks of 12 such that all blocks contained equal combinations of sentences following each word-initial syllable-strength rule. Within a given block, the presentation order of sentences was randomized once, and this order was not changed between subjects. The order of presentation of sentence blocks and their assignment to masker conditions was randomized for each subject, with the constraint that the masker condition alternated between blocks. In summary, there were 120 sentences in total, divided into 10 blocks of 12 sentences each. Five blocks were presented in each masker condition, thus 60 sentences were presented per condition to each subject.

D. Data analysis

Masking release was defined as the improvement in %-correct keyword identification performance from the steady condition to the +MR condition, i.e., the %-correct score in the +MR condition minus the %-correct score in the steady condition. To quantify the subjects' segmentation performance, two judges, including the first author, independently coded the lexical boundary errors in the transcripts of the subjects' verbal responses, following the procedure described by Liss *et al.* (2002) and Spitzer *et al.* (2009). Afterward, inconsistencies in coding between the two judges were either resolved or omitted from further analysis.

Boundary errors were classified according to whether the boundary error consisted of an incorrect insertion or deletion and the strength of the syllable following the boundary error. Examples of the four classes of boundary errors (IS, DW, IW, DS) are given in Table I. Strict adherence to the MSS should result in more IS and DW errors than IW and DS errors, and this adherence can be summarized using a simple ratio—referred to here as the MSS ratio—which is the ratio of the sum of the IS and DW errors to the total number of boundary errors. MSS ratios near 1 indicate robust perception of metrical stress cues and the use of these cues for segmentation in accordance with the MSS, while MSS ratios near 0.5 indicate little or no adherence to the MSS.

Out of 596 total judgments of lexical boundary errors, the two judges were inconsistent in 183 cases (~31%). Of those 183 cases, the judges reached agreement on 181, while 2 cases could not be resolved and were omitted from further analysis. After coding inconsistencies between judges were resolved, 563 total judgments of lexical boundary errors remained for further analysis.

III. RESULTS

Sentence keyword identification scores for each subject are shown in Fig. 2. Data are plotted with subjects in descending order with respect to observed masking release, from top to bottom. Average masking release for all subjects was moderate (mean = 10.7%-points), but performance was considerably variable across subjects, warranting further examination. In speech understanding in the +MR condition of 15 or greater %-points (mean = 25.6%-points), 5 subjects (CI1, CI6, CI7, CI8, CI9) showed an improvement, while the other 6 subjects each showed 5 or lower %-points of masking release (mean = -1.7%-points). Two subjects, CI2 and CI3, showed poorer sentence intelligibility in the +MR condition by 9 and 12%-points, respectively, which suggests that the effects of modulation interference outweighed the benefits of the extreme availability of speech in the +MR masker for these two subjects. Despite the procedural differences between the present and previous (Kwon *et al.*, 2012) studies, the individual patterns of masking release, or the lack thereof, are largely consistent for the eight subjects who were also subjects in the previous study, with one exception: CI8 previously showed evidence of modulation interference in the +MR condition, but presently shows substantial masking release. CI8 had only 4 months of CI experience at the time of the previous study compared to nearly 2 years of experience when he completed the present study; this difference in performance might reflect adaptation to listening with a CI (e.g., Gray *et al.*, 1995).

Counts of the boundary errors in the subjects' verbal responses are displayed in Fig. 3 and the MSS ratios calculated from these counts are shown in Fig. 4. As with Fig. 2, subjects are ordered with respect to observed masking release. All subjects produced MSS ratios > 0.5 in both masker conditions, which indicates that they perceived at least some of the metrical stress cues and made decisions about segmentation using the MSS, consistent with the findings of Spitzer *et al.* (2009). To inspect the data for

FIG. 3. The number of boundary errors of each class in subjects' verbal responses in the +MR and steady conditions are indicated by the gray and black bars, respectively. The boundary error classes shown in bold, insertion before a strong syllable (IS) and deletion before a weak syllable (DW), are expected to be more frequent in listeners using the MSS than the other two error classes, insertion before a weak syllable (IW) and deletion before a strong syllable (DS). Subjects are ordered according to observed masking release, descending from top to bottom.

differences in segmentation performance between the two masker conditions, the MSS ratios obtained in the steady condition were subtracted from the ratios in the +MR condition. Figure 5 displays a scatterplot of the differences in MSS ratios as a function of masking release. Greater values along the y-axis represent increasingly greater adherence to the MSS in the +MR condition than in the steady condition. Visual inspection of the figure reveals a trend of subjects showing more adherence to the MSS in the +MR condition (relative to the steady condition) as they show more masking release. This trend is consistent with the explanation that some CI users do not benefit from speech exposed in the dips of a fluctuating masker because they have difficulty segmenting the exposed speech.

MSS ratios calculated from individual subject data should be interpreted with caution, as some subjects made very few boundary errors. For those subjects, the distribution of errors among the four classes might be less reliable. For

FIG. 4. MSS ratios for each subject in the +MR and steady conditions are represented by the gray and black bars, respectively. MSS ratios closer to 1 indicate strong adherence to the MSS, while ratios close to 0.5 indicate little or no adherence to the MSS. Subjects are ordered according to observed masking release, descending from top to bottom.

example, CI3 and CI17 each made fewer than ten boundary errors in the steady condition. In the case of CI3, this was likely related to giving no response instead of guessing in 38 trials (excluding CI3, the average number of trials in which a

FIG. 5. A scatter plot of differences in MSS ratios between conditions (+MR - steady) as a function of masking release. Each point represents data from one subject. Higher values along the y-axis indicate greater adherence to the MSS in the +MR condition than in the steady condition.

TABLE III. Lexical boundary errors by subject group.^a

Condition	Subject group	Number of IS (%)	Number of DW (%)	Number of IW (%)	Number of DS (%)	Total LBE
Steady	Masking release	90 (56)	40 (25)	24 (15)	6 (4)	160
	No release	81 (55)	37 (25)	23 (16)	6 (4)	147
+MR	Masking release	67 (66)	25 (25)	7 (7)	3 (3)	102
	No release	75 (49)	42 (27)	25 (16)	12 (8)	154

^aSubjects showing $\geq 15\%$ -points of masking release (i.e., CI1, CI6, CI7, CI8, and CI9) comprise the masking release group, while the remaining subjects (i.e., CI2, CI3, CI4, CI15, CI16, and CI17) comprise the no release group.

subject gave no response was 2.1). CI7 gave no response in 9 trials, and in 55 trials his verbal responses consisted of only single words, greatly reducing the likelihood of lexical boundary errors.

To examine the relationship between masking release and segmentation and to increase the number of boundary errors for analysis, data from subjects were grouped according to whether a subject demonstrated substantial masking release ($\geq 15\%$ -points of identification improvement in the +MR condition), as reported above. Table III contains counts of boundary errors summed in each condition for all subjects in a group. Figure 6 displays the MSS ratios calculated from the summed boundary error counts. Both groups produced similar MSS ratios in the steady condition, but differed in their adherence to the MSS in the +MR condition. Specifically, the group that showed masking release also showed greater adherence to the MSS in the +MR condition.

Following Spitzer *et al.* (2009), Pearson's chi-square goodness-of-fit tests were conducted in R (R Core Team, 2014) to determine if there were statistically reliable differences in the distributions of lexical boundary errors made between the two subject groups within each condition. For the purposes of the tests, the counts of errors of each class

made by the subjects showing masking release were treated as observed cell-values, while the proportions of errors of each class made by subjects who did not show masking release were treated as the expected cell-values. The null hypothesis for each test was that the distribution of lexical boundary errors made by subjects who showed masking release is consistent with having sampled from the distribution of errors made by the subjects who did not show masking release. In the steady condition, the null hypothesis was not rejected at the $\alpha=0.05$ level [$\chi^2(3)=0.13$, $p=0.99$], indicating that subjects, regardless of group, produced similar distributions of lexical boundary errors in the steady condition. In contrast, the null hypothesis was rejected in the +MR condition [$\chi^2(3)=14.92$, $p=0.002$], indicating that the subjects who showed masking release produced a different distribution of lexical boundary errors in the +MR condition compared to those subjects who showed little or no release. These results reveal that the masker fluctuations have different effects on subjects' use of segmentation cues, even when the subjects otherwise show similar adherence to the MSS in the steady condition, and that this difference is related to whether the subjects demonstrated masking release.

IV. DISCUSSION

Multiple reports in the literature of reduced or absent masking release in CI listeners (e.g., Nelson *et al.*, 2003; Stickney *et al.*, 2004; Loizou *et al.*, 2009), including the recent finding in a condition where the potential for masking release was artificially promoted by reducing energetic masking (Kwon *et al.*, 2012), indicate that a fluctuating masker might impair the speech recognition of CI listeners beyond the effects of energetic masking alone. In an effort to understand this phenomenon, the present study examined the possibility that CI listeners' use of metrical segmentation cues might be influenced by masker fluctuations. The present results support the hypothesis that CI users who show reduced or absent masking release also show disruption of their use of segmentation cues in a fluctuating masker. Accordingly, a detailed look at the acoustic correlates of metrical segmentation cues in the signal should provide insight into the problem of masking release in CI listeners.

Prosodic stress can be conveyed by syllable duration and intensity, as well as by fundamental frequency (F0) and vowel quality (Fry, 1955; Lieberman, 1960; Lehiste, 1970; Fear *et al.*, 1995; Plag *et al.*, 2011). Stressed syllables typically have greater durations and intensities, full vowels, and

FIG. 6. MSS ratios in each masker condition calculated after grouping subjects according to their demonstrated masking release and summing the counts of boundary errors in each group. Subjects showing $\geq 15\%$ -points of masking release (i.e., CI1, CI6, CI7, CI8, and CI9) comprise the masking release group, shown in gray, while the remaining subjects (i.e., CI2, CI3, CI4, CI15, CI16, and CI17) comprise the no release group, shown in black.

higher F0s. Regarding metrical segmentation specifically, [Spitzer et al. \(2007\)](#) manipulated vowel quality, F0, and duration cues to reduce their salience. The manipulations of vowel quality and F0 elicited lower MSS ratios in NH listeners than manipulations of vowel duration, indicating that NH listeners are more reliant on vowel quality and F0 cues than duration cues for metrical segmentation, at least in a laboratory setting. The ecological validity of F0 as a cue for prosodic stress in naturally spoken English has been disputed ([Silipo and Greenberg, 2000](#); [Kochanski et al., 2005](#)).

No similar investigation has been conducted with CI listeners, so it is uncertain exactly which cues in the signal are most relevant to the question of metrical segmentation of English in CI listeners. This uncertainty notwithstanding, duration and intensity cues should be of particular interest in interpreting the present results for two reasons. First, because spectral information is highly degraded while temporal information is relatively preserved in modern CI systems (e.g., [Shannon, 1992](#); [Friesen et al., 2001](#)), vowel duration and intensity cues should be better represented by CI systems than either vowel quality or F0 cues and thus might be stronger, more reliable segmentation cues for CI listeners ([O’Halpin, 2010](#); [Meister et al., 2011](#); [Morris et al., 2013](#)). Second, information about vowel quality and F0 would be most salient when the speech energy was high, and in the +MR condition these portions of time had relatively little energetic masking. It is therefore unlikely that the +MR masker substantially interfered with spectral cues in the vowels of the speech.

On the other hand, duration and intensity cues could easily be disrupted by the +MR masker. Masker energy flanking a speech segment introduces uncertainty about the start or end points (or both) of the segment. Peaks in a fluctuating masker can reduce the salience of intensity contrasts between syllables. In cases where speech and masker are perceptually fused, moments dominated by speech energy may be easily confused with moments dominated by masker energy, which would exaggerate the disruptive effects of masker fluctuations on intensity and duration cues. Flanking masker energy could be interpreted as continuations of the speech segment, leading to greater ambiguity over a larger time scale about the start or end points of speech segments. Moments of high masker intensity might be interpreted as syllables, particularly when masker fluctuations are speech-like, leading the listener to make segmentation decisions based on an erroneous metrical pattern. Weak sound source segregation diminishes the salience and reliability of duration and intensity cues for segmentation, which could explain why the subjects who showed little or no masking release also did not show a difference in their use of segmentation cues between the +MR and steady conditions. Speech exposed in the dips of a masker is of little benefit if the listener cannot reliably segment the speech.

This analysis explains the performance of CI users who do not show masking release, but does not differentiate between explanations of why some CI users did show masking release. On one hand, it is possible that for CI users who demonstrated masking release, speech and masker were more perceptually segregated, thus providing more reliable information about segment duration and intensity. On the

other hand, it is also possible that speech and masker were perceptually fused for the subjects who showed masking release, but they were able to rely on other segmentation cues besides duration and intensity, such as vowel quality or F0, which are less available in the steady condition and more available in the +MR condition. Additional investigation would be needed to evaluate the relative merit of these two explanations. An expanded study with HI subjects could also shed light on the reduced benefit from masker fluctuations reported in the literature for HI listeners.

The present study explored the use of metrical stress cues for segmenting spoken English. Generally, listeners may not benefit from following a segmentation strategy identical to the MSS if the metrical patterns and probabilities of their language differ from those of English. However, for any spoken language, if a listener is likely to confuse a competing sound with a speech signal, even when the competing sound occurs primarily during moments when speech energy is low, he or she could experience difficulties in speech segmentation and understanding. Sound source segregation and speech segmentation are crucial processes for understanding any spoken language.

A primary goal of the present study was to investigate the processes by which amplitude fluctuations in a masker might disrupt speech understanding in CI users. It would not necessarily be surprising that listeners who achieve higher speech understanding in quiet also show higher speech understanding in fluctuating noise, perhaps reflecting a general advantage in understanding speech. However, the variability of outcomes among CI users undermines an explanation based on a general advantage in speech understanding. For example, four subjects were tested at an SNR of 3 dB (CI8, CI7, CI16, and CI4) and had roughly similar performance in the steady condition, yet two of these subjects showed masking release and two did not. Differences in general speech understanding performance do not explain such data neatly. A more detailed explanation of the mechanisms underlying individual differences in masking release among CI listeners is necessary.

Finally, with regard to future efforts to help CI users in noisy environments, which are more likely to have fluctuating rather than steady amplitudes, the results of the present study suggest that a signal processing algorithm facilitating auditory scene analysis would be more effective than noise reduction algorithms like those used in hearing aids. For CI listeners in adverse listening conditions, improving the SNR by a few dB during moments when speech energy is high might not translate into meaningful increases in speech understanding. This expectation is in line with information-theoretic frameworks of speech processing which posit that salient cues for speech understanding are not uniformly distributed across the speech signal ([Greenberg et al., 2003](#); [Christiansen et al., 2007](#); [Stilp and Kluender, 2010](#); [Chen and Loizou, 2012](#)).

V. CONCLUSION

CI listeners experience particular difficulty, compared to NH listeners, in understanding speech in the presence of a

fluctuating-amplitude masker. If masker energy is present but separated in time from speech energy, some CI listeners show masking release—an improvement in speech recognition relative to a steady-amplitude noise baseline, though other CI listeners still do not demonstrate masking release. The present study examined segmentation errors made by CI listeners in steady noise and in fluctuating noise to evaluate the influence of masker amplitude fluctuations on speech segmentation. CI subjects who showed substantial masking release also showed greater use of the MSS in the fluctuating masker than in the steady masker, while CI subjects who showed no benefit from masker dips produced similar segmentation errors regardless of the presence of masker fluctuations. These two groups differed specifically in their ability to segment speech exposed by masker fluctuations, as indicated by the similarity of their patterns of segmentation errors in steady noise. The present results re-affirm that the effect of noise is more than energetic masking *per se*; noise fluctuations can disrupt speech segmentation and thereby hinder the ability to understand speech, even when speech energy is highly audible in the dips of the noise. Momentary reductions in noise energy, which would typically be extremely advantageous to a NH listener, can even produce modulation interference in CI listeners. This serves as evidence that auditory scene analysis (or segregation of speech in a complex sound environment) is a fundamental challenge faced by CI listeners struggling to understand speech in the presence of competing sounds.

ACKNOWLEDGMENTS

A portion of data included in this report was presented at the 16th Conference on Implantable Auditory Prostheses, Lake Tahoe, California, in July, 2013. This study was funded in part by NIH/NIDCD (R03 DC009061, B.J.K.) and The Ohio State University Medical School. The authors thank Cassie L. Wilhelm for her assistance with data collection and Chelsea Matesi for her assistance with data coding.

- Bacon, S. P., Opie, J. M., and Montoya, D. Y. (1998). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," *J. Speech. Lang. Hear. Res.* **41**, 549–563.
- Bernstein, J. G., and Grant, K. W. (2009). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **125**, 3358–3372.
- Carhart, R., Tillman, T. W., and Greetis, E. S. (1969). "Perceptual masking in multiple sound backgrounds," *J. Acoust. Soc. Am.* **45**, 694–703.
- Chatterjee, M., Peredo, F., Nelson, D., and Baskent, D. (2010). "Recognition of interrupted sentences under conditions of spectral degradation," *J. Acoust. Soc. Am.* **127**, EL37–EL41.
- Chen, F., and Loizou, P. C. (2012). "Contributions of cochlea-scaled entropy and consonant-vowel boundaries to prediction of speech intelligibility in noise," *J. Acoust. Soc. Am.* **131**, 4104–4113.
- Christiansen, C., and Dau, T. (2012). "Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise," *J. Acoust. Soc. Am.* **132**, 1655–1666.
- Christiansen, T. U., Dau, T., and Greenberg, S. (2007). "Spectro-temporal processing of speech—An information-theoretic framework," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Berlin), pp. 515–523.
- Cutler, A. (1990). "Exploiting prosodic probabilities in speech segmentation," in *Cognitive Models of Speech Processing*, edited by G. T. M. Altman (MIT Press, Cambridge, MA), pp. 105–121.
- Cutler, A., and Butterfield, S. (1992). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *J. Mem. Lang.* **31**, 218–236.
- Cutler, A., and Carter, D. M. (1987). "The predominance of strong initial syllables in the English vocabulary," *Comput. Speech Lang.* **2**, 133–142.
- Cutler, A., and Norris, D. (1988). "The role of strong syllables in segmentation for lexical access," *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 113–121.
- Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (2002). "Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition," *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 218–244.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). "Note on informational masking," *J. Acoust. Soc. Am.* **113**, 2984–2987.
- Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (1995). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," *J. Speech. Hear. Res.* **38**, 222–233.
- Fear, B. D., Cutler, A., and Butterfield, S. (1995). "The strong/weak syllable distinction in English," *J. Acoust. Soc. Am.* **97**, 1893–1904.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Fry, D. B. (1955). "Duration and intensity as physical correlates of linguistic stress," *J. Acoust. Soc. Am.* **27**, 765–768.
- Fu, Q. J., and Nogaki, G. (2005). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," *J. Assoc. Res. Otolaryngol.* **6**, 19–27.
- Fu, Q. J., Shannon, R. V., and Wang, X. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," *J. Acoust. Soc. Am.* **104**, 3586–3596.
- Gaudrain, E., and Carlyon, R. P. (2013). "Using Zebra-speech to study sequential and simultaneous speech segregation in a cochlear-implant simulation," *J. Acoust. Soc. Am.* **133**, 502–518.
- Glasberg, B. R., Moore, B. C., and Bacon, S. P. (1987). "Gap detection and masking in hearing-impaired and normal-hearing subjects," *J. Acoust. Soc. Am.* **81**, 1546–1556.
- Gray, R., Quinn, S., Vanat, Z., and Baguley, D. (1995). "Patient performance over eighteen months with the Ineraid intracochlear implant," *Ann. Otol. Rhinol. Laryngol.* **104**, 275–277.
- Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). "Temporal properties of spontaneous speech—A syllable-centric perspective," *J. Phon.* **31**, 465–485.
- Jin, S. H., and Nelson, P. B. (2006). "Speech perception in gated noise: The effects of temporal resolution," *J. Acoust. Soc. Am.* **119**, 3097–3108.
- Jin, S. H., and Nelson, P. B. (2010). "Interrupted speech perception: The effects of hearing sensitivity and frequency resolution," *J. Acoust. Soc. Am.* **128**, 881–889.
- Jin, S. H., Nie, Y., and Nelson, P. (2013). "Masking release and modulation interference in cochlear implant and simulation listeners," *Am. J. Audiol.* **22**, 135–146.
- Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (2005). "Loudness predicts prominence: Fundamental frequency lends little," *J. Acoust. Soc. Am.* **118**, 1038–1054.
- Kwon, B. J. (2012). "AUX: A scripting language for auditory signal processing and software packages for psychoacoustic experiments and education," *Behav. Res. Methods* **44**, 361–373.
- Kwon, B. J., Perry, T. T., Wilhelm, C. L., and Healy, E. W. (2012). "Sentence recognition in noise promoting or suppressing masking release by normal-hearing and cochlear-implant listeners," *J. Acoust. Soc. Am.* **131**, 3111–3119.
- Kwon, B. J., and Turner, C. W. (2001). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," *J. Acoust. Soc. Am.* **110**, 1130–1140.
- Lehiste, I. (1970). *Suprasegmentals* (MIT Press, Cambridge, MA), pp. 1–194.
- Lieberman, P. (1960). "Some acoustic correlates of word stress in American English," *J. Acoust. Soc. Am.* **32**, 451–454.

- Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (2002). "The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria," *J. Acoust. Soc. Am.* **112**, 3022–3030.
- Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., and Edwards, B. (1998). "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech," *J. Acoust. Soc. Am.* **104**, 2457–2466.
- Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., and Edwards, B. W. (2000). "Lexical boundary error analysis in hypokinetic and ataxic dysarthria," *J. Acoust. Soc. Am.* **107**, 3415–3424.
- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (2009). "Speech recognition by bilateral cochlear implant users in a cocktail-party setting," *J. Acoust. Soc. Am.* **125**, 372–383.
- Mattys, S. L., Brooks, J., and Cooke, M. (2009). "Recognizing speech under a processing load: Dissociating energetic from informational factors," *Cogn. Psychol.* **59**, 203–243.
- Mattys, S. L., White, L., and Melhorn, J. F. (2005). "Integration of multiple speech segmentation cues: A hierarchical framework," *J. Exp. Psychol. Gen.* **134**, 477–500.
- Meister, H., Landwehr, M., Pyschny, V., Wagner, P., and Walger, M. (2011). "The perception of sentence stress in cochlear implant recipients," *Ear Hear.* **32**, 459–467.
- Morris, D., Magnusson, L., Faulkner, A., Jönsson, R., and Juul, H. (2013). "Identification of vowel length, word stress, and compound words and phrases by postlingually deafened cochlear implant listeners," *J. Am. Acad. Audiol.* **24**, 879–890.
- Nelson, P. B., and Jin, S. H. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286–2294.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- O'Halpin, R. (2010). "The perception and production of stress and intonation by children with cochlear implants," Ph.D. thesis, University of London.
- Oxenham, A. J., and Kreft, H. A. (2014). "Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing," *Trends Hear.* **18**, 1–14.
- Oxenham, A. J., and Simonson, A. M. (2009). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," *J. Acoust. Soc. Am.* **125**, 457–468.
- Plag, I., Kunter, G., and Schramm, M. (2011). "Acoustic correlates of primary and secondary stress in North American English," *J. Phon.* **39**, 362–374.
- R Core Team (2014). "R: A language and environment for statistical computing" (R Foundation for Statistical Computing, Vienna, Austria), <http://www.R-project.org/> (Last viewed 7/10/2014).
- Shannon, R. V. (1992). "Temporal modulation transfer functions in patients with cochlear implants," *J. Acoust. Soc. Am.* **91**, 2156–2164.
- Silipo, R., and Greenberg, S. (2000). "Prosodic stress revisited: Reassessing the role of fundamental frequency," in *Proceedings of the NIST Speech Transcription Workshop*, College Park, MD. <http://www1.icsi.berkeley.edu/~steveng/2000/PDF/Prosody.pdf> (Last viewed 10/2/2014).
- Spitzer, S. M., Liss, J. M., and Mattys, S. L. (2007). "Acoustic cues to lexical segmentation: A study of resynthesized speech," *J. Acoust. Soc. Am.* **122**, 3678–3687.
- Spitzer, S., Liss, J., Spahr, T., Dorman, M., and Lansford, K. (2009). "The use of fundamental frequency for lexical segmentation in listeners with cochlear implants," *J. Acoust. Soc. Am.* **125**, EL236–EL241.
- Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Stilp, C. E., and Kluender, K. R. (2010). "Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility," *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12387–12392.
- Takahashi, G. A., and Bacon, S. P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech. Hear. Res.* **35**, 1410–1421.
- Wilson, R. H., and Carhart, R. (1969). "Influence of pulsed masking on the threshold for spondees," *J. Acoust. Soc. Am.* **46**, 998–1010.
- Woodfield, A., and Akeroyd, M. A. (2010). "The role of segmentation difficulties in speech-in-speech understanding in older and hearing-impaired adults," *J. Acoust. Soc. Am.* **128**, EL26–EL31.