Sentence recognition in noise promoting or suppressing masking release by normal-hearing and cochlear-implant listeners^a

Bomjun J. Kwon,^{b)} Trevor T. Perry, and Cassie L. Wilhelm

Department of Otolaryngology-Head and Neck Surgery, Eye and Ear Institute, The Ohio State University, Columbus, Ohio 43212

Eric W. Healy

Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43212

(Received 24 August 2011; revised 1 February 2012; accepted 1 February 2012)

Normal-hearing (NH) listeners maintain robust speech understanding in modulated noise by “glimpsing” portions of speech from a partially masked waveform—a phenomenon known as masking release (MR). Cochlear implant (CI) users, however, generally lack such resiliency. In previous studies, temporal masking of speech by noise occurred randomly, obscuring to what degree MR is attributable to the temporal overlap of speech and masker. In the present study, masker conditions were constructed to either promote (+MR) or suppress (−MR) masking release by controlling the degree of temporal overlap. Sentence recognition was measured in 14 CI subjects and 22 young-adult NH subjects. Normal-hearing subjects showed large amounts of masking release in the +MR condition and a marked difference between +MR and −MR conditions. In contrast, CI subjects demonstrated less effect of MR overall, and some displayed modulation interference as reflected by poorer performance in modulated maskers. These results suggest that the poor performance of typical CI users in noise might be accounted for by factors that extend beyond peripheral masking, such as reduced segmental boundaries between syllables or words. Encouragingly, the best CI users tested here could take advantage of masker fluctuations to better segregate the speech from the background. © 2012 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.3688511>]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.71.Es [PBN]

Pages: 3111–3119

I. INTRODUCTION

Normal-hearing (NH) listeners typically demonstrate better speech understanding when speech is presented in a fluctuating masker relative to that in a steady-amplitude masker. This phenomenon, known as masking release, is accounted for by the ability to “glimpse” speech information when the instantaneous amplitude of the background is low and to integrate these glimpses to restore the content of speech. The amount of masking release is typically indicated by the improvement in either the percent-correct recognition score at a given signal-to-noise ratio (SNR) or by the decrease in SNR corresponding to a certain level of speech understanding. The speech reception threshold (SRT) uses the latter approach and a typical criterion of 50% correct. Figure 1 illustrates a shift in the performance-intensity function resulting from a fluctuating masker and the two methods of assessing masking release. The improvement in SRT for NH listeners in modulated backgrounds is typically about 6–10 dB and can be as large as 20 dB, depending on experimental conditions (e.g., Wilson and Carhart, 1969; Festen and Plomp, 1990).

In individuals with sensorineural hearing loss (SNHL), the amount of masking release is generally reduced (Wilson and Carhart, 1969; Festen and Plomp, 1990; Takahashi and Bacon, 1992; Eisenberg *et al.*, 1995; Bacon *et al.*, 1998; Bernstein and Grant, 2009). Furthermore, in cochlear implant (CI) users, masking release is largely absent (Nelson *et al.*, 2003; Stickney *et al.*, 2004; Loizou *et al.*, 2009). In fact, the possibility for the opposite phenomenon has been suggested (Kwon and Turner, 2001). In this “modulation interference,” performance is poorer in fluctuating than in steady backgrounds. Data obtained with CI users have displayed some indication of this phenomenon (Nelson *et al.*, 2003; Stickney *et al.*, 2004), which is in accord with anecdotal observations that CI users experience particular difficulty in fluctuating backgrounds.

Several mechanisms have been proposed to account for the reduced or absent masking release in the impaired population. First, it was suggested that the phenomenon in listeners with SNHL is attributable to the poor audibility of speech during brief time windows in which masker intensity is low. However, reduced masking release remained even after audibility was controlled (Takahashi and Bacon, 1992; Eisenberg *et al.*, 1995; Bacon *et al.*, 1998). Another interpretation involves a longer-than-normal recovery from forward masking in SNHL listeners (Glasberg *et al.*, 1987). This effect would tend to shorten the already brief time windows corresponding to the valleys of the masker, making speech information less accessible. The effect can be further complicated by broadened auditory filters in SNHL (Moore and Glasberg,

^{a)}Portions of this study were presented at the 34th MidWinter Meeting of the Association for Research in Otolaryngology, Baltimore, Maryland, February, 2011, and 2011 Conference on Implantable Auditory Prostheses, Pacific Grove, California, July 2011.

^{b)}Author to whom correspondence should be addressed. Electronic mail: bjkwon@gmail.com

FIG. 1. Idealized example of a shift in the performance-intensity function obtained with a fluctuating masker (top line) versus a steady-amplitude masker (bottom line). Masking release can be quantified either as an improvement in percent-correct score at a given SNR or as an improvement in the SNR corresponding to a given level of speech understanding, such as in the speech reception threshold (SRT), which typically corresponds to a score of 50%.

1986), which would cause masking to occur in the spectral domain as well. These same arguments can be extended to CI listeners, who can exhibit longer recovery times for forward masking (Nelson and Donaldson, 2001) and wide excitation patterns and abnormal growth of masking (Kwon and van den Honert, 2009). A third, more recent suggestion is that a limited coding of fundamental frequency (F0) might also result in reduced masking release in the SNHL population (Summers and Leek, 1998) and in CI users (Stickney *et al.*, 2004). However, the hypothesis that access to resolved F0 in the speech target leads to greater masking release in NH listeners has yet to be confirmed (cf. Oxenham and Simonson, 2009). In summary, the exact causes for reduced masking release in the hearing-impaired populations remain unclear.

The purpose of the present study was to examine masking release in CI users and a cohort of NH subjects for comparison under conditions in which the temporal overlap between speech and masker was controlled. In most previous studies, temporal masking of speech segments occurred randomly. While incidental masking could influence the results depending on the degree of temporal overlap between the speech and masker, it is unclear to what degree this influence has shaped general conclusions concerning speech-noise masking. A specific aim of the present study was to examine the effect of temporal overlap between speech and masker by measuring

speech understanding in contrasting conditions of speech-masker overlap. That is, in one condition, the noise masker was manipulated to avoid temporal overlap with speech to the extent possible, whereas in the other condition, noise was adjusted to temporally overlap with speech. In the former condition, masking release should be promoted, whereas in the latter, it should be suppressed. As a reference, a third condition was tested using noise with steady amplitude. This approach allows the examination of the best and worst cases of masking release. In addition, the effect of modulation interference could be measured when masking release was not dominant.

This paper reports the results of two experiments that differ in signal processing methods, procedures, and dependent variables but share the rationale in the preceding text. Specifically the following questions were addressed: (1) Will CI users demonstrate masking release when the listening condition promotes it? And (2) will modulation interference be demonstrated when masking release is suppressed? Answers to these questions should help us better understand the mechanisms underlying speech understanding by CI users in challenging everyday listening environments. If CI users cannot demonstrate masking release under the current condition that promotes masking release, then their recognition difficulties in fluctuating maskers likely extend beyond energetic masking. In addition, answers involving modulation interference could help clarify the extent to which CI users' lack of masking release is due to interference beyond energetic masking.

II. EXPERIMENT 1. MODULATED NOISE MASKERS

A. Subjects

Nine CI subjects (CI1 through CI9 in Table I) implanted with Nucleus 24, Freedom, or Nucleus 5 devices and 10 young-adult NH subjects with pure tone thresholds of 20 dB HL or better (ANSI, 2004) participated in expt. 1. Detailed subject information for CI participants is shown in Table I. The NH listeners were females aged 22–29 yr (mean = 26). The use of human subjects was approved by the Institutional Review Board at The Ohio State University.

TABLE I. Cochlear implant subject demographic and device information. The stimulation rate was the same as that in the everyday clinical setting.

	Age	CI experience (yr)	Gender	Etiology	Device	Stimulation rate (Hz)
CI1	35	5	M	Unknown; possible noise-induced	CI24RE	900
CI2	69	2.5	F	Auto-immune	CI24RE	1200
CI3	63	1.2	M	Unknown; possible genetic component	CI24RE	900
CI4	63	6 (L), 2 (R)	F	Otosclerosis	CI24R (L), CI24RE (R)	500
CI5	77	1	M	Noise-induced	CI24RE	900
CI6	59	3.3 (L), 5 (R)	F	Genetic	CI24RE (L&R)	900
CI7	54	4	M	Ototoxicity	CI24RE	900
CI8	61	0.3	M	Meniere's disease	CI512	500
CI9	66	2	M	Ototoxicity	CI24RE	900
CI10	60	4.7	F	Unknown	CI24R	250
CI11	64	3.8	F	Ototoxicity	CI24RE	250
CI12	53	1.7	M	Congenital	CI24RE	900
CI13	73	3	M	Unknown; possible noise-induced	CI24RE	900
CI14	38	1.6	M	Unknown; possible genetic component	CI24RE	900

B. Stimuli

The IEEE sentence materials (IEEE, 1969) were employed. They are composed of 72 lists with 10 sentences in each list and 5 keywords in each sentence. For the current experiment, a male speaker was selected (average $F_0 = 116$ Hz). As mentioned in Sec. I, there were three listening conditions: two conditions that differed in terms of speech-masker overlap and a reference (“steady”) condition having noise with steady amplitude. Masking release was promoted in one condition and suppressed in the other—these conditions will be referred to as +MR (“plus MR”) and −MR (“minus MR”), respectively. In each condition, the spectrum of the noise was matched with that of the sentence presented in each trial, as the testing software (see following text) analyzed the speech signal using a series of 1/3-octave band-pass filters centered from 120 to 7680 Hz and shaped the spectrum of broadband white noise accordingly.

In the +MR condition, the short-term amplitude of the speech-shaped noise masker was adjusted in every 50-ms window in inverse proportion to the speech envelope as described in the following formula:

$$N = L - X, \quad (1)$$

where N and X are the short-term RMS levels of noise and speech in the same time window, respectively. L was a constant related to the level of the noise floor and was empirically set at −50 dB relative full scale. This equation specifies that the noise amplitude is adjusted in the opposite direction of the speech amplitude (i.e., the less intense the speech, the more intense the noise, and vice versa). In addition, the amplitude of the noise (N) was limited to 10 dB below full scale to avoid clipping during windows having low speech levels. Figure 2 provides an example of a speech waveform (A) and a noise having the opposite temporal pattern of amplitude excursion (B). The mixture of A and B, shown in the right panel, was the stimulus for the +MR condition.

For the −MR condition, to maximize the overlap between speech and masker and to disallow “dip-listening”

using the least noise energy, the noise envelope was adjusted to grossly follow the speech envelope. First, the envelope of the speech signal was obtained by full-wave rectification and low-pass filtering at 20 Hz (elliptic filter, $N = 3$). It was then multiplied with the speech-shaped noise, resulting in speech-modulated noise. This modulated noise is illustrated as C in Fig. 2, and the mixture of A + C in the right panel is the stimulus for the −MR condition. Note that signal C is a wideband speech-shaped noise modulated by the speech envelope, similar to a 1-channel noise vocoder (Shannon *et al.*, 1995) or speech-correlated noise (Schroeder, 1968). Although the noise might offer a small amount of temporal speech information in the speech envelope, this information yields no open-set speech recognition (Shannon *et al.*, 1995; Turner *et al.*, 1995). Finally, the steady condition was created simply by mixing the speech-shaped noise (D in Fig. 2) with the speech (A + D). In all three masker conditions, the SNR was defined as the RMS average level across a given sentence relative to the RMS average across the noise employed to mask that sentence.

C. Procedure and apparatus

An adaptive procedure (1-down/1-up) was employed to obtain the SRT, defined as the SNR required for 50% correct responses. The SNR was reduced if the subject correctly identified the sentence and increased if the response was inaccurate. As all NH subjects were able to identify all keywords in the sentences in quiet, correct-response trials were defined by accurate identification of three or more keywords (of 5).¹ Because there was considerable variability among CI subjects’ performance in quiet, correct-response trials were defined as those exceeding half of each individual’s average score in quiet: For example, the quiet score was 78% for CI2, so any trial with two or more correctly identified keywords ($2/5 = 40\%$, which exceeds half of 78%) was considered correct. Identification scores for each CI subject in quiet were measured prior to testing using 20 IEEE sentences not used in the experimental sessions. Therefore, this procedure

FIG. 2. Illustration of target speech and noise maskers for three conditions examined in Expt. 1. (A) The original sentence (*The hogs were fed chopped corn and garbage*). (B) Noise with amplitude fluctuations in inverse proportion to the speech envelope. (C) Noise with amplitude fluctuations coherent with the speech. (D) Noise with steady envelope. +MR (plus MR): Speech-noise mixture with the least temporal overlap (A+B). −MR (minus MR): Speech-noise mixture with the most temporal overlap (A+C). Steady: Reference condition containing speech mixed with steady noise (A+D). The noise is displayed in gray and the speech in black. The long-term RMS energy of the noise displayed in this illustration, in all conditions [(B) to (D)], matches that of speech, i.e., the nominal signal-to-noise ratio of the mixture is 0 dB.

converged on the SNR estimating 50% performance (Levitt, 1971), relative to the score in quiet for all listeners.

The procedure began with a favorable SNR, i.e., +20 dB for CI subjects and +5 dB for NH subjects. For CI subjects, SNR was adjusted in 8-dB steps for the first two reversals and in 4-dB steps for the remaining six reversals. For NH subjects, the adjustment was 4 and 2 dB for the first two and subsequent reversals, respectively. The final four values at reversal points were averaged and taken as the SRT for the block of trials.

The presentation order of conditions was randomized as was the sentence-to-condition correspondence. At least three blocks of trials were employed for each condition. If the standard deviation of the three SNR's exceeded 3 dB, additional blocks, up to a total of six blocks, were completed until the SD was below 3 dB, and the last three blocks were accepted. All NH subjects showed relatively small test-retest variability and standard deviations were all less than 3 dB. For CI subjects, 4 of 11 ears required additional sessions in at least one condition.

A generalized psychoacoustic testing software tool, (Psycon[®], version 2.0), was used to retrieve the individual sentence.wav files, prepare the stimuli on-line, and administer the adaptive procedure. NH subjects were seated in a double-walled sound-attenuated booth and presented with the stimuli via Sennheiser HD 280 headphones. The PSYCON software delivered the acoustic output via an ECHO Gina 3G D/A converter. The overall presentation level (speech plus noise) was set at 65 dBA in a flat-plate coupler (Larson Davis AEC101).

For CI subjects, the interface and the accompanying software were designed to emulate typical operation of clinical processors within the Nucleus system. Another software module, ACEplayer[®] (version 1.6), converted the .wav file processed by Psycon to a format consisting of interleaved pulse trains in accord with ACE[™] processing. Patient-specific information was employed, including threshold and comfort levels, stimulation rate, and electrodes used. Finally, the prepared stimulus was delivered by ACEplayer to the POD interface that relayed the prepared RF signal to the CI subjects. ACE[™] processing and delivery of the stimuli to subjects in ACEplayer was based on the Nucleus MATLAB Toolbox (NMT; version 4.2) and the Nucleus Implant Communicator 2 library (NIC2), respectively, from Cochlear Ltd (Lane Cove, Australia). This direct-stimulation setting has the advantages of (1) precise specification of processing parameters and stimulus control, not bound by general-use clinical software and (2) not requiring a sound-attenuated booth for testing.

Psycon[®] and ACEplayer[®] were developed by the first author and are available for download at the author's website (Kwon, 2011). The former is open to the public under Academic Free License 3.0, whereas the latter requires NIC2 and NMT, both of which are available through license and a non-disclosure agreement with Cochlear Ltd. Signal preparation in Psycon[®] is achieved with a new scripting language, namely AUX (AUDitory syntaX). Both the program and AUX are described in detail in a recent paper (Kwon, 2012). Also, AUX codes used for the present study are available on the download website, so that readers can generate stimuli for the +MR and -MR conditions for their own exploration.

D. Results and discussion

Masking release was defined as the performance improvement from the steady to the +MR condition. Modulation interference was defined by performance reductions from the steady to the -MR or +MR conditions. Figure 3 displays SRTs measured in the young-adult NH subjects. Overall, the results were highly similar across subjects. In the steady condition, SRTs ranged from -5.8 to -4.2 dB with an average of -5.1 dB. This is roughly consistent with the study of Nelson *et al.* (2003), using the same recordings of the IEEE materials, where keyword identification was 15% for -8 dB SNR and 80% for 0 dB. Scores in the -MR condition were similar to those in the steady condition (range: -6.5 to -4.5 dB, mean = -5.3 dB). The lack of an average performance reduction indicates that modulation interference was absent. As expected, NH listeners demonstrated very robust performance in the +MR condition: SRTs ranged from -19.5 to -23.0 dB with an average of -21.6 dB. Masking release for these subjects averaged 16.5 dB.

Figure 4 displays the results for the CI subjects in each of the three masker conditions. Also displayed in the following text, the abscissa are mean IEEE sentence scores measured in quiet. The order of appearance in the figure is based on these values. First, the mean IEEE quiet intelligibility score across subjects was 79.6%, which was similar to that in the study of Nelson *et al.* (2003). This value, combined with the fact that three users scored 94% and above, while only two scored below 69%, indicates that these were generally "successful" CI users. As anticipated, a great deal of variability was observed. SRTs in the steady condition varied from -1.0 to +10.5 dB, and the standard deviations are generally larger for CI subjects than for NH subjects.

Because the data for the CI subjects are heterogeneous, individual planned comparisons (paired *t*-tests) were used to evaluate the presence of masking release or modulation

FIG. 3. SRTs measured in NH subjects in Expt. 1. Error bars indicate one standard deviation.

interference for each subject. The results of this analysis are displayed in Fig. 4. Asterisks below individual bars denote masking release (as reflected by performance in the +MR condition significantly exceeding that in the steady condition), whereas asterisks above individual bars indicate modulation interference (as reflected by performance in the steady condition significantly exceeding that in either MR condition). Values are significant at $P < 0.05$ unless made explicit. The marginally-significant values $P < 0.06$ and $P < 0.07$ were displayed in the figure and included in the interpretation of these results due to the low power of the tests, which in turn results from the limited number of values comprising each mean.² The leftmost subjects in the figure had the best speech understanding and displayed a pattern of results most like that of their younger NH counterparts. Masking release was significant in these users, averaging 8.2 dB in the +MR condition, and masking interference was absent. A vertical dotted line in Fig. 4 separates these best users from the others. No subjects to the right of this line display MR. For two of these users (three ears), no significant difference was observed between the steady condition and either MR condition. That is, the manipulation of noise in the +MR condition, which produced a profound effect for the NH subjects, had no significant effect for these CI subjects. The remaining four CI subjects demonstrated poorer performance in at least one of the modulated conditions. This modulation interference, which was absent for NH subjects, occurred in both –MR and +MR conditions.

It was expected that modulation interference would be demonstrated, if present, in the –MR condition. It is therefore unexpected that modulation interference appeared in the +MR condition for two of the CI users. While the mechanisms underlying this performance are unclear, it is apparent that the influence of a masker is not always in the direction expected from the physical characteristics of the acoustic signals, except for the best-performing CI subjects here. As discussed later, this implies that factors beyond peripheral masking might potentially account for some of the consequences of maskers on speech understanding by CI users.

III. EXPERIMENT 2. GATED NOISE MASKERS

Experiment 2 was driven by the same objective as expt. 1—to evaluate masking release and/or modulation interference in contrasting masking conditions—but employed a number of methodological modifications to test the general-

ity of the findings. Although the same terms “+MR” and “–MR” are used, the construction of the noise maskers in expt. 2 differed from that in expt. 1, as explained in the following text.

A. Noise masker preparation

In expt. 2, instead of modulating the noise masker based on the short-term amplitude of the speech, a noise masker with a constant but gated (on or off) amplitude was employed. Unlike the gated noise used in previous studies (e.g., Nelson *et al.*, 2003), the gating was not regular in timing but was instead determined by the short-term amplitude of the speech, according to the rule specified in the following text. While the noise was present (“gated on”) for 50% of the speech duration in both +MR and –MR conditions, for the +MR condition, the noise was arranged such that temporal overlap between speech and noise was minimized and for the –MR condition the temporal overlap was maximized. Thus the amplitude of the noise masker was constant for a given SNR and only the timing of the noise relative to the speech was different across +MR and –MR conditions. In this approach, the absence of amplitude fluctuations in the masker caused the masker to be perceptually distinct from speech, particularly for the –MR condition.

Figure 5 illustrates this signal preparation. Specifically, (1) the RMS energy of the speech was computed during each 50-ms segment, (2) the segments comprising the lower and upper 50th percentile were identified, and (3) the noise was gated on or off for each segment in the +MR or –MR condition to minimize or maximize the temporal overlap, respectively. It should be noted that this approach was based on the presumption that the amount of speech information is proportional to the RMS energy in each short time period (ANSI, 1997). Although this notion should not be taken as a precise premise, it is in line with the rough approximation that the intelligibility of speech presented in noise is monotonically related to the relative energy ratio of the speech to noise (e.g., Oxenham and Simonson, 2009).

B. Subjects, procedure, and apparatus

Six CI subjects (CI9 through CI14 in Table I) implanted with Nucleus 24 or Freedom devices and 12 subjects with NH [pure tone thresholds of 20 dB HL or better (ANSI,

FIG. 4. SRTs measured in CI subjects in Expt. 1. Error bars indicate one standard deviation. Subjects are ordered by IEEE sentence intelligibility score in quiet, which is provided. Asterisks below individual bars reflect significant masking release (steady > +MR), whereas asterisks above individual bars indicate significant modulation interference (steady < +MR or –MR). The dotted line separates subjects who displayed significant masking release.

FIG. 5. Illustration of target speech and noise masker preparation in Expt. 2. (A) The original sentence, as Fig. 2. (B) Noise gated off during the 50% of 50-ms windows in which the speech target contained the most energy. (C) Noise gated off during the 50% of 50-ms windows in which the speech target contained the least energy. +MR: Speech-noise mixture with the least temporal overlap (A+B). -MR: Speech-noise mixture with the most temporal overlap (A+C). The noise signal is displayed in gray and the speech in black. The long-term RMS energy of the noise displayed in this illustration matches that of the speech, i.e., the signal-to-noise ratio of the mixture is 0 dB.

2004)] participated in expt. 2. Note that CI9 participated in both expts. 1 and 2. Detailed information for CI participants is shown in Table I. The NH listeners were aged 20–38 yr (mean = 25 yr) and 10 were female. The use of human subjects was approved by the Institutional Review Boards at the University of Utah and The Ohio State University.

For expt. 2, the CUNY sentences (Boothroyd *et al.*, 1988) were used. The SNR was fixed for each subject, and intelligibility was based on recognition of constituent words. As in expt. 1, the improvement from the steady condition reflected masking release and decrements in either gated condition reflected modulation interference. Six NH subjects were tested at an SNR of –9 dB (speech at 65 dBA and noise at 74 dBA), and six subjects were tested at an SNR of –25 dB (speech at 50 dBA and noise at 75 dBA). For CI subjects, recognition was measured in quiet, followed by a short pilot exploration to determine an appropriate SNR. Although individual CI listeners' speech understanding in noise can be difficult to predict, individual SNR's were chosen for each subject between 5 and 15 dB in an attempt to minimize floor and ceiling effects.

NH listeners received a brief familiarization (six sentences) in each condition, followed by 24 sentences (two lists) in each of the three masking conditions (steady, –MR, +MR) at –9 dB SNR or 48 sentences (four lists) in the –MR and +MR conditions at –25 dB SNR. Cochlear-implant users received a brief familiarization, followed by 48 sentences (four lists) in each of the three masker conditions. The presentation order of conditions and the sentence list-to-condition correspondence was randomized for each subject. After each presentation of a sentence, subjects were asked to repeat what they heard and the number of words correctly identified was recorded by the experimenter. The

testing apparatus was the same as in expt. 1 except that a program called TOKEN replaced Psycon, as it is more suitable for testing with prepared lists of speech materials. This program is also available for download at the website indicated earlier.

C. Results and discussion

Figure 6 displays the intelligibility scores for the NH subjects. At –9 dB SNR (left panel), the steady noise produced substantial masking, with scores ranging from 20.6 to 43.6%. The scores improved substantially in the two gated masker conditions (scores of 87% and higher). In contrast with expt. 1, where scores in the –MR condition were equivalent to those in the steady condition for NH subjects, the –MR condition here resulted in substantial masking release. This is attributed to the difference in the masker preparation across experiments. The –MR masker left speech exposed during the time windows having lower short-term RMS levels. These least-intense speech segments appear to have been sufficient to allow the NH listeners to achieve high levels of speech understanding. Given this robust performance in the –MR condition, it is no surprise to observe NH performance at ceiling in the +MR condition. The +MR masker provided no decrement in performance even at an SNR of –9 dB.

Figure 6 also shows performance at –25 dB SNR (right panel). This SNR was employed to examine potential differences across the –MR and +MR conditions. Note that the speech was completely unintelligible in the steady condition at this SNR, so it was not tested. At this SNR, scores in the –MR condition decreased substantially, ranging from 17.4 to 32.6%. However, scores in the +MR condition remained at ceiling.

FIG. 6. Sentence intelligibility for NH subjects in Expt. 2. Sentences were presented in gated noise at SNRs of –9 dB (left panel) and –25 dB (right panel). The gating arrangement masked either the most intense (–MR) or least intense (+MR) 50-ms segments of speech. Error bars indicate one standard deviation.

Given the choice of signal preparation in expt. 2, it seems fair to conclude that masking release for the young-adult NH subjects in the +MR condition is too great to quantify in terms of the improvement in percent-correct scores from the steady condition. The average amount of masking release (the improvement in intelligibility from the steady to +MR condition) was limited only by the performance ceiling and the selection of SNR. The average score was 67.6% at -9 dB SNR and immeasurable at -25 dB SNR. It is notable that considerable masking release also appeared in the $-MR$ condition, albeit to a lesser degree.

Figure 7 displays the results from the CI subjects, including scores in quiet and the SNR employed. Sentence identification ranged from 22% (CI10) to 100% (CI9) with an average of 75.5%. It is clear from the figure that the CI users did not demonstrate the profound masking release shown by the NH subjects. Again the subjects were examined individually, here using planned comparisons (paired t tests) on the RAU-transformed data (Studebaker, 1985) to evaluate the presence of masking release. The results of this analysis are displayed in Fig. 7 as asterisks above conditions in which intelligibility exceeded that in the steady condition. As in expt. 1, values are significant at $P < 0.05$ unless made explicit. Only two users (CI9 and CI11) demonstrated masking release in both gated masker conditions, consistent with the pattern displayed by the NH listeners. However, the magnitude of the release is far smaller than that observed in the NH subjects. Also note that CI9 was one of the subjects who demonstrated masking release in expt. 1. Significant masking release is observed in only the +MR condition for two additional listeners (CI10 and CI14), but the size of this effect is extremely small for CI 10. Thus, masking release was considerably reduced, relative to that observed in NH listeners. Modulation interference, i.e., the steady condition showing a better score than either $-MR$ or +MR, was non-existent in these listeners as well.

To summarize, the overall pattern of results was similar across expts. 1 and 2 despite differences in procedures and signal preparation. One-third of the CI users tested were able to display masking release in expt. 1, and these corresponded to the users having the best sentence intelligibility scores in quiet. Some remaining users displayed an indication of interference in a modulated masker. In expt. 2, again one-third of

the CI users displayed a pattern of results mimicking that of the NH listeners, but the relationship between intelligibility in quiet and masking release was not as strong, and the amount of masking release was far reduced, relative to the NH counterparts.

IV. GENERAL DISCUSSION

Although masking release (i.e., “dip-listening”) is a well-known phenomenon, the present study demonstrates that the effect of a noise masker on speech recognition can vary dramatically depending on the degree of temporal overlap between speech and masker. While previous studies demonstrating masking release have provided the average effects of incidental masking across speech-noise combinations, the present study describes a range of temporal masking effects of speech by noise. A unique finding in the present study is that some CI listeners—approximately one third of the CI subjects tested—are capable of listening in the dips or “glimpsing” to achieve better speech understanding in a condition that facilitates masking release. Figure 4 shows this clearly—first, the pattern of results changes from left to right, as only the best users (defined based on sentence intelligibility in quiet) display masking release. Second, note that overall SRT values change rather dramatically, from negative values for subjects who display masking release, to positive values for subjects who do not. But it is equally important to note that good quiet intelligibility scores alone are not sufficient to determine whether a user will display masking release. Indeed, the line dividing masking-release from non-masking-release users in Fig. 4 separates quiet intelligibility scores by only 1 percentage point.

For the NH group, modulation interference, the phenomenon opposite to masking release, was not observed. This is not surprising as modulation interference rarely occurs in NH listeners unless there is substantially decreased redundancy in the speech signal such as restrictions or manipulation of spectral or temporal information (Kwon and Turner, 2001). On the other hand, CI listeners are subject to modulation interference due to reduced spectral representations of speech signals and their increased reliance on envelope modulation cues, which could potentially be disrupted by fluctuating maskers. Indeed, modulation interference was observed in several of the CI users in expt. 1, although this phenomenon did not appear in expt. 2. While it could be considered puzzling that modulation interference was equally distributed across the $-MR$ and +MR conditions in expt. 1, this result might suggest that factors beyond peripheral masking, such as a disruption of word boundaries in the sentence, might play a role in modulation interference and ultimately in CI users’ speech understanding in noise.

When considering differences in performance across listener groups, it is potentially important to consider differences in SNR at which the groups are tested (cf. Bernstein and Grant, 2009). Although it is possible that the amount of masking release would decrease, as the SNR increases (Bernstein and Grant, 2009), the influence on modulation interference is unclear. While an effect of SNR cannot be ruled out, Bernstein and Grant’s account was based on temporal

FIG. 7. Same as Fig. 6, but for CI subjects in Expt. 2. Sentence intelligibility scores in quiet are shown as is the SNR at which each subject was tested.

masking of speech information with noise occurring randomly. It is unclear how applicable it is to the +MR and –MR conditions, in which timing was controlled artificially.

Another aspect to be considered involves age differences between the NH and CI subjects. The average age of the NH group tested in this study was 25 yr, whereas that of the CI group was 60 yr. A body of literature suggests that older listeners with normal hearing sensitivity exhibit less speech recognition benefit from interrupted or modulated background than do younger NH individuals (Dubno *et al.*, 2002, 2003; Gifford *et al.*, 2007; Grose *et al.*, 2009). However, the effects of age in these studies are in general far smaller than the vast differences observed between NH and CI groups in the present study. Therefore, while the age difference might contribute to the present results to a certain degree, it should not be a primary factor.

Overall, the degree of speech-noise overlap had far less impact on speech understanding for CI users than for the NH listeners. This result might have implications for auditory scene analysis. The +MR condition provided a good representation of the speech signal—essentially all of the speech was available in the +MR condition, interspersed with noise only when the speech signal was least present. Therefore in this condition, listeners did not have the burden of simultaneous segregation of the target signal from the background. Instead a sequential grouping involving the selection and integration of information during time windows containing the target was required for good performance. Robust performance by NH and by the best CI subjects in this condition suggests that they successfully completed (1) formation of the auditory objects and (2) selection of the target (see Shinn-Cunningham, 2008).

In contrast, the results displayed by the typical CI subjects tested here—a smaller effect of masking release and incidences of modulation interference—suggest an important deficit in the typical CI users' perception of speech. One possible account is that perceptual differences between speech and noise are less clear for these listeners, and the boundaries of words and syllables may have been obscured by the noise that occurred during the quieter portions of speech in the +MR condition. Considering that amplitude fluctuations in the speech waveform are often markers of segmental boundaries and play an important role in speech recognition (e.g., Drullman *et al.*, 1994a,b), compromised word and syllable boundaries in the +MR condition combined with less perceptual difference between speech and noise, might account for the lack of MR in CI listeners. This interpretation is in accord with the suggestion that the detrimental influence of reverberation on CI-user performance can be attributed to a filling in of low speech-intensity segments and resulting segmentation difficulties (Kokkinakis *et al.*, 2011; Hazrati and Loizou, 2012).

In everyday listening environments, speech and maskers typically overlap in random fashion and to moderate degrees, producing conditions intermediate to the extreme conditions, +MR and –MR, tested here. The results of the present study suggest that the poor speech understanding in noise exhibited by typical CI users, often far poorer than that predicted by acoustic models such as the speech intelligibility index

(ANSI, 1997), might be partially attributable to the detrimental effect of noise on the segmental (word or syllable) boundaries of speech.

V. CONCLUSIONS

In the current study, conditions were constructed to promote or suppress masking release by controlling the degree of temporal overlap between speech and masker. These conditions involved modulated (expt. 1) and gated noise (expt. 2) that overlapped minimally or maximally with the sentence materials in conditions labeled +MR and –MR. Listeners with NH displayed vast amounts of masking release in the condition designed to promote it. These conditions also maximized the opportunity to observe masking release in CI users to the extent that these users possess such a mechanism. In sharp contrast to their younger NH counterparts, the CI users displayed far less masking release overall. In fact, some displayed an opposite phenomenon, modulation interference, in which performance was poorer in modulated than in steady backgrounds. But encouragingly, approximately one-third of the CI users tested here did display patterns of performance approximating those of the NH listeners, although their masking release was substantially reduced in magnitude. In expt. 1, it was the users with the best sentence intelligibility in quiet scores that displayed the ability to take advantage of modulated maskers having minimal overlap with speech.

Overall, it appears that some CI users can take advantage of dips in the masker to improve intelligibility when conditions are designed to maximize this effect. But it also appears that the average user lacks this critical aspect of normal hearing. Further, it appears that some users may suffer interference from modulated backgrounds. These deficits may potentially be related to a reduced perceptual difference between speech and noise and to reduced segmental boundaries of speech (syllable or word boundaries) due to noise, resulting in poorer than normal ability to perceptually segregate speech from noise.

ACKNOWLEDGMENTS

This study was supported by the NIH/NIDCD (Grant Nos. R03DC009061 to B.J.K. and R01DC008594 to E.W.H.). The authors thank Carla Berg and Vauna Olmstead for their assistance in data collection, Sarah Yoho for executing the statistical analyses, and Susan Nittrouer for her helpful comments.

- ANSI. (1997). S3.5, *Methods for the Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).
- ANSI. (2004). S3.6, *Specifications for Audiometers* (Acoustical Society of America, New York).
- 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.
- Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (1988). "Voice fundamental frequency as an auditory supplement to the speech-reading of sentences," *Ear Hear.* **9**, 306–312.
- Drullman, R., Festen, J. M., and Plomp, R. (1994a). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.

- Drullman, R., Festen, J. M., and Plomp, R. (1994b). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053–1064.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2002). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," *J. Acoust. Soc. Am.* **111**, 2897–2907.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2003). "Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing," *J. Acoust. Soc. Am.* **113**, 2084–2094.
- 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.
- 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.
- Gifford, R. H., Bacon, S. P., and Williams, E. J. (2007). "An examination of speech recognition in a modulated background and of forward masking in younger and older listeners," *J. Speech Lang. Hear. Res.* **50**, 857–864.
- 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.
- Grose, J. H., Mamo, S. K., and Hall, J. W., 3rd (2009). "Age effects in temporal envelope processing: speech unmasking and auditory steady state responses," *Ear Hear.* **30**, 568–575.
- Hazrati, O., and Loizou, P. C. (2012). "Tackling the combined effects of reverberation and masking noise using ideal channel selection," *J. Speech Lang. Hear. Res.* (in press).
- IEEE. (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.
- Kokkinakis, K., Hazrati, O., and Loizou, P. C. (2011). "A channel-selection criterion for suppressing reverberation in cochlear implants," *J. Acoust. Soc. Am.* **129**, 3221–3232.
- Kwon, B. J. (2011). *Psycon* [Computer program], <http://auditorypro.com/download> (Retrieved December 19, 2011).
- Kwon, B. J. (2012). "AUX: A scripting language for auditory signal processing and software packages for psychoacoustic experiments and education," *Behav. Res. Methods* (in press).
- 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.
- Kwon, B. J., and van den Honert, C. (2009). "Spatial and temporal effects of interleaved masking in cochlear implants," *J. Assoc. Res. Otolaryngol.* **10**, 447–457.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- 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.
- Moore, B. C., and Glasberg, B. R. (1986). "Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments," *J. Acoust. Soc. Am.* **80**, 93–107.
- Nelson, D. A., and Donaldson, G. S. (2001). "Psychophysical recovery from single-pulse forward masking in electric hearing," *J. Acoust. Soc. Am.* **109**, 2921–2933.
- 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.
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.* **95**, 1085–1099.
- 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.
- Schroeder, M. R. (1968). "Reference signal for signal quality studies," *J. Acoust. Soc. Am.* **44**, 1735–1736.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Shinn-Cunningham, B. G. (2008). "Object-based auditory and visual attention," *Trends Cogn. Sci.* **12**, 182–186.
- 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.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Summers, V., and Leek, M. R. (1998). "FO processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," *J. Speech Lang. Hear. Res.* **41**, 1294–1306.
- 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.
- Turner, C. W., Souza, P. E., and Forget, L. N. (1995). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," *J. Acoust. Soc. Am.* **97**, 2568–2576.
- Wilson, R. H., and Carhart, R. (1969). "Influence of pulsed masking on the threshold for spondees," *J. Acoust. Soc. Am.* **46**, 998–1010.