Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners

Peggy B. Nelson[a)]
Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

Su-Hyun Jin
Department of Communication Disorders, Department of Otolaryngology, University of Minnesota, Minneapolis, Minnesota 55455

Arlene Earley Carney
Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

David A. Nelson
Department of Otolaryngology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 15 March 2002; accepted for publication 4 November 2002)

Many competing noises in real environments are modulated or fluctuating in level. Listeners with normal hearing are able to take advantage of temporal gaps in fluctuating maskers. Listeners with sensorineural hearing loss show less benefit from modulated maskers. Cochlear implant users may be more adversely affected by modulated maskers because of their limited spectral resolution and by their reliance on envelope-based signal-processing strategies of implant processors. The current study evaluated cochlear implant users' ability to understand sentences in the presence of modulated speech-shaped noise. Normal-hearing listeners served as a comparison group. Listeners repeated IEEE sentences in quiet, steady noise, and modulated noise maskers. Maskers were presented at varying signal-to-noise ratios (SNRs) at six modulation rates varying from 1 to 32 Hz. Results suggested that normal-hearing listeners obtain significant release from masking from modulated maskers, especially at 8-Hz masker modulation frequency. In contrast, cochlear implant users experience very little release from masking from modulated maskers. The data suggest, in fact, that they may show negative effects of modulated maskers at syllabic modulation rates (2–4 Hz). Similar patterns of results were obtained from implant listeners using three different devices with different speech-processor strategies. The lack of release from masking occurs in implant listeners independent of their device characteristics, and may be attributable to the nature of implant processing strategies and/or the lack of spectral detail in processed stimuli. [DOI: 10.1121/1.1531983]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ts, 43.64.Me [CWT]

I. INTRODUCTION

Many natural background noises are temporally fluctuating, such as clattering dishes or background conversations. Listeners with normal hearing sensitivity take advantage of gaps in these fluctuating or modulated maskers. They are able to ''listen in the dips'' of the modulated masker to extract information about the speech signal. These extracted pieces of the message, then, are often sufficient to provide full understanding of the message. This improvement in speech recognition provided by modulated maskers compared to steady maskers is referred to as a ''release from masking.'' The amount of release from masking in normal-hearing listeners ranges in published reports from less than 5 dB to as much as 20 dB, depending on the stimuli and the temporal characteristics of the maskers (e.g., Bacon *et al.*, 1998). For most speech stimuli, the optimal masker modulation rates for observing masking release fall between 10 and 32 Hz (e.g., Gustafson and Arlinger, 1994). At slower modulation rates, whole syllables or words may occasionally be masked by a cycle of noise. At faster modulation rates, forward masking may perceptually fill the nominal silent interval, resulting in performance similar to that of a continuous masker.

Listeners with hearing loss are less able than normal listeners to obtain release from modulated maskers (e.g., Festen and Plomp, 1990; Takahashi and Bacon, 1992; Eisenberg *et al.*, 1995; Bacon *et al.*, 1998). Eisenberg and colleagues tested listeners with normal hearing and listeners with hearing loss for their understanding of consonants in steady and fluctuating noise. Listeners with normal hearing were tested with shaped noise designed to simulate the hearing sensitivity of the impaired listeners. Their results suggested that listeners with true hearing loss obtained far less release from modulated maskers than did normal-hearing listeners with or without simulated hearing losses. Amplification restored some, but not all, of the expected release from masking for impaired listeners. Eisenberg and colleagues concluded that audibility alone cannot explain the additional masking experienced by listeners with sensorineural hearing loss.

In contrast, Trine (1995) hypothesized that the primary problem for listeners with hearing loss was, in fact, reduced

[a)]Electronic mail: nelso477@umn.edu

audibility of signals that occurs in the dips of the fluctuating maskers. He noted a high negative correlation between masking release and the degree of hearing loss of his impaired listeners. He also noted that when amplification was provided to the impaired listeners, especially in the high-frequency region, the amount of release from masking increased, and approached that obtained by normal-hearing listeners. He postulated that if it were possible to amplify all signals, such that the temporal dips in the modulated maskers resulted in full audibility of the signals during that cycle, then impaired listeners might obtain normal release from masking.

Subsequently, Bacon *et al.* (1998) reported that some listeners with hearing loss obtained less release from temporally fluctuating maskers than did normal-hearing listeners with and without simulated hearing loss. They evaluated listeners' understanding of sentences in speech-shaped noise that was modulated by the envelope of one of the following: steady-state noise, multitalker babble, single-talker babble, and a 10-Hz square wave with 100% modulation depth. They observed that for normal-hearing listeners, the square-wave modulation provided the greatest release from masking. In addition, they found that the impaired listeners obtained significantly less release from masking than did their normal-hearing counterparts. Noise-masked normal-hearing listeners obtained somewhat less masking release than they had with full access to the signals. However, six of the 11 impaired listeners obtained significantly less release from masking than did their counterparts with simulated hearing loss. They concluded that audibility accounts for some loss of masking release, but additional factors, such as excessive forward masking in impaired ears, may account for the additional loss of masking release.

More recently, Dubno, Horowitz, and Ahlstrom (2002) suggested that audibility explained only a small percentage of the variability in older and younger listeners' identification of consonants in modulated noise. Older and younger listeners with normal or near-normal hearing sensitivity were matched for their thresholds using threshold-matching noise. Significant differences in consonant identification between groups were found. They further noted a significant correlation between forward masking and masking release in older listeners with near-normal hearing sensitivity, suggesting that factors other than audibility can affect masking release.

Kwon and Turner (2001) investigated consonant identification in normal-hearing listeners' understanding of implant simulations. They suggested that two opposing factors may influence hearing-impaired listeners' understanding of speech in modulated noise. First, these listeners may benefit from the same release from masking that is observed in normal-hearing subjects. As a result, their performance may improve when noise is modulated rather than constant. Second, listeners with hearing loss may be negatively affected by modulation masking because listeners with reduced spectral resolution rely on natural amplitude modulations for speech recognition (e.g., Hedrick and Jesteadt, 1996; Hedrick and Carney, 1997). The modulated noise may actually interfere with the acoustic envelope cues, at syllabic or segmental levels, that are used by the listener with hearing loss. Kwon and Turner (2001) evaluated the effects of modulated noise on understanding spectrally impoverished signals (12-band noise simulations). Their listeners apparently experienced a mix of masking release and modulation masking. When the signals and/or the maskers were bandlimited, they found that midfrequency modulated maskers provided the listeners some masker release, resulting in improved consonant recognition when compared to unmodulated maskers. In contrast, a modulated high-frequency masker sometimes caused reduced consonant identification when compared to an unmodulated masker. They concluded that high-frequency modulated maskers can cause some interference in consonant recognition that may offset any benefit provided by the masking release.

Listeners with cochlear implants have well-documented difficulties understanding speech in steady noise (e.g., Fu *et al.*, 1998). Most realistic noise, however, is fluctuating in nature, and a listeners' ability to ''listen in the dips'' is important for communication in these realistic environments. It is not known whether listeners with cochlear implants obtain release from masking when listening in fluctuating noise. If Kwon and Turner's (2001) hypothesis is true, that modulated maskers can cause both masking release and modulation masking (interference), then listeners with cochlear implants may not benefit from masker temporal fluctuations. Instead, implant listeners who use speech processors with envelope-extracting processor algorithms may be adversely affected by fluctuating maskers like individual competing talkers. A lack of masking release might cause additional difficulty in day-to-day situations.

The current experiment was designed to evaluate the ability of cochlear implant listeners to take advantage of temporal gaps in background noise. Listeners with implants were compared to listeners with normal hearing sensitivity for the understanding of sentences in background noise, when the noise was either steady or square-wave modulated across a range of modulation frequencies.

II. METHODS

A. Subjects

Subjects were eight young adult listeners with normal hearing sensitivity who listened to typical full-spectrum speech (normal group), eight additional young adult listeners with normal hearing sensitivity who listened to implant simulations (simulation group), and nine adult listeners with hearing loss who were cochlear implant users (implant group). Characteristics of listeners in the implant group are shown in Table I. All implant users were postlingually deafened. Their mean age was 49 years (range: 34 to 64 years), and their average length of deafness prior to implantation was 16 years (range 1 to 44 years). All listeners had worn their implants for more than 2 years (mean: 5 years, range 2 to 11 years) and derived significant benefit from their devices. As shown in Table I, three listeners used the Nucleus 22 device with a spectral-peak (SPEAK) speech-processing strategy, three used the Clarion 1.2 device with a continuous interleaved sampling (CIS) strategy, and three used the Clarion HiFocus device with a CIS strategy. Listeners in the

TABLE I. Summary of subject characteristics.

Listener	CI/Processor	Age at test	Age at onset of deafness	Age at implantation
N12	Nucleus 22/SPEAK	53	32	42
N14	Nucleus 22/SPEAK	58	49	50
N32	Nucleus 22/SPEAK	34	5	29
C02	Clarion 1.2/CIS	42	18	37
C03	Clarion 1.2/CIS	53	22	49
C05	Clarion 1.2/CIS	47	42	43
C14	Clarion HiFocus/CIS	64	16	60
C15	Clarion HiFocus/CIS	42	33	40
C16	Clarion HiFocus/CIS	48	29	43

implant group used their own speech processors with typical sensitivity and volume settings, and no noise reduction. At the beginning of each session, the users set the sensitivity and/or volume controls while listening to practice lists, and they were instructed not to change the settings during the test session.

B. Stimuli

Speech stimuli consisted of IEEE (1969) sentence materials spoken by five male and five female talkers. Stimuli were recorded on digital audio tape at 44 kHz. They were digitized, downsampled to 20k samples per second, and normalized for long-term rms amplitude using COOLEDIT PRO©. Sentences contained an average of five key words. Blocks of ten sentences were presented, each block containing one sentence spoken by each talker, in random order.

Noise stimuli were generated in real time using the Tucker-Davis waveform generator (TDT WG1). The noise was passed through a Rane 30-band equalizer so that the spectrum of the resulting noise matched the long-term spectrum of the IEEE sentences. Noise stimuli were presented either continuously (steady), or gated with 2-ms cosine-squared ramps. Gating was implemented with 50% duty cycles and 100% modulation depths. Six gate frequencies ranged from 1 to 32 Hz, resulting in noise bursts that ranged in duration from 16 ms (32 Hz) to 500 ms (1 Hz). Signal-to-noise ratios (SNRs) were computed based on the long-term rms of the noise and the speech. SNRs were +16, +8, 0, −8, or −16 dB, depending upon the listener and the condition.

Sentences from two talkers were modified to create four-channel simulations of implant processing. Sentences were filtered into four narrow bands (after Shannon *et al.*, 1995): 100–300, 300–500, 500–1700, and 1700–6000 Hz. The envelope of each filter output was extracted and narrow-band noises of the same frequency region were modulated by the respective envelope (low-pass filtered at 500 Hz).

C. Test procedures

Listeners were seated in the center of a sound-treated chamber. Speech signals were delivered diotically through two Bose 301 speakers at an overall level of 65 dBA. Speech stimuli were presented in blocks of ten sentences, using all ten talkers in random order for each list. All SNR and gating conditions were randomized prior to the beginning of each subject's testing. The listeners responded verbally to each

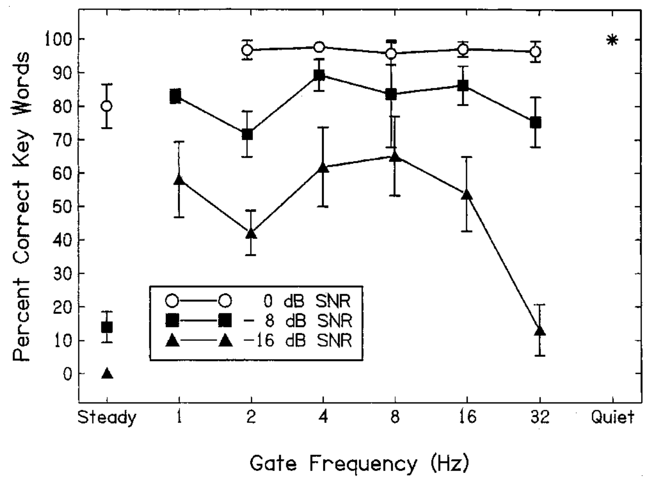

FIG. 1. Average percent-correct key word identifications are shown as a function of noise gate frequency for normal-hearing listeners at SNRs of 0, −8, and −16 dB. Error bars indicate one standard deviation from the mean.

sentence, and the experimenter scored the key words correct for each sentence, circling the correct answers on an answer form. Each listener's results (percent-correct key words) for each condition were later entered into computer files.

Key word identification was evaluated in steady and gated noise. On each trial, the masking noise started initially, with the sentence beginning after a random delay that ranged from 10 to 100 ms. The noise was either steady or gated. The level of the noise varied depending upon the condition being tested. Listeners in the implant and simulation groups heard the noise at +8 and +16 dB SNR. (Listeners in the simulation group also heard the noise at 0 dB SNR. Pilot testing with three high-performing implant listeners indicated that performance was near 0% for all gate conditions at 0 dB SNR and lower.) Listeners in the normal group heard the noise at 0, −8, and −16 dB SNR. All listeners also completed two blocks of sentences in quiet.

III. RESULTS

A. Normal group

Results from listeners with normal hearing sensitivity are shown in Fig. 1. These listeners were able to repeat nearly 100% of the key words in quiet. When the SNR was 0 dB, they obtained scores of approximately 80% correct for steady noise, and near 100% for all gated noise conditions. When the SNR was −8 dB, their performance in steady noise was only 10% correct key words, while in gated noise their mean performance ranged from 70% to 90% correct. When the SNR was −16 dB, they scored 0% correct in steady noise, with average gated noise performance ranging between 15% and 65% correct. Performance was dependent upon the gate frequency. Although there was considerable variability among listeners in the normal group, release from masking was maximal for gate frequencies between 1 and 16 Hz, and was reduced at gate frequencies at or above 16 Hz.

Figure 2 shows the normal group's masking release, or improvement in scores for gated vs steady noise, for the three SNR conditions. For SNR of 0 dB, improvement from gating was approximately 20% for all gate frequencies and

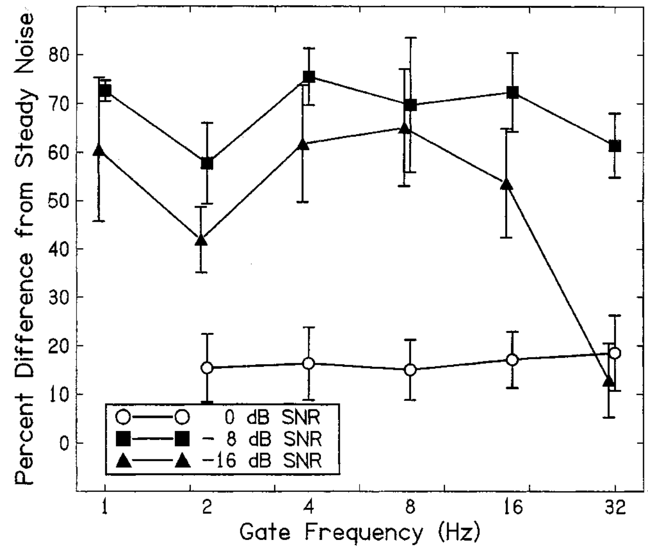

FIG. 2. Average percent improvement from steady noise is shown as a function of noise gate frequency for normal-hearing listeners, at SNRs of 0, −8, and −16 dB. The amount of release from masking is especially large for the SNR of −8 dB.

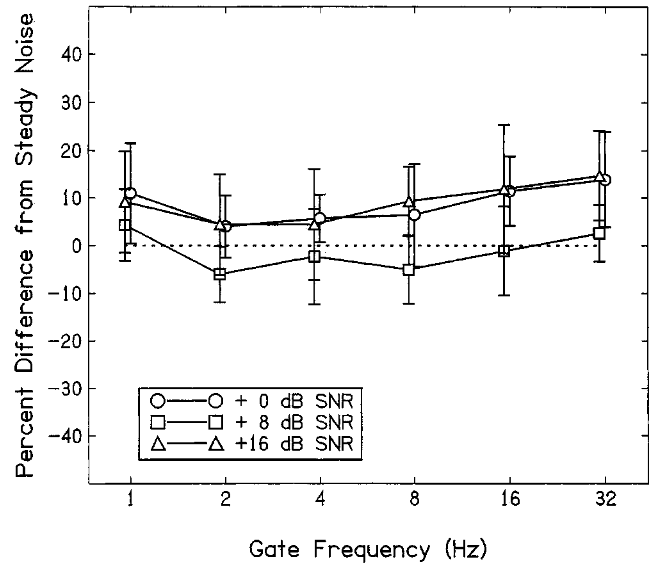

FIG. 4. Average percent improvement from steady noise is shown as a function of noise gate frequency for simulation listeners, at SNRs of 0, +8, and +16 dB. Little release from masking is seen for simulation listeners.

was limited by a ceiling effect for the gated conditions (see Fig. 1). The maximum release from masking occurred at the SNR of −8 dB, with improvement ranging from 60%–80%. Masking release at −8 dB SNR was relatively independent of gate frequency, with a possible minimum at 2 Hz. For −16 dB SNR, release from masking ranged from 10%–60% and was strongly affected by gate frequency. Normal-hearing listeners' release showed the same apparent minimum at 2 Hz and was reduced for very fast (32 Hz) gate frequencies.

B. Simulation group

Results from listeners in the simulation group are shown in Fig. 3. The stimuli for these listeners were 4-band modulated noise replicas of the IEEE sentences from one talker. Their mean key word identification score in quiet was approximately 55%, indicating that these listeners showed performance somewhat typical of implant listeners. Mean scores in steady noise dropped to approximately 30% correct for the SNR of +16 dB, 25% for SNR of +8 dB, and 5% correct for SNR of 0 dB, suggesting that all levels of noise had a significant negative effect on word understanding. When the noise was gated, mean scores for an SNR of +16 dB were near 40% correct for all gate frequencies, still poorer than the mean score correct in quiet and only slightly better than their performance in steady noise. Mean scores in gated noise at +8 and 0 dB SNR showed a similar pattern; scores in gated noise were very close to those in steady noise and were independent of gate frequency.

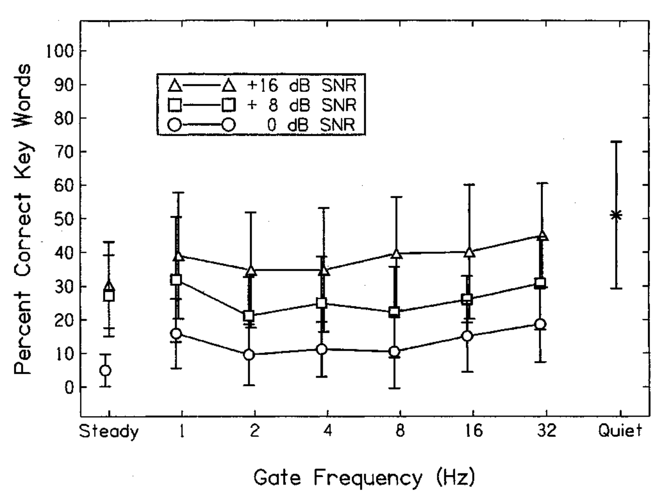

FIG. 3. Average percent-correct key word identification is shown as a function of noise gate frequency for simulation group normal listeners for SNRs of 0, +8, and +16 dB.

The data from the simulation group are replotted in Fig. 4 showing release from masking, or percent improvement in scores for gated versus steady noise for their three SNR conditions. For conditions with an SNR of +16 and 0 dB, the gated noise provided a slight benefit over the steady noise, except perhaps for the fastest gate rates. For a +8-dB SNR, no masking release was observed. No effect of gate frequency was seen. An analysis of variance (ANOVA) indicated that there was a significant effect of SNR [$F(1,10) = 18.91$, $p < 0.01$], but no significant effect of gate frequency [$F(1,10) = 1.43$, $p > 0.05$].

C. Implant group

Results from listeners in the implant group are shown in Fig. 5. Their mean key word identification score in quiet was 80%, indicating that these listeners were successful implant users. Mean scores in steady noise dropped to 60% correct for an SNR of +16 dB, and to 35% correct for an SNR of +8 dB, suggesting that both levels of noise had a significant negative effect on word understanding. When the noise was gated, mean scores for an SNR of +16 dB ranged from 55% to 65% correct, still significantly poorer than the mean 80% correct in quiet and not different from their performance in steady noise.

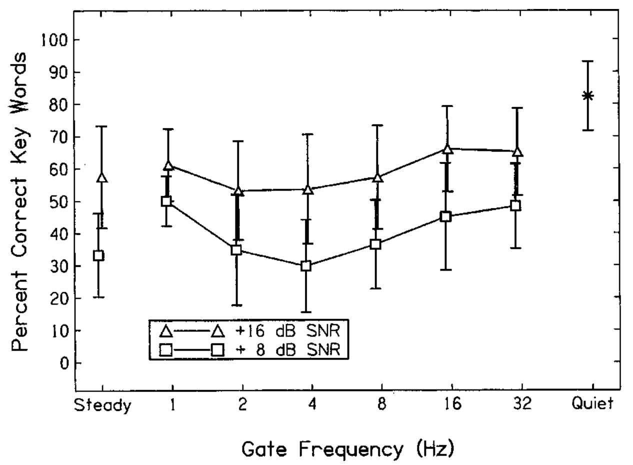

FIG. 5. Average percent-correct key word identification is shown as a function of noise gate frequency for listeners with cochlear implants for SNRs of +8 and +16 dB.

The data from the implant group are replotted in Fig. 6 showing release from masking, or percent improvement in scores for gated versus steady noise for their two SNR conditions. For conditions with an SNR of +16 dB, the gated noise provided little benefit over the steady noise, except perhaps for the fastest gate rates. For the +8-dB SNR condition, performance was slightly better for the slowest and fastest gate frequencies than for the steady noise, with a minimum in masking release seen at 2-, 4-, and 8-Hz gate frequencies.

The amount of masking release obtained by the different listener groups was compared using analysis of variance. To determine whether normal listeners had significantly more masking release than the other groups, results had to be compared at different SNRs because normal listeners were tested at SNRs that were different from the other two groups. Masking release results (pooled across gate frequencies between 2 and 16 Hz) were compared at −8-dB SNR for the

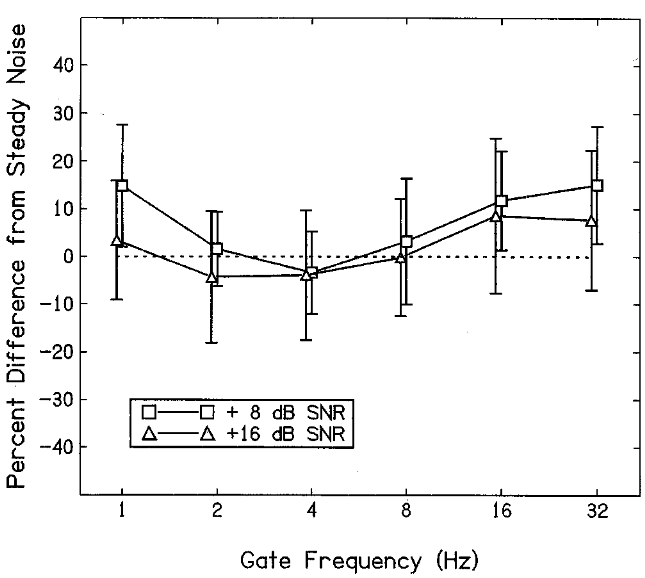

FIG. 6. Average percent improvement from steady noise is shown as a function of noise gate frequency for listeners with cochlear implants, at SNRs of +8 and +16 dB. Little release from masking is seen for cochlear implant users.

normal group, +8-dB SNR for the simulation group, and +8-dB SNR for the implant group. Results indicated that normal listeners obtained significantly more masking release than did implant and simulation listeners [$F(2,18)=46.4$, $p<0.0001$]. More detailed comparisons were possible between simulation and implant groups. Repeated measures analysis of variance was applied to the masking release data for the two groups (simulation and implant), two SNRs (+8 and +16 dB), and six gate frequencies. No significant difference between groups was noted [$F(1,12)=0.3$, $p>0.05$], suggesting that the implant and simulation listeners had similar release from masking. A significant effect of SNR was noted [$F(1,12)=6.1$, $p<0.05$] with no significant group by SNR interaction [$F(1,12)=4.3$, $p>0.05$]. A significant effect of gate frequency was also noted [$F(4,48)=7.6$, $p<0.01$]; however, there was a significant gate frequency by group interaction [$F(4,48)=3.6$, $p<0.05$]. No higher-order interactions were significant.

As noted in the previous section, for the simulation group, no significant effect of gate frequency was found. Analysis of the implant group indicated that the effect of gate frequency on masking release approached, but did not reach significance [$F(1,10)=4.68$, $p=0.056$] across both SNRs. Multiple regression analysis indicated that gate frequency accounted for 32% of the overall variance in implant listeners' performance, while SNR accounted for 16%. Both gate frequency and SNR accounted for 47% of the variance in implant listeners' masking release. When the gated noise was presented at +8-dB SNR, mean scores ranged from 35% to 50%. Some improvement over steady noise was seen at the slowest (1 Hz) and fastest (16 and 32 Hz) gate frequencies, but performance remained low at moderate gate frequencies (2 to 8 Hz). Paired *t*-tests for the data from the +8-dB SNR condition indicated that mean performance in 1-Hz gated noise (500-ms alternating cycles of noise and silence) was significantly better than performance in steady noise ($t[6]=-2.72$, $p=0.017$). Performance in 16-Hz ($t[7]=-3.7$, df=7, $p=0.0037$) and 32-Hz ($t[7]=-3.26$, $p=0.007$) gated noises were significantly better than performance in steady noise. When corrected for multiple comparisons, these individual comparisons retain their significance.

Figure 7 shows mean data for the subgroups of implant listeners with different devices. This figure shows the improvement in performance for listeners divided by implant type for gated vs steady noise at an SNR of +8 dB. Clearly, although there were overall performance differences between listeners, the trend was that all listeners obtained minimal to no benefit of gated noise over steady noise for all devices, with an apparent minimum in performance at gate frequencies around 4 Hz. This suggests that specific characteristics of a given implant device (Nucleus 22 vs Clarion 1.2) or speech processing strategy (SPEAK vs CIS) were not primarily responsible for implant listeners' failure to demonstrate release from masking. Examination of individual data functions revealed that only one implant listener did *not* show the characteristic minimum performance near the 4-Hz gate frequency. That listener showed a relatively flat performance function for 1–8-Hz gate frequencies, with increased masking release at 16 to 32-Hz gate frequencies. All other implant

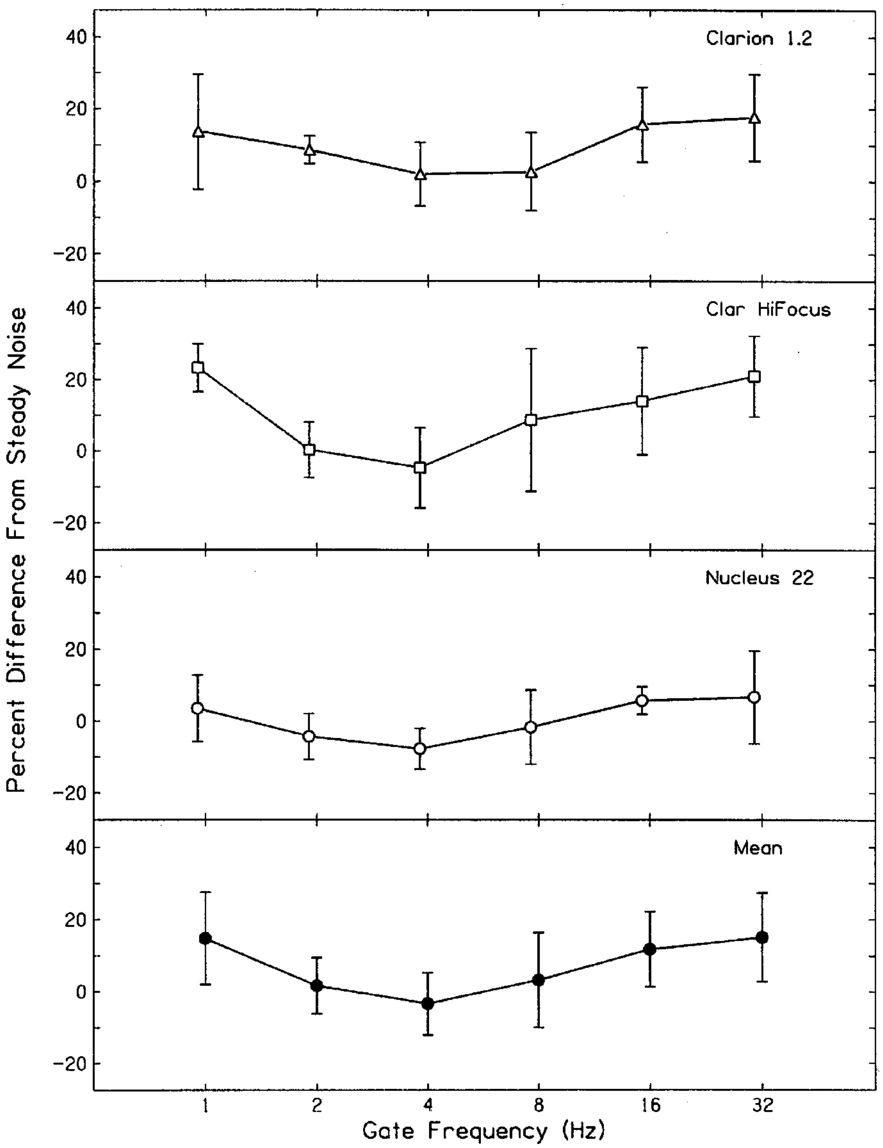

FIG. 7. Average percent improvement from steady noise is shown as a function of noise gate frequency for listeners with different implant devices. Users with Clarion 1.2 processors using CIS strategy are shown with open triangles; users with Clarion Hi-Focus processors are shown with open squares; users with Nucleus N22 processors using the SPEAK strategy are shown with open circle symbols. The overall mean performance for all implant users is shown with filled circles. No meaningful differences between processor types can be seen.

listeners showed a minimum in the performance function at 2 or 4 Hz, with improved performance at slower and faster gate frequencies.

IV. GENERAL DISCUSSION

A. Normal group

As expected, listeners in the normal group obtained significant release from masking from the gated maskers. The greatest amount of masker release was obtained for SNRs of −8 dB, at which the speech signals were an approximate 65 dB A and the noise was 73 dB A. At those levels, the words were very difficult to hear in steady noise, and mean performance was approximately 10% correct. When the noise was gated, however, performance improved considerably to mean levels of approximately 80% correct for gate frequencies of 4 Hz or higher. Presumably, some minimal amount of speech information was audible in the presence of the steady noise because performance was better than chance (mean scores were approximately 10% correct). When parts of the signals were made fully audible during silent intervals in the gated noise, performance improved considerably. Clearly, listeners were using bits of information to fill in the message, and as a result they were understanding a majority of the key words. The IEEE sentences that were used in this investigation have been shown to have relatively small linguistic context effects (Nittrouer and Boothroyd, 1990). Nevertheless, the partial acoustic and linguistic cues obtained by the normal group were used to understand most of the key words.

The results were somewhat different for the normal group at an SNR of −16 dB, when the speech and noise signals were at 65 and 81 dBA, respectively. At these levels, none of the key words was identifiable in steady noise. Gated noise maskers again provided listeners with significant masker release, but in this case, the amount of release was related to the masker's gate frequency. At the slowest gate frequencies (1–2 Hz, corresponding to alternating 500- or 250-ms periods of noise and silence) approximately 40%–50% of the key words were identified. In this condition, whole words and syllables were presumably completely inaudible, and listeners were unable to extract more than 50% of the information. However, at 4- and 8-Hz gate frequencies (125- and 62-ms periods), parts of many syllables and words were probably audible. Listeners used these parts to identify approximately 60% of the key words. At faster gate frequencies, the periods of silence were only approximately 30 and 15 ms in duration. Because the noise levels significantly exceeded the level of the speech signals, we presume that some forward masking occurred, at least partially obscuring the speech during these short silent intervals. At the gate frequency of 32 Hz, performance was greatly reduced to a mean score of 15%.

B. Implant and simulation groups

Listeners with cochlear implants and normal-hearing listeners responding to implant simulations were much more affected by background noise than were normal-hearing listeners. Initial pilot results had suggested that none of the best implant users could understand any key words at a 0-dB SNR for either gated or continuous noise. As a result, implant and simulation listeners were tested at SNRs that were different from those used with the normal group. Even though the implant listeners demonstrated very good performance in quiet (around 80% for difficult stimuli), they were greatly affected by noise. Even at the favorable SNR of +16 dB, performance dropped by more than 20%. At an SNR of +8 dB, an SNR typical of many environmental situations, key word identification dropped by about 50%. These results indicate that when context is low and speakers unfamiliar, even low levels of background noise affect implant listeners substantially.

Noise affected simulation listeners in a similar way. Simulation group listeners understood approximately 55% of words in quiet, somewhat poorer than the implant group results, but typical of some implant listener performance. At the favorable SNR of +16 dB, their performance also dropped by more than 20%, indicating a significant effect of the steady background noise.

It seems likely that noise (modulated or steady) disrupts the ideal amplitude envelope cues that are coded by the implant processors. Even when the noise occurred 16 dB below the speech signal, one can imagine that the random envelope of the noise could disrupt natural envelope cues extracted by the implant processor. This may explain the significant drop in performance from quiet to steady noise, even at an SNR of +16 dB. However, this does not explain a lack of the ability to use intervals of quiet speech within gated noise to extract some key words.

Interestingly, the implant and simulation group listeners did not show significant masking release from temporal gaps in noise. Simulation group listeners showed very little masking release (10%) for SNRs of +16 and 0 dB, and no masking release for SNRs of +8 dB. No effect of gate frequency was seen, suggesting that listeners responding to four-channel implant simulations do not take advantage of temporal gaps in noise, even when that gap is as long as 500 ms. For implant listeners at an SNR of +16 dB, there was no difference in performance between steady and gated noises at any gate frequency. For an SNR of +8 dB (a condition quite typical of conversational settings), there seems to be some slight release from masking at extremely slow (1 Hz) and fast (16 and 32 Hz) modulation rates, with a minimum in performance between 2 and 8 Hz.

We do not attribute the lack of masking release to either a lack of audibility in the "dips" nor to forward masking. At the SNRs used by the implant and simulation groups the signal level greatly exceeded the level of the maskers. Implant listeners set the sensitivity of their devices so that the quiet sentences were at a comfortable and audible level. When noise was introduced, it was always at a level 8 or 16 dB below the level of the speech. Also, because the simulation group showed a lack of masking release similar to that of the implant group, inaudibility cannot be the primary cause. Clearly those normal-hearing listeners had full access to the sentence information in the temporal dips in noise. Thus, we do not expect that an inability to repeat key words was due to a lack of audibility of the quiet stimuli. Similarly, because of the low-level noise we did not expect, nor did we see, any decrement in performance at the fastest gate frequencies (like that observed at −16 dB SNR and 32 Hz for the normal group) that might be attributed to forward masking.

We had presumed, however, that 250-ms silent intervals (the 2-Hz gating condition) would be sufficient for at least some implant listeners to identify some key words. Because initial pilot data had shown no masker release even at 2 Hz, the 1-Hz condition was added. Based on the results for the 1-Hz condition, it seems that most implant listeners were able to take advantage of 500-ms silent intervals to identify some key words, at least for the 8-dB SNR condition. Eight of nine individual implant users showed some release from masking at 1-Hz gate frequency. It was surprising that one remaining implant listener and all simulation listeners did not show significant word understanding with silent intervals as long as 500 ms in the noise. None of the implant or simulation group listeners could take advantage of 250-ms silent intervals to identify at least some key words. In fact, performance was the same for steady noise and for maskers with 2-, 4-, and 8-Hz gate frequencies.

One logical explanation for this effect is that gated maskers at those syllabic-like rates were actually a distraction or interference, rather than a benefit to the implant listener. In fact, some implant group users reported anecdotally that the gated noise mixed with the sentences sounded like additional syllables, perhaps in another language. When the gated noise was presented alone, one listener described the noise appropriately as bursts of noise at slow modulation rates, and as "fluttering" noise at faster rates. This confirmed that gaps in the noise were perceived by the listeners. However, when the noise was mixed with the speech at moderate modulation rates, he reported that he heard it as a strange competing talker. This would support the Kwon and Turner (2001) hypothesis that gated maskers can provide some release (seen here at 1-, 16-, and 32-Hz gating for 8-dB SNR) as well as some interference (seen here at 2-, 4-, and 8-Hz gating for 8-dB SNR maskers).

There seems to be no significant performance difference between users of different implant devices or speech-processing algorithms, at least among the pulsatile strategies evaluated here (CIS and SPEAK). Also, there was very little difference in performance between the implant and simulation groups. Thus, the specific processing characteristics of the implant devices such as the processing algorithm, the number of electrodes stimulated, the automatic gain control, or the range of acoustic amplitudes encoded (input dynamic range), do not seem to account for the lack of masking release. Listeners' performance was not apparently restricted by implant processing hardware. The implant processor was providing them with the temporal envelope information at sufficiently high (at least 250 Hz) rate (Kwon, 2002).

Listeners with both devices (Clarion and Nucleus) showed the minimum in performance at gate frequencies between 2 and 8 Hz. The lack of release from masking is apparently, then, not related to characteristics of the implant devices themselves.

In addition, it seems unlikely that these results can be explained on the basis of abnormal forward masking of the implant users. Previous studies (e.g., Nelson and Donaldson, 2001) have suggested that most cochlear implant users demonstrate rapid-recovery time constants of less than 7 ms. That rapid recovery should allow implant users to take advantage of temporal gaps in the noise that were as long and 250 and 500 ms in some conditions. Performance functions from the implant group (seen in Figs. 3 and 4) do not show the characteristic shape seen in the data from the normal group listeners (Figs. 1 and 2). While listeners from the normal group show decreased release from masking as modulation rate increases, the listeners from the implant group do not. The decrease in benefit from modulation noise at rapid modulation rates (temporal gaps <30 ms) can be attributed to forward masking perceptually filling the gaps for the normal-hearing listeners. Forward masking, then, cannot explain the relatively flat functions of the implant group.

It may be more likely that the implant and simulation listeners receive such an impoverished spectral code that they are unable to integrate the speech information into a well-defined auditory image, or to segregate the speech signal from the background noise. Because of the limited acoustic cues available to the listener, a longer temporal ''glimpse'' of the quiet signal (longer than 500 ms) is needed before the speech stream can be integrated and whole words extracted. Additional testing of auditory stream segregation by implant users is warranted and is underway. In addition, further evaluation of the role of spectral resolution (increased numbers of spectral channels) in masking release is underway and will be reported in a companion paper.

These results suggest that listeners with cochlear implants are likely to be extremely disrupted in acoustic situations with a single competing talker, even when the level of the talker's voice is significantly less than the target signal. Similarly, they may be quite affected by a reverberant room where the envelope cues are disrupted by echoes. Further study is needed to explain and understand these results.

V. CONCLUSIONS

Although normal-hearing listeners are able to obtain release from masking from modulated noise, listeners with cochlear implants cannot. Implant and simulation listeners are significantly affected by background noise, even at very favorable signal-to-noise ratios. When noise is modulated, even with 250-ms silent intervals, implant and simulation listeners are unable to take advantage of a silent gap to extract meaningful words. Performance of implant users seems poorest at modulation frequencies between 2 and 8 Hz, encompassing rates corresponding to syllables and words. These results imply that modulation interference, or masking, may be responsible for the lack of masking release in the implant group listeners. Performance does not seem to vary with implant device or processing strategy, and may be due to a disruption in the envelope cues extracted by the devices and used by the listeners. The lack of masking release, then, may be attributable to general characteristics of the implant processing, including the lack of spectral information in the processed signal. Implant listeners may have noticeable difficulty in situations with fluctuating noise, such as in restaurants or with single competing talkers.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. DC00110, by the Lions 5M International Hearing Foundation, and by the University of Minnesota. The authors wish to thank Gail Donaldson and Benjamin Munson for their assistance on early versions of this manuscript.

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 Hear. Res. **41**, 549–563.

Dubno, J., Horwitz, A., and Ahlstrom, J. (**2002**). ''Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing,'' J. Acoust. Soc. Am. **111**, 2897–2907.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). ''Speech recognition in amplitude modulated noise of listeners with normal and impaired hearing,'' J. Speech Hear. Res. **38**, 222–233.

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuation noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

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.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). ''Masking of speech by amplitude-modulated noise,'' J. Acoust. Soc. Am. **95**, 518–529.

Hedrick, M. S., and Carney, A. E. (**1997**). ''Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants,'' J. Speech Lang. Hear. Res. **40**, 1445–1457.

Hedrick, M. S., and Jesteadt, W. (**1996**). ''Effect of relative amplitude, presentation level, and vowel duration on perception of voiceless stop consonants by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **100**, 3398–3407.

IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' IEEE Trans. Audio Electroacoust. **17**(3), 225–246.

Kwon, B. J. (**2002**). Personal communication.

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.

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.

Nittrouer, S., and Boothroyd, A. (**1990**). ''Context effects in phoneme and word recognition by young children and older adults,'' J. Acoust. Soc. Am. **87**, 2705–2715.

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

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.

Trine, T. D. (**1995**). ''Speech recognition in modulated noise and temporal resolution: Effects of listening bandwidth,'' Unpublished doctoral dissertation, University of Minnesota, Twin Cities.