Masking release for hearing-impaired listeners: The effect of increased audibility through reduction of amplitude variability

Joseph G. Desloge, Charlotte M. Reed,[a] Louis D. Braida, Zachary D. Perez, and Laura A. D'Aquila
Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

(Received 13 July 2016; revised 5 April 2017; accepted 23 May 2017; published online 15 June 2017)

The masking release (i.e., better speech recognition in fluctuating compared to continuous noise backgrounds) observed for normal-hearing (NH) listeners is generally reduced or absent in hearing-impaired (HI) listeners. One explanation for this lies in the effects of reduced audibility: elevated thresholds may prevent HI listeners from taking advantage of signals available to NH listeners during the dips of temporally fluctuating noise where the interference is relatively weak. This hypothesis was addressed through the development of a signal-processing technique designed to increase the audibility of speech during dips in interrupted noise. This technique acts to (i) compare short-term and long-term estimates of energy, (ii) increase the level of short-term segments whose energy is below the average energy, and (iii) normalize the overall energy of the processed signal to be equivalent to that of the original long-term estimate. Evaluations of this energy-equalizing (EEQ) technique included consonant identification and sentence reception in backgrounds of continuous and regularly interrupted noise. For HI listeners, performance was generally similar for processed and unprocessed signals in continuous noise; however, superior performance for EEQ processing was observed in certain regularly interrupted noise backgrounds.

© *2017 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4985186]

[VMR] Pages: 4452–4465

I. INTRODUCTION

A major complaint of hearing-aid and cochlear-implant users is the difficulty of understanding speech in competing backgrounds. This problem may be compounded by the reduced or even absent ability of hearing-impaired (HI) listeners, relative to that of normal-hearing (NH) listeners, to take advantage of temporal fluctuations in background interference to understand the target speech (e.g., Festen and Plomp, 1990; Moore *et al.*, 1999; Lorenzi *et al.*, 2006; Bernstein and Grant, 2009; Desloge *et al.*, 2010; Léger *et al.*, 2015). The release from masking observed in fluctuating noise for NH listeners has been described in terms of a decrease in the signal-to-noise ratio (SNR) required to understand the target speech at a given level of performance (e.g., Festen and Plomp, 1990), or as an increase in performance (Lorenzi *et al.*, 2006) for a fluctuating noise background compared to continuous noise at the same long-term root-mean-square (rms) level. For NH listeners, the size of this effect can be as great as a 15–25 dB reduction in SNR for sentence reception (George *et al.*, 2006; Rhebergen *et al.*, 2006) or a 30–65 percentage point increase in consonant identification (e.g., Füllgrabe *et al.*, 2006; Lorenzi *et al.*, 2006), depending on various characteristics of the fluctuating noise. Release from masking has also been shown to be dependent on speech-to-noise ratio (SNR), with a tendency to increase with a decrease in SNR (Bernstein and Grant, 2009; Oxenham and Simonson, 2009). For HI listeners, these effects are greatly reduced if not absent. For example, Festen and Plomp (1990) observed essentially no difference in SNR for 50%-correct sentence reception for HI listeners, compared to improvements on the order of 4 to 8 dB for NH listeners in a variety of fluctuating noise backgrounds.

The release from masking experienced by NH listeners has been attributed to the ability to perceive audible glimpses of the target speech during dips in the fluctuating noise (Cooke, 2006). Following from this explanation, release from masking may be reduced in HI listeners due to a lack of audibility of the signal present during dips in the noise. That is, the increased thresholds of HI listeners prevent them from listening in the dips where interference is relatively weak. A more recent alternative explanation of reduced masking release in HI listeners is based on their reduced sensitivity to inherent modulations in continuous noise. According to this theory (Stone *et al.*, 2012; Stone and Moore, 2014), the source of masking in a continuous Gaussian noise is modulation masking arising from inherent fluctuations in the noise. Better performance in an interrupted noise (leading to masking release) arises due to a reduction in modulation masking and not because of increased energy of the signal in the dips of an interrupted noise. According to this hypothesis, the reason that HI listeners do not experience a similar amount of masking release is because they are less sensitive to the fluctuations

[a] Electronic mail: cmreed@mit.edu

in the continuous noise to begin with, and thus do not experience release from it in the presence of an interrupted noise. In fact, the additional fluctuations in the interrupted noise could make performance worse than in the case of a continuous background noise. The reduced sensitivity of HI listeners to the fluctuations in the continuous noise is explained by Oxenham and Kreft (2014) as arising from reduced frequency selectivity, which leads to a flattening of the temporal envelope.

The work reported here is concerned with further exploration of the role of audibility in the reception of speech in temporally fluctuating background noise. This research derives from previous studies which suggest that the amplitude variations present in the speech signal may play a role in determining the size of masking release in HI listeners (Léger *et al.,* 2015; Reed *et al.*, 2016). Specifically, by reducing amplitude variation for speech in interrupted noise, the audibility of lower-energy speech present in the noise dips is increased leading to higher intelligibility.

Léger *et al.* (2015) examined the ability of HI and NH listeners to identify consonants in backgrounds of continuous and square-wave modulated noise for unprocessed speech and for speech that was processed using the Hilbert transform to convey either envelope (ENV) or temporal fine-structure (TFS) cues. The ENV speech conveys the slowly varying changes in signal amplitude while being stripped of the rapidly varying changes in temporal fine structure. Conversely, the TFS speech is stripped of the slowly varying amplitude changes while retaining the rapidly varying changes in temporal fine structure (see Gilbert and Lorenzi, 2006 for details of the processing). Léger *et al.* (2015) studied masking release [defined as the difference in consonant identification scores (in percent correct, PCT) between fluctuating and continuous noise, and denoted as $\mathrm{MR_{PCT}}$] for unprocessed speech, for a 40-band ENV signal, and for TFS signals created from wide-band speech or speech that was bandpass filtered into four logarithmically spaced bands. For the HI listeners, $\mathrm{MR_{PCT}}$ was generally negligible for unprocessed speech and ENV speech, both of which contain naturally occurring variations in the amplitude envelope. On the other hand, positive $\mathrm{MR_{PCT}}$ was observed for both types of TFS speech, in which amplitude variation had been removed. One fault should be pointed out: the positive $\mathrm{MR_{PCT}}$ was due primarily to a decrease in performance in continuous noise rather than to an increase in performance in fluctuating noise.

Similar effects of positive $\mathrm{MR_{PCT}}$, again tempered by a corresponding decrease in continuous-noise performance compared to unprocessed speech, were observed by Reed *et al.* (2016) for two other types of processing which effectively removed variations in amplitude of the speech signal. These methods included infinite peak clipping to achieve a reduction in amplitude variation similar to that present in TFS speech and a 40-band envelope signal that was reprocessed with Hilbert-transform-based TFS processing to remove the amplitude variations.

The practicality of the techniques that yielded positive $\mathrm{MR_{PCT}}$ in both Léger *et al.* (2015) and Reed *et al.* (2016) is limited by the fact that the increase in $\mathrm{MR_{PCT}}$ resulted from a decrease in speech intelligibility in continuous noise as opposed to an increase in intelligibility in interrupted noise. This decrease in continuous-noise speech intelligibility is largely due to distortions introduced by TFS processing and peak clipping. This paper investigates whether it is possible to reduce amplitude variation (and increase audibility of speech in interrupted noise) without introducing disruptive distortions. This would increase release from masking by increasing performance in interrupted noise while maintaining performance in continuous noise. A signal-processing technique was developed to achieve these aims and results are presented for consonant identification and sentence reception in backgrounds of continuous noise, square-wave interrupted noise, and sinusoidally amplitude-modulated noise. The experiments employed listeners with sensorineural hearing loss as well as those with normal hearing.

II. GENERAL METHODOLOGY

The experimental protocol for testing human subjects was approved by the internal review board of the Massachusetts Institute of Technology. All testing was conducted in compliance with regulations and ethical guidelines on experimentation with human subjects. All listeners provided informed consent and were paid for their participation in the experiments.

Aspects of the methods that apply to both the consonant-identification experiments (Sec. III) and the sentence-reception tests (Sec. IV) are described below.

A. Participants

Seven listeners with bilateral, symmetric, moderate-to-severe sensorineural hearing loss (5 male–M and 2 female–F) participated in these experiments. They were all native speakers of American English and ranged in age from 20 to 75 years (mean age of 37 years). All but one (HI-10) had participated in a previous study of consonant identification in noise conducted by Léger *et al.* (2015) and these six were assigned the same listener number as used in that paper. The five-frequency (0.25, 0.5, 1, 2, and 4 kHz) audiometric pure-tone average (PTA) ranged from 27 dB hearing level (HL) to 75 dB HL across listeners (and averaged 48.3 dB HL). Pure-tone thresholds in dB sound-pressure level (SPL) for each HI listener are shown in Fig. 1, where listeners are arranged in order of increasing PTA. These measurements were obtained with Sennheiser HD580 headphones for 500-ms stimuli in a three-alternative forced-choice adaptive procedure which estimates the threshold level required for 70.7%-correct detection (see Léger *et al.*, 2015). The panel of results for each HI listener in Fig. 1 includes thresholds for the left and right ears, test ear, listener's age, PTA in dB HL, and the speech levels and SNRs employed in experiment 1. The ear with the better PTA was selected for use in the monaural experiments reported below.

Four listeners (all M) with normal hearing (defined as 15 dB HL or better in the octave frequencies between 250 and 8000 Hz) also participated in the study. They were native speakers of American English, ranged in age from 22 to 53 years (mean age of 35 years), and had participated

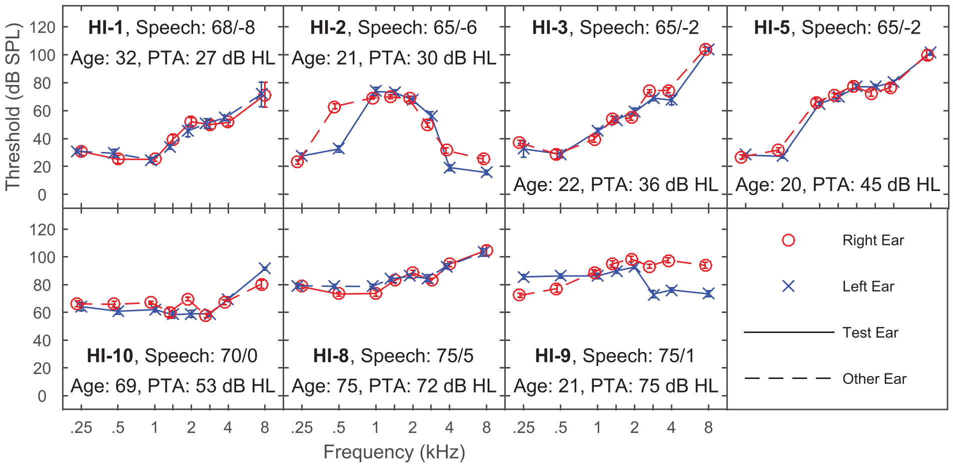

FIG. 1. (Color online) Detection thresholds in dB SPL as a function of frequency in kHz for seven HI listeners. Thresholds were measured using 500-ms tones in a three-alternative, forced-choice, adaptive procedure. Also provided in the panels for each listener are the speech level in dB SPL prior to amplification and the signal-to-noise ratio (SNR) in dB used in the consonant tests (e.g., 68/−8), age in years, and pure-tone average in dB HL (averaged over the five octave frequencies between 0.25 and 4.0 kHz).

in previous studies of speech identification in noise. A test ear that met the audiometric criteria was selected for each listener (2 left ear and 2 right ear).

B. Signal processing to reduce amplitude variation with minimal distortion: EEQ processing

Energy Equalization (EEQ) is a homogeneous signal processing technique that was developed to reduce amplitude variation while introducing minimal distortion to the audio signal. The general concept is very simple: create running estimates of short-term (∼10 ms) and long-term (∼1000 ms) energy of the speech-plus-noise stimulus and apply a scale factor to equate short-term to long-term energy. In the presence of continuous noise, the short-term and long-term energies are roughly the same and the processing has little effect (and minimal distortion). In the presence of interrupted noise, the short-term energy oscillates around the long-term energy and so equating short-term to long-term energy reduces amplitude variation and increases the level of the speech in the noise dips, which may result in improved intelligibility.

Energy equalization of a speech-plus-noise signal $x(t)$ is achieved using the following steps.

(1) Form running short- and long-term moving-average estimates of the energy of the speech-plus-noise stimulus,

$$E_{\text{short}}(t) = \text{AVG}_{\text{short}}\left[x^2(t)\right] \tag{1}$$

and

$$E_{\text{long}}(t) = \text{AVG}_{\text{long}}\left[x^2(t)\right], \tag{2}$$

where AVG is a moving-average operator that uses specified short- and long-term time constants to provide an estimate of speech-plus-noise energy.

(2) Determine the scale factor $SC(t)$,

$$SC(t) = \sqrt{\frac{E_{\text{long}}(t)}{E_{\text{short}}(t)}}, \tag{3}$$

where care is taken to prevent dividing by zero during quiet intervals.

(3) Apply the scale factor to the original speech-plus-noise stimulus,

$$y(t) = SC(t)x(t). \tag{4}$$

(4) Form the output $z(t)$ by normalizing $y(t)$ to have the same energy as $x(t)$,

$$z(t) = Ky(t), \tag{5}$$

where K is chosen such that

$$\text{AVG}_{\text{long}}[z(t)] = \text{AVG}_{\text{long}}[x(t)]. \tag{6}$$

The implementation of the technique requires choices of parameters for the moving-average operators and for the scaling term. For the moving-average operations, time constants (λ_{short} and λ_{long}) must be specified taking into account that the short-term average should reflect the rapidly changing nature of speech (e.g., 1–10 ms) and the long-term average should reflect the overall across-syllable energy (e.g., 200–1000 ms). The scaling term must be constrained on the basis of several relevant considerations such as preventing over-attenuation of the speech or over-amplification of the noise floor. Finally, the processing could be applied to either the broadband stimulus or independently to band-pass filtered components of the speech-plus-noise stimulus that are then recombined to create the EEQ signal.

The following parameters were selected for the specific wide-band, non-real-time implementation of EEQ processing used in the current study.

- The AVG operation for computing E_{short} of Eq. (1) consisted of dividing the input into non-overlapping blocks of 5.3 ms and computing the mean energy within each block, which yielded a time constant of $\lambda_{\text{short}} = 5.3$ ms.
- Instead of a moving-average, the non-real-time AVG operation for computing E_{long} of Eq. (2) simply computed the mean energy over the entire speech-plus-noise stimulus (either disyllables or sentences, depending on the experiment), which yielded a time constant λ_{long} equal to the duration of the signal.
- The scale factor, $SC(t)$, was calculated according to Eq. (3), and then limited to be within the range of 0 to 20 dB. The lower limit of 0 dB was selected to prevent attenuation of

stronger signal components (such as the speech signal at higher SNRs) and the upper limit of 20 dB was selected to prevent over-amplification of the noise floor.

- The final operations employed in the processing [Eqs. (5) and (6)] were to normalize the energy of the processed stimulus to make it equal to that of the original unprocessed stimulus.

C. Background noises

The background noises selected to evaluate the EEQ processing technique included speech-shaped continuous noise as well as two types of temporally fluctuating noises (square-wave interruption and sinusoidal-amplitude modulation) at a rate of 10 Hz. These two different types of modulation in the temporally fluctuating noises were selected to contrast the discrete periods of masker offset introduced by the square-wave interruptions with the continuous changes in masker level introduced by the sinusoidal modulation. The fluctuation rate of 10 Hz was selected based on previous studies indicating that masking release is maximal for interruption rates in the vicinity of 8–16 Hz (e.g., Füllgrabe *et al.*, 2006).

Four types of speech-shaped noises (spectrally shaped to match the average of the spectra of the speech materials used in each experiment) were added to the stimuli.

Baseline Noise Condition (BAS): a continuous noise at 30 dB SPL was added to all speech stimuli in order to mask recording noise and provide a common noise floor for the stimuli.

Continuous Noise Condition (CON): an additional continuous noise was added to the *BAS* condition. The level of the *CON* noise is described for the specific experiments in Secs. III and IV below.

Square-Wave Interrupted Noise Condition (SQW): an additional *SQW* noise at a rate of 10 Hz and a duty cycle of 50% was added to the *BAS* condition and its overall root mean square (rms) level was adjusted to be equal to that of the *CON* noise.

Sinusoidal Amplitude Modulation Noise Condition (SAM): an additional *SAM* noise with a rate of 10 Hz was added to the *BAS* condition and its overall rms was adjusted to be equal to that of the *CON* noise.

D. Speech-plus-noise stimuli

Speech-plus-noise stimuli for the four noise conditions described above were created for an unprocessed (UNP) condition and for an EEQ condition. For each stimulus, a segment of speech-shaped noise of the same duration was selected randomly from a 30-s segment of digitized speech-shaped noise and added to the speech to create the initial speech-plus-noise stimulus. For the UNP condition, the speech-plus-noise stimuli materials were presented with no further processing other than the application of an individual NAL-RP frequency-gain characteristic (Dillon, 2001) for each HI listener. For the EEQ condition, noise was added to the stimuli as described above in Sec. III D prior to processing using the specific EEQ implementation described above in Sec. III B and followed by NAP-RP amplification (for HI listeners only).

E. Effects of EEQ processing on speech-plus-noise stimuli

The effects of the processing are demonstrated in Fig. 2 through waveforms (top) and level distributions (bottom) for the vowel-consonant-vowel (VCV) utterance /ɑ/-/p/-/ɑ/ produced by a male talker. Speech at a level of 70 dB SPL is shown in four different backgrounds of noise for UNP signals (upper left of Fig. 2) compared to the signals after EEQ processing (upper right). Plots are shown for speech combined with each of the four types of noise described above.

For purposes of illustration, the same frozen noise sample was used in each of the waveform plots to make it easier to observe the effects of the processing. In the *CON* noise with SNR of +40 dB (the *BAS* condition), the low-level energy associated with the plosive burst following the silent interval after the initial /ɑ/ production and leading into the final /ɑ/ production is seen to be higher in the EEQ compared to the UNP plots. When the *CON* noise level is increased to yield an SNR of −10 dB, the signal is dominated by the random fluctuations in the noise, and the waveforms are similar whether or not the EEQ processing is employed. In the two modulated noises (*SQW* and *SAM* both at a rate of 10 Hz and added to the signal to yield an SNR of −10 dB), the level of the noise changes periodically as a function of time. For example, the *SQW* modulation alternates between intervals of 50 ms off and 50 ms on, resulting in the noise dominating the signal during its on state and speech energy dominating during its off state. By comparing the plots of the UNP and EEQ-processed signals, it can be seen that the effect of the processing is to increase the level of the signal during the dips in the noise.

The effects of EEQ processing on signal level are demonstrated further in the level distribution histograms shown in the bottom of Fig. 2. The histograms were derived from sampling the waveform at a rate of 32 kHz, converting the individual sample magnitudes to dB, and generating a histogram of unit bin size with normalized probabilities. The dotted vertical bar indicates the rms level of each of the signals, and the solid vertical bar indicates the median of the level distribution. The histograms for UNP signals may be compared to those for EEQ signals in terms of their medians and standard deviations. With the exception of the *CON* case, EEQ signals demonstrate higher medians and smaller standard deviations than the corresponding UNP signals. For UNP compared to EEQ signals, the medians of the level distribution increased from 61 to 65 dB for the *BAS* case, from 72 to 76 dB for *SQW*, and from 74 to 77 dB for *SAM*, while the corresponding standard deviations decreased from 19 to 13 dB for *BAS*, 19 to 14 dB for *SQW*, and 13 to 10 dB for *SAM*. The characteristics of the level distributions were similar for both UNP and EEQ processing in the *CON* background, both with medians of 77 dB and standard deviations of 9 dB. Both the increase in medians and the decrease in standard deviations for EEQ compared to UNP demonstrate the decrease in amplitude variation as a result of the

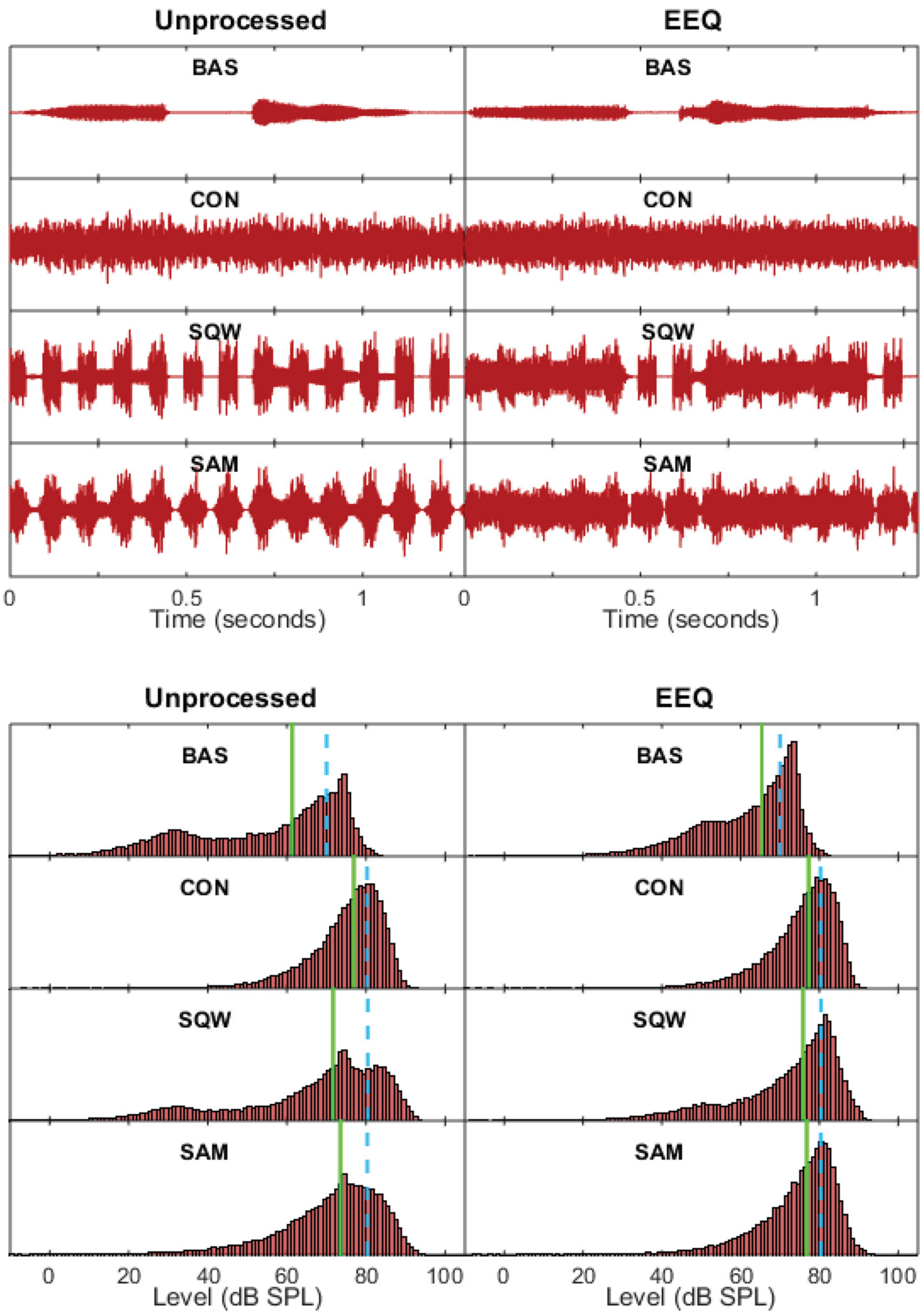

FIG. 2. (Color online) Stimulus waveforms (upper half of plot) and level distributions (lower half) for unprocessed (UNP) signals, shown on left, and EEQ signals, shown on right, depicting the utterance /ɑ/-/p/-/ɑ/ (produced by a male talker) at a level of 70 dB SPL in four different noise backgrounds. In each half of the plot, the rows represent speech in (a) the baseline (*BAS*) condition with a continuous-noise background of 30 dB SPL, i.e., speech-to-noise ratio (SNR) of +40 dB, (b) continuous (*CON*) noise with SNR of −10 dB, (c) square-wave interrupted (*SQW*) noise with SNR of −10 dB, and (d) sinusoidally amplitude-modulated (*SAM*) noise with SNR of −10 dB. The level distribution histograms were derived from sampling the waveform at a rate of 32 kHz, converting the samples to dB, and generating a histogram of unit bin size with normalized probabilities. The dashed vertical bar indicates the rms level of each of the signals; the solid vertical bar indicates the median of the level distribution.

processing designed to amplify the low-energy portions of the signal.

F. Experimental control

The evaluations of the EEQ processing scheme included tests of consonant identification as well as tests of sentence reception, described below in Secs. III and IV, respectively. All experiments were controlled by a desktop PC equipped with a 24-bit PCI sound card (E-MU 0404 by Creative Professional) and using MATLAB™ software. The digitized speech-plus-noise stimuli were played through a D/A converter and passed through a programmable attenuator (Tucker-Davis PA4) and a headphone buffer (Tucker-Davis HB6) before being presented monaurally to the listener in a soundproof booth over headphones (Sennheiser HD580). A monitor, keyboard, and mouse located within the soundproof booth allowed interaction with the control PC.

III. CONSONANT-IDENTIFICATION EXPERIMENTS

A. Speech materials and procedure

The speech materials were consonant stimuli in vowel-consonant-vowel (VCV) disyllables taken from the corpus of Shannon *et al.* (1999). These included recordings by 4 M and 4 F talkers of /ɑ/-C-/ɑ/ syllables with C=/p t k b d g f s ʃ v z dʒ m n r l/. The training set consisted of 64 syllables (one utterance of each of the 16 syllables by 2 M and 2 F talkers) and the test set consisted of a separate set of 64 syllables (one utterance of each of the 16 syllables by two different M and two different F talkers). The speech-shaped noise that was added to the VCV stimuli was derived from the average spectra of the 128 VCV tokens and included the four types of noise described above in Sec. II C (*BAS, CON, SQW*, and *SAM*). The recordings were digitized with 16-bit precision at a sampling rate of 32 kHz and filtered to a bandwidth of 80–8020 Hz for presentation. For the HI listeners, linear-gain amplification was applied to the speech-plus-noise stimuli using the NAL-RP formula (Dillon, 2001).

Consonant identification was tested using a one-interval, 16-alternative, forced-choice procedure without correct-answer feedback. On each 64-trial run, one of the 64 tokens (from either the training set or from the test set, whose particular use is described further below) was selected randomly without replacement. The listener's task was to identify the medial consonant of the syllable that had been presented by selecting a response (using a computer mouse) from a 4 × 4 visual array of orthographic representations associated with the consonant stimuli. No time limit was imposed on the listeners' responses. Each run lasted roughly 4–8 min depending on the listener's response times. Chance performance was 6.25%-correct.

1. Experiment 1

NH listeners were tested using a speech level of 60 dB SPL with an SNR of -10 dB for the *CON, SQW*, and *SAM* conditions. For each HI listener, a comfortable speech level was selected for listening to UNP speech in the *BAS* condition and an SNR was selected to yield roughly 50%-correct scores for UNP speech in a *CON* noise background. These speech levels and SNRs are provided in the panels for each HI listener in Fig. 1. Stimuli were presented either UNP or EEQ-processed. The UNP conditions were tested prior to the EEQ conditions.

The four noise conditions were tested in order of *BAS* first, then *CON* or *SQW* in a randomly selected order, and *SAM* noise last. Eight 64-trial runs were presented at each condition. The first three runs used the 64 tokens from the training set and the final five runs used the 64 tokens from the test set. The three training runs and the first test run were considered as practice and discarded. The final four test runs were used to calculate the percent-correct score for each of the eight conditions (two speech types by four noises).

2. Experiment 2

Experiment 2 was conducted to examine the effects of EEQ processing as a function of SNR and to compare these effects to those obtained on UNP materials. Four of the HI listeners (HI-1, HI-3, HI-8, and HI-9) were tested at three values of SNR: the value used in experiment 1 to yield 50%-correct identification for UNP speech in *CON* noise and two additional values (one lower by 4 or 5 dB and one higher by 4 or 5 dB). This testing was conducted with UNP and EEQ speech in three types of noise: *CON, SQW*, and *SAM*. The test order for UNP and EEQ processing was selected randomly for each listener. For each processing type, the test order of the three types of noise was selected at random and within each noise, the three values of SNR were presented in random order. Five 64-trial runs were presented at each condition using the tokens from the test set. The first run was discarded as practice and the final four runs were used to calculate the percent-correct score on each of the 18 conditions (2 processing types by 3 noises by 3 SNRs).

For both experiments 1 and 2, a normalized measure of masking release (NMR) was employed. In this metric, the $\mathrm{MR}_{\mathrm{PCT}}$ (i.e., the difference in scores between either *SQW* or *SAM* noise conditions and the *CON* noise condition) is represented as a fraction of the total possible amount of improvement defined as the difference in scores between *BAS* and *CON*. Thus, the measure reflects the fraction of the performance "lost" when going from baseline to continuous noise that was subsequently "restored" by going from continuous to fluctuating noise. This normalization process takes into account the SNR at which a given listener was tested (as reflected in the *CON* score) and thus allows for comparisons among listeners tested at different values of SNR. NMR was calculated from the percent-correct scores as

$$\mathrm{NMR} = \frac{SQW \text{ or } SAM \text{ Score} - CON \text{ Score}}{BAS \text{ Score} - CON \text{ Score}}. \qquad (7)$$

B. Results

1. Experiment 1

The results of experiment 1 are reported in Table I and summarized in Figs. 3 and 4. Table I provides the %-Correct scores for each HI listener in each of the four noise conditions for UNP and EEQ processing. In Fig. 3, consonant identification scores for EEQ signals are plotted as a function of those obtained for UNP signals. Figure 3(A) shows results obtained in the two continuous-noise conditions (*BAS* and *CON*) and Fig. 3(B) in the two modulated-noise conditions (*SQW* and *SAM*) for each of the seven HI listeners and for the mean across NH listeners. The diagonal line in each plot indicates equal performance on both types of processing.

For the NH listeners, performance was similar for both types of processing in all four noise backgrounds (filled data points in Fig. 3). The UNP and EEQ scores in each noise were *BAS*: 98.4%- and 98.6%-correct; *CON*: 50.5%- and 50.9%-correct; *SQW*: 92.5%- and 93.6%-correct; and *SAM*: 85.5%- and 86.5%-correct. Variability was low across NH listeners as indicated by a mean standard deviation of 1.9

TABLE I. Consonant identification scores in %-Correct for UNP and EEQ processing for four noise conditions: baseline *(BAS)*, continuous *(CON)*, square-wave interrupted *(SQW)*, and sinusoidally amplitude-modulated *(SAM)*. The speech levels and SNRs used in testing each of the seven HI listeners are provided in Fig. 1.

	UNP				EEQ			
	BAS	*CONT*	*SQW*	*SAM*	*BAS*	*CONT*	*SQW*	*SAM*
HI-1	98.4	54.7	85.9	82.8	99.2	53.1	96.5	86.3
HI-2	100.0	54.3	80.9	59.8	93.8	47.3	67.6	59.8
HI-3	93.7	56.6	62.9	60.9	87.5	60.6	78.5	71.5
HI-5	93.7	56.2	60.9	74.2	95.7	57.4	77.7	77.3
HI-10	93.7	50.0	53.1	51.6	87.1	47.7	71.1	63.3
HI-8	84.4	49.6	50.8	47.7	87.9	51.2	70.3	66.4
HI-9	73.4	48.1	39.4	51.8	88.3	45.3	67.6	57.4

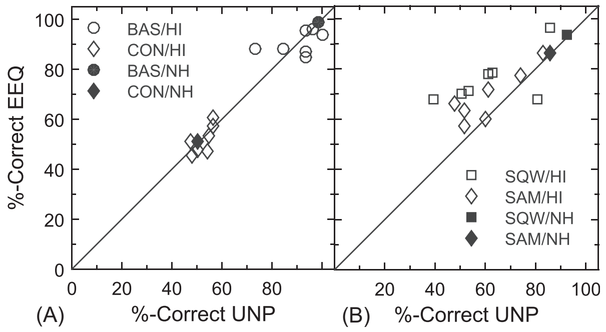

FIG. 3. Consonant identification scores in %-Correct: EEQ scores plotted as a function of UNP scores for each HI listener (unfilled symbols) and for the mean across four NH listeners (filled symbols). Scores for the two continuous noise backgrounds (*BAS* and *CON*) are shown in the panel on the left and scores for the two fluctuating noise backgrounds (*SQW* and *SAM*) are on the right. The diagonal line in each panel indicates equivalent performance for the two types of processing. (Abbreviations as defined in Fig. 2.)

percentage points averaged over the eight listening conditions.

For the HI listeners, performance was similar for UNP and EEQ processing in the *BAS* and *CON* noises but was generally higher for EEQ than UNP in *SQW* and *SAM* (see Fig. 3). Averaged over the HI listeners, the UNP and EEQ scores in each noise were *BAS*: 91.0%- and 91.4%-correct; *CON*: 52.8%- and 51.8%-correct; *SQW*: 62.0%- and 75.6%-correct; and *SAM*: 61.3%- and 68.9%-correct, respectively. An analysis of variance (ANOVA) was conducted on the percent-correct scores of the HI listeners after transformation to rationalized arcsine units (RAU) (Studebaker, 1985) to test for main effects of speech-processing type and noise condition and their interaction (using a two-factor within-subjects design with subjects as a random variable). The main effect of speech type was not significant [F(1, 6) = 2.75, p = 0.148], but significant effects were observed for noise condition [F(3, 18) = 57.5, p < 0.0001], and for the interaction of speech by noise [F(3, 18) = 7.00, p = 0.003]. *Post hoc* Tukey comparisons of the noise effect indicated that *BAS* scores were significantly higher than those of the other three noises, *CON* scores were significantly lower than those of the other three noises, and that *SQW* and *SAM* were intermediate between *BAS* and *CON* and significantly different from these two noise types but not from each other. The *post hoc* comparisons of the speech-by-noise interaction indicated that the EEQ scores were significantly higher than UNP scores for the *SQW* noise condition but not significantly different from each other for the other three noise types.

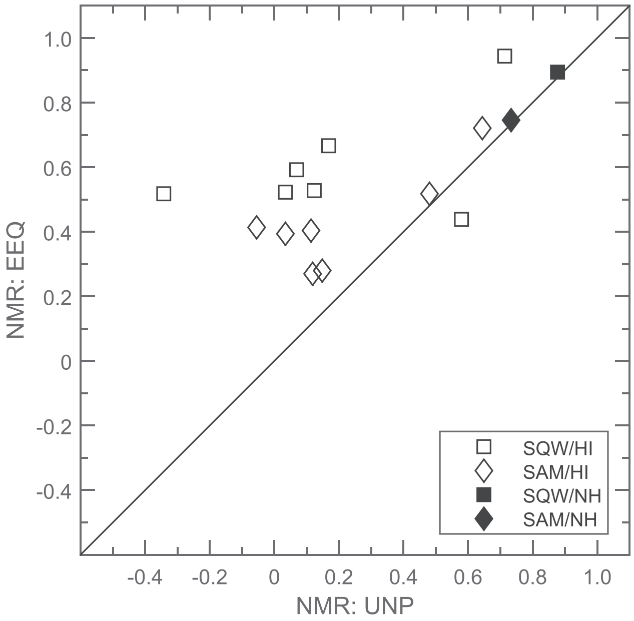

FIG. 4. Normalized masking release (NMR) for EEQ plotted as a function of UNP for each HI listener (unfilled symbols) and for the mean across four NH listeners (filled symbols). Results with *SQW* noise are shown by squares and *SAM* noise by diamonds. (Abbreviations as defined in Fig. 2.)

In Fig. 4, NMR for EEQ is plotted as a function of NMR for UNP for both types of modulated noises (*SQW* and *SAM*). Individual results are provided for the seven HI listeners and mean results for the NH listeners. For the NH listeners, NMR was essentially the same for UNP and EEQ processing for both *SQW* (0.88 for UNP vs 0.90 for EEQ) and *SAM* (0.73 for UNP vs 0.75 for EEQ). For the HI listeners, NMR computed with either type of modulated noise was higher for EEQ (mean NMR of 0.60 for *SQW* and 0.43 for *SAM*) than UNP (mean NMR of 0.19 for *SQW* and 0.21 for *SAM*). The results of a two-way ANOVA (using a two-factor within-subjects design with subjects as a random variable) of the NMR values of the HI listeners indicated a significant effect of speech-processing type [F(1, 6) = 19.01, p = 0.0048] but not of noise [F(1, 6) = 1.41, p = 0.280] or for the interaction between the two main variables [F(1, 6) = 2.61, p = 0.157]. Thus, NMR was significantly greater for EEQ than UNP for both SQW and SAM noise.

2. Experiment 2

The psychometric functions obtained on four HI listeners for consonant identification in *CON*, *SQW*, and *SAM* noises for EEQ and UNP are shown in the upper panels of Fig. 5. Each panel shows results from an individual listener where consonant-identification scores in %-Correct are plotted as a function of SNR for each of the three noise backgrounds. Several trends are evident in the data from each of the four listeners. The *CON* functions are generally overlapping for UNP and EEQ, and the EEQ functions for the *SQW* and *SAM* noises lie above those for UNP. Furthermore, for HI-3, HI-8, and HI-9, there is considerable overlap among the UNP functions for all three types of noise. Using the data shown in these figures, $\mathrm{MR}_{\mathrm{PCT}}$ was calculated as a function of SNR for *SQW* and *SAM* and is plotted in the lower panels of Fig. 5. The effect of SNR seen here is similar to that observed in previous studies (e.g., Bernstein and Grant, 2009; Oxenham and Simonson, 2009; Desloge *et al.*, 2010) indicating a tendency for an increase in masking release as SNR decreases and for negligible masking release at SNR > 0 dB. For HI-3, HI-8, and HI-9, the plots indicate greater $\mathrm{MR}_{\mathrm{PCT}}$ for EEQ compared to UNP for both *SQW* and *SAM* noise across a wide range of SNR (roughly −7 to +10 dB). For HI-1, $\mathrm{MR}_{\mathrm{PCT}}$ was greatest for EEQ processing in the *SQW* noise, similar for EEQ/*SAM* and UNP/*SQW*, and lowest for UNP/*SAM*.

The NMR was also calculated from the results at each SNR for *SQW* and *SAM* noise for UNP and EEQ processing

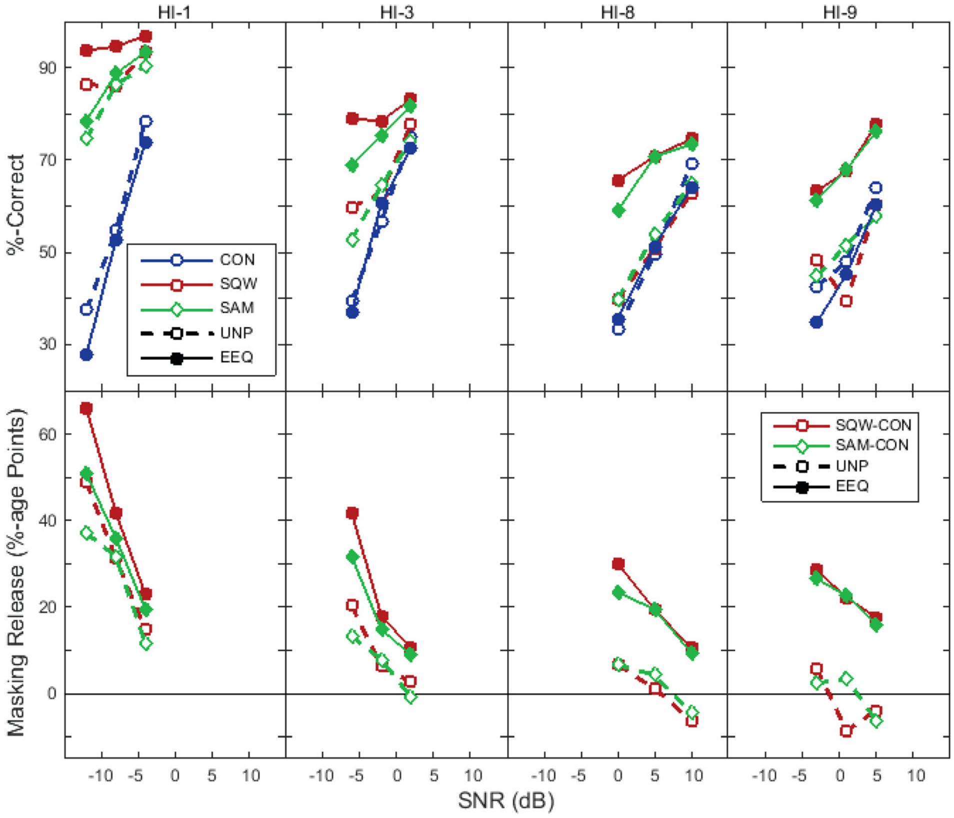

FIG. 5. (Color online) Upper panels: %-Correct scores plotted as a function of SNR in dB for UNP (unfilled symbols) and EEQ (filled symbols) processing in three different noises*:* *CON* (circles), *SQW* (squares), and *SAM* (diamonds). The level of the speech prior to NAL amplification is provided in the panels of Fig. 1 corresponding to each of the four HI listeners. Lower panels: Masking release (in percentage points, MR_{PCT}) as a function of SNR for each of the four HI listeners. MR_{PCT} is shown for four conditions: *SQW* and *SAM* noise for UNP and EEQ processing. (Abbreviations as defined in Fig. 2.)

for each of the four HI listeners and is plotted in Fig. 6. (Note that three data points for *SQW* noise are clustered in the vicinity of UNP = 0.8 and EEQ = 0.9, all arising from data for HI-1 at the three values of SNR.) These data indicate a strong tendency for larger NMR for EEQ for both types of modulated noise for all subjects regardless of SNR and, as expected, for higher NMR for *SQW* compared to *SAM* noise. For *SQW* noise, NMR across subjects for low, medium, and high SNR averaged 0.72, 0.71, and 0.67, respectively, for EEQ compared to 0.37, 0.26, and 0.05 dB, respectively, for UNP. Likewise, for *SAM* noise these mean values were 0.57, 0.58, and 0.59, respectively, for EEQ compared to 0.26, 0.15, and −0.09 for UNP. An ANOVA was conducted on the NMR values using a three-factor within-subjects design with subjects as a random variable. Significant effects were observed for each of the three main factors of speech type [F(1, 3) = 15.44, p = 0.0294], modulating noise [F(1, 3) = 15.08, p = 0.0303], and SNR [F(2, 6) = 7.57, p = 0.0229]. Of the interaction effects among these three variables, only the effect of speech type by SNR was significant [F(2, 23) = 14.28, p = 0.0001]. *Post hoc* Tukey-Kramer comparisons indicated a significant decrease in NMR with an increase in SNR for UNP (ranging from 0 at high SNR to 0.32 at low SNR) while no significant differences were observed in NMR across the three values of SNR for EEQ (where NMR was 0.64, 0.64, and 0.63 for low, mid, and high SNR, respectively). Neither of the other two interactions reached significance: speech by modulating noise type [F(1, 23) = 0.0039, p = 0.9544] or modulating noise type by SNR [F(1, 23) = 0.00026, p = 0.9664].

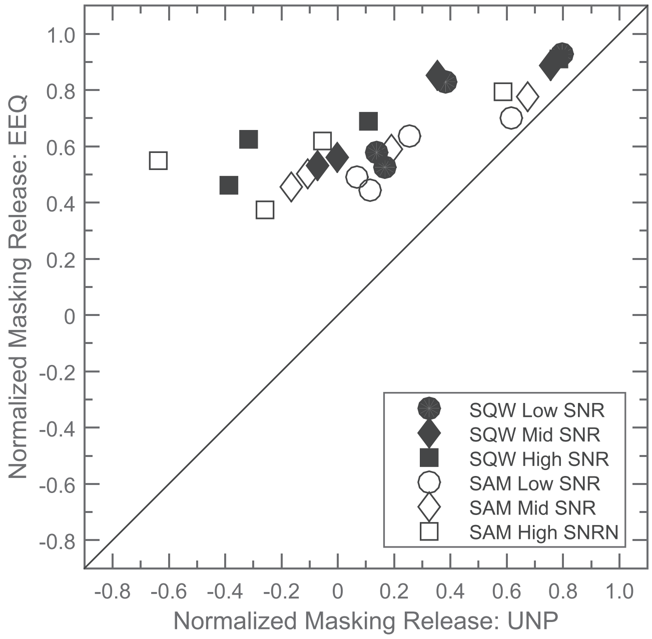

FIG. 6. Normalized masking release (NMR) for EEQ plotted as a function of UNP for SQW (filled symbols) and *SAM* noise (unfilled symbols) based on the results shown in Fig. 5 for 4 HI listeners. NMR is shown for three values of SNR: circles represent the lowest SNR, diamonds the middle SNR, and squares the highest SNR tested for each of the listeners. (Abbreviations as defined in Fig. 2.)

Thus, better performance was observed with EEQ processing compared to UNP in both modulated background noises across a wide range of SNR values. The higher values of both MR_{PCT} and NMR for EEQ compared to UNP arise from higher levels of performance on the modulated noises, while performance on the *CON* noise is similar for both EEQ and UNP (see psychometric functions in upper panels of Fig. 5). The plots of MR_{PCT} (lower panels of Fig. 5) show that after reaching a peak value in the negative SNR range, MR_{PCT} decreases with increasing SNR for both UNP and EEQ processing. The normalized NMR metric (see Fig. 6), however, remained fairly constant across SNR for EEQ processing in each of the two types of fluctuating background interference but decreased with an increase in SNR for UNP conditions.

IV. SENTENCE-RECEPTION TESTING

A. Materials and test procedure

The sentence-reception testing employed the materials of the hearing in noise test (HINT) (Nilsson *et al.*, 1994), which consists of 26 phonetically balanced lists of 10 sentences each, recorded by a single male talker. The sentences, which are conversational in nature and contain six or seven syllables each, were presented in four different backgrounds of speech-shaped HINT noise (matched to the long-term spectrum of these recorded sentences) as described in Sec. II B: *BAS*, *CON*, *SQW*, and *SAM*. The recordings and noises were digitized with 16-bit precision at a sampling rate of 24 kHz. Stimuli were presented either UNP or EEQ-processed. For the HI listeners, linear amplification was applied to the speech-plus-noise stimuli using the NAL-RP formula (Dillon, 2001).

The speech-to-noise ratio (SNR) necessary for 50%-Correct sentence identification was measured using an adaptive procedure in which the noise level was fixed and the sentence level was adapted using a one-up one-down rule. Following each sentence presentation, the listener was instructed to repeat the sentence word-for-word. The sentence was scored as being correct only if all words were identified correctly (with minor exceptions for articles and verb tense). The first sentence of each list was presented on consecutive trials with increasing level in 8-dB steps until it was correctly identified. The remaining nine sentences were presented once each, with the presentation level increased following an incorrect response and decreased following a correct response. For these sentences, the level was adapted using a one-up one-down rule with a 4-dB step size until the first reversal and a 2-dB step size thereafter. The SNR was computed as the average of the SNR (i.e., the speech presentation level minus the noise level) on the final six sentences. The experimenter was present in the soundproof booth with the participant to record the oral response to each sentence as either correct or incorrect.

The speech-shaped HINT noise was set to a level of 30 dB SPL for the *BAS* condition and to 70 dB SPL for the *CON, SQW,* and *SAM* conditions (with the exception of HI-9 who was tested at 80 dB SPL at these three conditions to yield at least 8 dB of masking in the *CON* condition compared to *BAS*). Three HINT lists were presented consecutively at each noise condition for each speech-processing type, and SNRs were averaged to yield a mean SNR for each speech-in-noise condition. The test order for UNP and EEQ was selected randomly for each listener. For each processing type, the *BAS* noise condition was always presented first and the other three noises (*CON, SQW*, and *SAM*) were presented in random order. A total of 24 HINT lists were employed in the testing of each listener. [It should be noted that six of the listeners (all but HI-10) had been tested with HINT materials roughly one year prior to the current study. Given the length of time between tests, it is unlikely that they would have benefitted from this previous exposure.]

TABLE II. SNR in dB for 50%-Correct reception of HINT sentences for UNP and EEQ processing for four noise conditions: baseline *(BAS)*, continuous *(CON)*, square-wave interrupted *(SQW)*, and sinusoidally amplitude-modulated *(SAM)*.

	UNP				EEQ			
	BAS	*CONT*	*SQW*	*SAM*	*BAS*	*CONT*	*SQW*	*SAM*
HI-1	5.9	−3.9	−16.6	−8.3	9.7	−2.8	−26.6	−9.4
HI-2	10.3	3.2	3.4	3.7	10.8	4.8	−10.6	2.1
HI-3	13.4	3.7	−2.1	−3.9	20.8	3.2	−8.6	−1.2
HI-5	12.8	2.3	−1.0	−3.9	17.0	1.9	−12.6	−1.0
HI-10	27.7	1.9	3.7	3.0	29.7	10.6	−2.8	1.7
HI-8	44.3	2.0	4.8	5.6	46.6	5.8	4.1	4.1
HI-9	28.8	3.4	6.3	3.7	31.2	6.8	−1.4	2.1

The data were summarized as the mean SNR for each of the four noise conditions for UNP and EEQ and as the masking release in dB (MR_{dB}) for the 70 dB SPL (or 80 dB SPL, for HI-9) conditions for UNP and EEQ (i.e., SNR in *CON* noise minus SNR in either *SQW* or *SAM* noise).

B. Results

The results of the HINT test are summarized in Table II for individual HI listeners. In Fig. 7, the SNR obtained on each of the four noise conditions for EEQ is plotted as a function of that obtained for UNP for individual HI listeners and as the mean across the NH listeners. For the *BAS* noise,

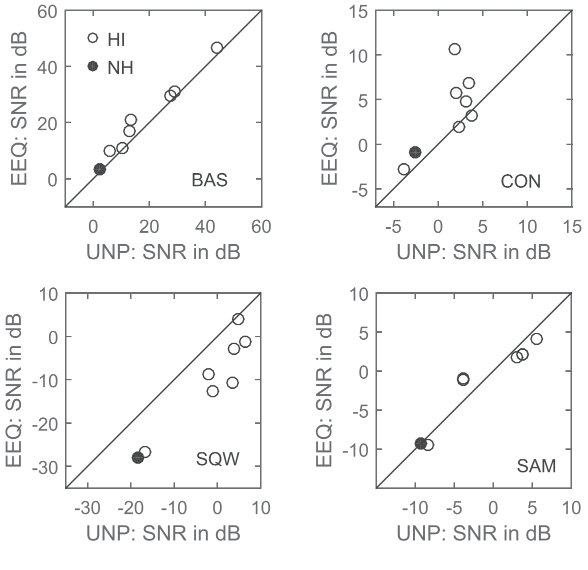

FIG. 7. Speech-to-noise ratio (SNR) for 50%-Correct HINT sentence reception: EEQ scores plotted as a function of UNP scores for each HI listener (unfilled symbols) and for the mean across four NH listeners (filled symbols). SNRs for the two continuous noise backgrounds (*BAS* and *CON*) are shown in the upper two panels and scores for the two fluctuating noise backgrounds (*SQW* and *SAM*) are shown in the lower two panels. The diagonal line in each panel indicates equivalent performance with the two speech-processing types. (Abbreviations as defined in Fig. 2.)

lower SNR for UNP (i.e., better performance) than for EEQ was observed for each HI listener with a mean SNR of 20.4 dB for UNP compared to 23.7 dB for EEQ. For the *CON* noise, mean SNR for HI listeners for UNP and EEQ, respectively, was 1.8 and 4.3 dB; for *SQW,* −0.2 and −8.4 dB, respectively; and for *SAM*, −0.22 and −0.19 dB, respectively. (In the subplot for *SAM* in Fig. 7, note the overlap of two data points at UNP = 3.7 and EEQ = 2.1 dB, and another two points at UNP = −3.9 and EEQ = −1.) The results of a two-way ANOVA (using a two-factor within-subjects design with subjects as a random variable) conducted on the SNR values of the HI listeners indicated no main effect of speech-processing type [F(1, 6) = 1.17, p = 0.3209] but showed significant effects for noise condition [F(3, 18) = 30.18.01, p < 0.0000] and for the interaction of the two main variables [F(3, 18) = 20.62, p < 0.0000]. *Post hoc* Tukey-Kramer multiple comparisons for the main effect of noise indicated that SNRs were significantly different among all four noises. *Post hoc* comparisons of the interaction effect arose from significantly lower SNR for EEQ than UNP for *SQW* noise but no significant differences between the two processing types for each of the other three noise conditions.

For the NH listeners, values of SNR for UNP and EEQ, respectively, were 2.4 and 3.3 dB for *BAS*, −2.6 and −0.9 dB for *CON,* −9.3 and −9.7 dB for *SAM,* and −18.2 and −28.1 dB for *SQW* noise. An ANOVA conducted on the SNR values of the four NH listeners indicated significant main effects of speech-processing type [with lower SNR for EEQ than UNP; F(1, 3) = 36.1, p = 0.0092] and noise condition [F(3, 9) = 253.04, p < 0.0000], as well as their interaction [F(3, 9) = 18.54, p = 0.0003]. *Post hoc* Tukey-Kramer comparisons of the noise effect indicated significant differences among all noise conditions and indicated that the speech by noise interaction arose from lower SNR for EEQ than for UNP for *SQW* noise but not for the other three noises.

In Fig. 8, MR_{dB} for EEQ processing is plotted as a function of that for UNP for both types of fluctuating noise (*SQW* and *SAM*). Mean HI values of MR_{dB} in *SQW* noise for EEQ and UNP were 12.7 versus 1.8 dB, respectively [a statistically significant difference with t(6) = 6.74, p < 0.001] and in *SAM* noise were 4.5 versus 1.8 dB, respectively [not reaching statistical significance with t(6) = 1.51, p = 0.18]. The same trend was observed in the NH data, with a larger MR_{dB} observed for EEQ compared to UNP in *SQW* noise (27.1 versus 15.7 dB) than in *SAM* noise (8.7 versus 6.7 dB).

V. DISCUSSION

A. Comparison of EEQ and UNP signals: Local changes to SNR

For low-to-moderate SNRs and fluctuating noise, EEQ tends to amplify the higher-SNR stimulus segments present in the dips when noise energy is low relative to the lower-SNR stimulus segments when the noise energy is high. By doing this, EEQ changes the effective SNR of the stimulus, and so it is possible that the observed increase in NMR might be explained simply by an increase in SNR. This hypothesis was examined in an analysis employing the VCV stimuli.

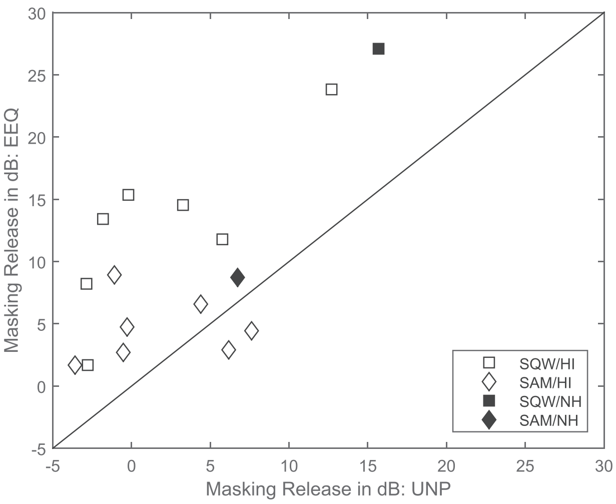

FIG. 8. Masking release in dB (MR_{dB}) for HINT sentences. MR_{dB} for EEQ processing is plotted as a function of MR_{dB} for UNP. The diagonal line indicates equivalent performance with the two speech-processing types. Data are plotted for individual HI listeners (unfilled symbols) and as means across NH listeners (filled symbols), and show MR_{dB} for *SQW* and *SAM* noise. (Abbreviations as defined in Fig. 2.)

In EEQ processing, the scale factor is applied linearly to the speech and to the noise. Thus, it is possible to determine its effect on the speech and noise components of the signal separately for a particular stimulus at a particular input SNR, which allows for computation of the post-processing SNR for that input. The output SNR for a particular input sample [consisting of specific speech and noise samples $s(t)$ and $n(t)$ and a known input SNR, SNR_{UNP}] may be calculated as follows.

(1) Compute the EEQ scale factor $SC[x(t)]$ based upon

$$x(t) = s(t) + n(t). \tag{8}$$

(2) The EEQ output signal is given as

$$y(t) = x(t) * SC[x(t)] = y_s(t) + y_n(t), \tag{9}$$

where

$$y_s(t) = s(t) * SC[x(t)] \tag{10}$$

and

$$y_n(t) = n(t) * SC[x(t)]. \tag{11}$$

(3) The post-processed SNR for this combination of $s(t)$, $n(t)$, and SNR_{UNP} is

$$SNR_{EEQ} = 10\log_{10}\left(\overline{y_s^2(t)} \;/\; \overline{y_n^2(t)}\right), \tag{12}$$

where $\overline{y_s^2(t)}$ and $\overline{y_n^2(t)}$ are the mean values of $y_s^2(t)$ and $y_n^2(t)$, respectively.

Each of the 64 speech tokens used in the experiments was examined with various noise types (*CON*, *SQW*, and *SAM*) and values of SNR_{UNP} (ranging from −40 to +40 dB). For every combination of speech token, noise type, and SNR_{UNP}, ten noise samples $n(t)$ of length equal to $s(t)$ were randomly generated. The above procedure was used to calculate SNR_{EEQ} as a function of SNR_{UNP} and noise type averaged across each of 10 noise samples combined with each of the 64 speech tokens. These averages were used to assess the impact of EEQ processing on SNR for each noise type.

In Fig. 9 (panel on left), the average SNR improvement (SNR_{EEQ} – SNR_{UNP}) calculated over the entire syllable is shown as a function of SNR_{UNP}. When SNR_{UNP} is negative, EEQ processing provides an increase in SNR for the *SQW* and *SAM* noises. This benefit decreases as SNR_{UNP} increases and actually crosses over to become a detriment for SNR_{UNP} greater than 2.9 and 0.0 dB for the two noise types, respectively. For *CON* noise, no SNR benefit is evident at low values of SNR_{UNP}, and an SNR detriment arises as SNR_{UNP} increases above −7.3 dB. This relationship predicts improved EEQ performance for fluctuating noises at lower SNRs and reduced EEQ performance for all noises at higher SNRs.

While this prediction agrees with the experimental results at low SNRs, the data do not show the predicted decrease in performance at positive values of SNR, as seen in the psychometric functions shown in Fig. 5. These functions are monotonically increasing, even for HI-3, HI-8, and HI-9, whose data include points obtained at positive values of SNR. Furthermore, results obtained in the *BAS* condition (where SNR was in the range of 35 to 45 dB across listeners) indicated no significant difference in performance between EEQ and UNP speech.

The analysis of the effects of EEQ processing on SNR was repeated with the focus only on the consonant region of the VCV recordings. The medial consonant of each of the 64 VCV disyllables was segmented with the aid of visual inspection of waveforms and spectrograms. This analysis employed the same syllable-length calculations as described above for the left panel of Fig. 9, but examined the average SNR improvement (SNR_{EEQ} – SNR_{UNP}) as a function of SNR only over the duration of the segmented consonant regions. These results are shown in the right panel of Fig. 9. Due to the generally weaker energy of the consonant, the SNR of this region tends to be lower than the SNR of the syllable as a whole, and the SNR benefit for EEQ processing persists up to higher values of SNR_{UNP} for all noise types: 0.6 dB for *CON*, 12.9 dB for *SAM*, and 11.3 dB for *SQW*. This analysis appears to be more consistent with the benefits of EEQ processing observed here for HI listeners than does the whole-syllable analysis.

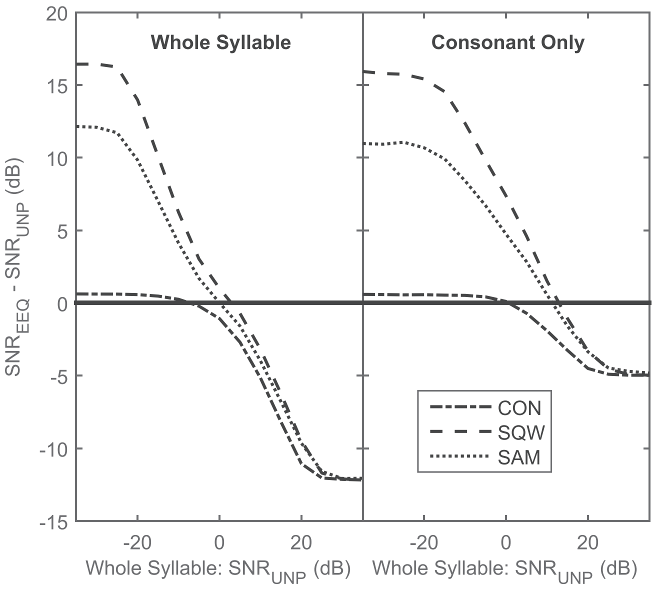

FIG. 9. The difference between the effective SNR after EEQ processing (SNR_{EEQ}) and the SNR before EEQ processing (SNR_{UNP}) is plotted as a function of SNR_{UNP}. Calculations were derived from 10 samples of noise in each of the 64 test syllables in *CON* (dash-dotted line), *SQW* (dashed line), and *SAM* noise (dotted line). A reference line of $SNR_{EEQ} - SNR_{UNP} = 0$ dB (solid horizontal line) is shown to highlight where the EEQ processing is lowering or raising the effective SNR. Calculations derived from the entire syllable are shown in the panel on the left and those restricted to the medial consonant segment are shown in the panel on the right. (Abbreviations as defined in Fig. 2.)

B. Comparison with other methods

The EEQ approach can be compared to other methods that address the problem of improved speech audibility in noise for HI listeners, including methods employed in the development of hearing aids, approaches in the area of noise reduction, and techniques developed within the context of automatic speech recognition (ASR). Several characteristics of EEQ processing are described below to highlight differences and similarities with other methods.

First, although EEQ processing is a form of compression, it operates on levels relative to the signal's own long-term energy and in this way can be contrasted with compression amplification. The goals of the two methods are similar in that they both attempt to match the range of speech levels into the reduced dynamic range of a listener with sensorineural hearing loss: specifically, lower-energy components receive greater amplification than higher-energy components (Lippmann *et al.*, 1981; Braida *et al.*, 1982; De Gennaro *et al.*, 1986). Compression amplification, however, is based on the actual SPL of the input as compared to the relative energy calculations performed in EEQ. Further, although compressive aids are often designed to use fast-attack and slow-release times resulting in compressive amplification that operates over multiple syllables, EEQ processing does not treat attack and release differently. It is designed to react rapidly based on the short-term energy estimate and can even operate within a single syllable to amplify less intense portions of the signal relative to more intense ones. Finally, in contrast to the improved performance in fluctuating noise shown here for EEQ processing, compression aids have not been shown to produce substantial benefits for HI listeners for speech in fluctuating noise compared to linear-gain aids. Houben (2006), for example, conducted a detailed study of a wide range of parameters associated with compression and did not observe any improvements for HI listeners in fluctuating versus continuous background noises.

Similar to EEQ processing and amplitude compression, peak clipping also reduces the dynamic range of speech, and has been in use for many years most frequently as a means of pre-processing speech to improve its intelligibility prior to being presented in background noise (Licklider and Pollack, 1948; Pollack and Pickett, 1958). While some speech intelligibility is maintained with peak clipping, this processing can introduce major distortions to the original signal and intelligibility and speech quality in quiet and in noise can be reduced as a result (Kates and Kozma-Spytek, 1994; Kates and Arehart, 2005; Arehart *et al.*, 2007). The advantage of EEQ processing over peak-clipping is that it does not introduce major distortions due to the processing itself. Evidence of the greater distortions present in peak-clipping compared to EEQ processing is seen by its lower performance compared to UNP speech in continuous noise backgrounds at a given SNR (see Reed *et al.*, 2016), whereas no significant differences were observed between *CON* scores for EEQ and UNP speech (see left panel of Fig. 3 above).

Another characteristic of EEQ processing is that it operates blindly on the speech-plus-noise stimulus without the use of segmentation. In this sense, it can be contrasted with techniques such as CV ratio enhancement which also attempts to improve speech intelligibility in HI listeners by reducing the dynamic range of speech. However, this is accomplished by amplifying lower-level energy (generally associated with consonant production) relative to the higher-energy portions of the signal (generally the vowels). Although seemingly similar to EEQ processing, there are important differences between the two methods. First, many implementations of CV ratio enhancement (e.g, Gordon-Salant, 1986; Kennedy *et al.*, 1998) require explicit segmentation of speech into consonant and vowel components, or attempt to approximate segmentation using approaches such as detection of voiced versus unvoiced segments (Skowronski and Harris, 2006; Saripella *et al.*, 2011) or cross-frequency-band energy comparisons (Preves *et al.*, 1991). The relative energies from these segments are used to explicitly adjust the CV ratio. EEQ processing, on the other hand, does not carry out such a segmentation operation. Additionally, unlike EEQ processing, CV ratio enhancement generally operates on clean speech prior to presenting the processed signal in noise and is not designed to operate on the speech-plus-noise signal as is used by EEQ processing.

In the more general area of noise reduction, signal-processing approaches have been addressed towards direct reduction of the noise relative to the speech. These include single-channel techniques such as "spectral subtraction" that subtracts the estimated noise spectrum from the speech-plus-noise spectrum (Lim and Oppenheim, 1979) and multi-channel techniques such as directional microphones or microphone-array processing that attempt to preserve sources arriving from a preferred "target" direction while attenuating sources arriving from non-target directions (Desloge *et al.*, 1997; Welker *et al.*, 1997). One technique that has been applied to the speech-plus-noise signal at a single microphone involves the estimation of the ideal binary mask (Hu and Wang, 2001; Brungart *et al.*, 2006; Wang *et al.*, 2009): spectro-temporal regions that are dominated by speech are retained and those dominated by noise are eliminated. Recently, Healy *et al.* (2013) implemented a machine-learning approach to estimate the binary mask and observed large improvements in sentence-reception thresholds in continuous and fluctuating noises for hearing-impaired listeners. While noise reduction attempts to change the speech-to-noise energy ratio at any given time and frequency by attenuating elements identified as "noise," EEQ processing amplifies lower-energy portions of the signal relative to higher-energy portions.

Various techniques to improve the estimation of the speech signal in a noisy background have also been developed within the context of automatic speech recognition (ASR). Relevant examples include level and/or frequency equalization techniques which attempt to transform the speech-plus-noise signal so that its features mimic a set of reference features calculated in the temporal, spectral, or cepstral domains. One line of work equalizes the noisy input signal to reflect the characteristics of the clean speech used to train the ASR system (Hilger and Ney, 2006; Joshi *et al.*, 2011), while other work has focused on undoing the characteristics of Lombard speech (Boril and Hansen, 2010). Other techniques involve more complex models of intelligibility with the explicit goal of enhancing some intelligibility metric and may operate on clean speech prior to the addition of noise (Chanda and Park, 2007). Compared to these techniques, the EEQ technique operates blindly without the need for reference signals and instead focuses on making the low-energy portions of the speech signal accessible so that the HI listener's own auditory system can use this information to assist in speech comprehension.

C. Effects of EEQ on consonant and sentence reception in noise

In experiment 1, the consonant-identification scores of the NH listeners were essentially unchanged for EEQ compared to UNP in all four background noise conditions, leading to similar values of NMR for both processing types. For the HI listeners, on the other hand, consonant-identification scores were significantly higher in the *SQW* modulated-noise background for the EEQ compared to UNP stimuli while not showing significant differences between UNP and EEQ in the remaining noise backgrounds (*BAS*, *CON*, *SAM*). However, significantly higher values of NMR were observed for the HI listeners in EEQ compared to UNP for both *SQW* and *SAM* modulated noise. The results of experiment 2 support the conclusions of experiment 1 across a wider range of SNR. In addition, because the order of UNP and EEQ was randomized in experiment 2, these data provide evidence that the results obtained in experiment 1 (where UNP conditions were tested before EEQ) were not due to an order effect. In experiment 2, both $\mathrm{MR}_{\mathrm{PCT}}$ and NMR were greater for EEQ than for UNP stimuli across a wide range of SNR. Although NMR decreased with an increase in SNR for UNP stimuli, it was independent of SNR for EEQ processing. As seen in Fig. 6, there was a large inter-subject variability in NMR for UNP stimuli, with NMR ranging from roughly −0.7 to +0.8 across the HI listeners. With EEQ processing, however, all HI listeners experienced release from masking and the range of NMR across HI listeners was reduced to roughly 0.4 to 0.9. Thus, the energy-equalization signal-processing technique employed here allows HI listeners to benefit from the momentary reductions in noise levels (and improved short-term SNRs) that occur in the modulated noises to a greater extent than is the case with UNP.

For reception of sentences, the NH listeners performed equally well for UNP and EEQ-processed materials in the *BAS*, *CON*, and *SAM* noises; however, EEQ processing led to a 10-dB improvement in SNR over UNP for the *SQW* noise (and thus to a larger $\mathrm{MR}_{\mathrm{dB}}$). The HI listeners showed improved performance for *SQW* noise but not for *SAM* noise with EEQ processing. The improved performance in *SQW* but not *SAM* noise with EEQ processing (for both groups of listeners) may be related to the combination of gap depth and duration as shown in Fig. 2. SQW modulation yields noise with regularly spaced, full-depth, 50-ms dips. SAM modulation, on the other hand, yields regularly spaced, variable-depth, 50-ms dips. Full depth is achieved only momentarily when the modulation reaches its nadir. As such, the duration of deep dips in the noise is shorter for SAM than for SQW, which may reduce the effectiveness of EEQ processing. In addition, the HI listeners required an SNR for 50%-correct sentence reception that was roughly 3 dB higher for EEQ than UNP in the *BAS* and *CON* noises (but reaching statistical significance only for *BAS*). The poorer performance of the HI listeners in these noises for EEQ processing compared to UNP may be related to the possible amplification of the background noise in instances where that noise exceeds the level of the speech signal.

These results lend support to the hypothesis that reduced audibility is an important factor in the reduced masking release for unprocessed speech in modulated noises for HI listeners. Other evidence in support of this hypothesis arises from studies comparing the performance of HI listeners to NH listeners under an audibility-based simulation of hearing loss (e.g., Zurek and Delhorne, 1987; Desloge *et al.*, 2010). Desloge *et al.* (2010), for example, observed that the reduced masking release values of HI listeners in a HINT sentence recognition task (Nilsson *et al.*, 1994) were generally well-matched by the results of NH listeners under a hearing-loss simulation that used a combination of threshold-elevating noise and multi-band expansion to reproduce a given pure-tone audiogram. Thus, in the simulation study, a reduction in audibility led to a reduction in masking release. In the current study, the application of EEQ processing led to greater audibility of the signal during dips in the modulated noises by the HI listeners, which in turn led to increased values of NMR, $\mathrm{MR}_{\mathrm{PCT}}$, and $\mathrm{MR}_{\mathrm{dB}}$.

Our attempt to improve the performance of HI listeners in interrupted noise was based on increasing the energy of the signal available during the dips in the interrupted noises. Results obtained with the EEQ processing technique indicated an improvement in consonant scores in interrupted noise compared to performance with unprocessed signals. In addition, performance in continuous noise backgrounds was similar for EEQ and unprocessed signals, indicating that improved NMR was based on better performance in interrupted noise (rather than lower performance on continuous noise). Similar trends were observed in the sentence-reception data, where significantly larger values of $\mathrm{MR}_{\mathrm{dB}}$ were observed for EEQ compared to UNP for SQW (but not SAM) interruption for both NH and HI listeners. These results suggest that HI listeners are able to "listen in the dips" of interrupted noise when the audibility of the signal present in the dips is sufficiently high. Thus, factors other than an insensitivity to modulation masking in HI listeners (as postulated by Stone *et al.*, 2012; Stone and Moore, 2014), appear to contribute to the reduced masking release observed with unprocessed signals. Our initial attempts to explain the improved performance of HI listeners for EEQ versus UNP in interrupted noise focused on the role of changes in local SNR brought about by EEQ processing. When examined for the consonant region of the VCV speech tokens, EEQ processing resulted in improved local SNR over a wide range of SNR for unprocessed speech, consistent with the general trends in the consonant-identification data reported here.

The current study demonstrates that EEQ processing leads to a greater release of masking for HI listeners in certain types of fluctuating noise (in particular, for square-wave interruptions), and that these benefits are observed for both segmental and connected-speech materials. However, these results must be tempered by various limitations of the study including the small number of HI listeners, the artificial nature of the interrupted noises, and the use of a non-real-time signal processing scheme. Further research is planned to address these issues. Finally, in addition to studying EEQ processing as a means of improving the performance of HI listeners in fluctuating noise, this technique (or a variant thereof) might also be adapted to the study of the mechanisms related to masking release in NH and HI listeners as well as those with cochlear implants.

VI. CONCLUSIONS

- For consonant identification, scores for hearing-impaired listeners were significantly higher for EEQ-processed than for unprocessed stimuli in square-wave modulated noise but not in backgrounds of continuous or sinusoidally amplitude-modulated noise. A normalized measure of masking release was greater for EEQ processing than for unprocessed stimuli for both types of modulated noise. Although normalized masking release decreased with an increase in SNR for unprocessed stimuli, it was independent of SNR for EEQ processing.
- The improved performance for EEQ processing compared to unprocessed stimuli in fluctuating noises may be explained in part by an increase in local SNR during the medial consonant segments of the vowel-consonant-vowel test stimuli after EEQ processing across a wide range of input SNR.
- For sentence reception, EEQ processing led to better performance of the hearing-impaired listeners in square-wave modulated noise, to similar levels of performance for EEQ-processed and unprocessed stimuli in continuous and sinusoidally modulated noises, and to worse performance in the baseline condition. Improvements in masking release in dB were observed in square-wave modulated noise for listeners with both normal and impaired hearing.
- Overall, these results suggest that, with EEQ processing, hearing-impaired listeners are able to listen more effectively in the dips of fluctuating noises when the energy of the signal present during the dips is sufficiently high.

ACKNOWLEDGMENTS

This research was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award No. R01DC000117. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Arehart, K. H., Kates, J. M., Anderson, M. C., and Harvey, L. O., Jr. (**2007**). "Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 1150–1164.

Bernstein, J. G. W., 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.

Boril, H., and Hansen, J. H. L. (**2010**). "Unsupervised equalization of Lombard Effect for speech recognition in noisy adverse environments," IEEE Trans. Audio Speech Lang. Process. **18**, 1379–1393.

Braida, L. D., Durlach, N. I., De Gennaro, S. V., Peterson, P. M., and Bustamante, D. K. (**1982**). "Review of recent research on multi-band amplitude compression for the hearing impaired," in *The Vanderbilt Hearing Aid Report: Monographs in Contemporary Audiology*, edited by G. A. Studebaker and F. H. Bess (York Press, Upper Darby), pp. 123–140.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Chanda, P. S., and Park, S. (**2007**). "Speech intelligibility enhancement using tunable equalization filter," in *International Conference on Acoustics, Speech, and Signal Processing—ICASSP 4*, pp. IV-613–IV-616.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

De Gennaro, S., Braida, L. D., and Durlach, N. I. (**1986**). "Multichannel syllabic compression for severely impaired listeners," J. Rehab. Res. Dev. **23**, 17–24.

Desloge, J. G., Rabinowitz, W. M., and Zurek, P. M. (**1997**). "Microphone-array hearing aids with binaural output. I. Fixed-processing systems," IEEE Trans. Speech Audio Process. **5**, 529–542.

Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., and Delhorne, L. A. (**2010**). "Speech reception by listeners with real and simulated hearing impairment: Effects of continuous and interrupted noise," J. Acoust. Soc. Am. **128**, 342–359.

Dillon, H. (**2001**). *Hearing Aids* (Thieme, New York), pp. 239–247.

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.

Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**, 74–84.

George, E. J., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

Gilbert, G., and Lorenzi, C. (**2006**). "The ability of listeners to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am. **119**, 2438–2444.

Gordon-Salant, S. (**1986**). "Recognition of time/intensity altered CVs by young and elderly subjects with normal hearing," J. Acoust. Soc. Am. **80**, 1599–1607.

Healy, E. W., Yoho, S. E., Wang, Y., and Wang, D. (**2013**). "An algorithm to improve speech recognition in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **134**, 3029–3038.

Hilger, F., and Ney, H. (**2006**). "Quantile based histogram equalization for noise robust large vocabulary speech recognition," IEEE Trans. Audio Speech Lang. Process. **14**, 845–854.

Houben, R. (**2006**). "The effect of amplitude compression on the perception of speech in noise by the hearing impaired," Doctoral Dissertation, Utrecht University, the Netherlands.

Hu, G., and Wang, D. (**2001**). "Speech segregation based on pitch tracking and amplitude modulation," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 79–82.

Joshi, V., Bilgi, R., Umesh, S., Garcia, L., and Benitez, L. (**2011**). "Sub-band level histogram equalization for robust speech recognition," Interspeech 2011, Florence, Italy, 28–31 August 2011, pp. 1672–1675.

Kates, J. M., and Arehart, K. H. (**2005**). "Coherence and the speech intelligibility index," J. Acoust. Soc. Am. **117**, 2224–2237.

Kates, J. M., and Kozma-Spytek, L. (**1994**). "Quality ratings for frequency-shaped peak-clipped speech," J. Acoust. Soc. Am. **95**, 3586–3594.

Kennedy, E., Levitt, H., Neuman, A. C., and Weiss, M. (**1998**). "Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners," J. Acoust. Soc. Am. **103**, 1098–1114.

Léger, A. C., Reed, C. M., Desloge, J. G., Swaminathan, J., and Braida, L. D. (**2015**). "Consonant identification in noise using Hilbert-transform temporal fine-structure speech and recovered-envelope speech for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **138**, 389–403.

Licklider, J. C. R., and Pollack, I. (**1948**). "Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech," J. Acoust. Soc. Am. **20**, 42–51.

Lim, J. S., and Oppenheim, A. V. (**1979**). "Enhancement and bandwidth compression of noisy speech," *Proceedings of IEEE 67*, pp. 1586–1604.

Lippmann, R. P., Braida, L. D., and Durlach, N. I. (**1981**). "Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss," J. Acoust. Soc. Am. **69**, 524–534.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. USA **103**, 18866–18869.

Moore, B. C. J., Peters, R. W., and Stone, M. A. (**1999**). "Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips," J. Acoust. Soc. Am. **105**, 400–411.

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 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.

Pollack, I., and Pickett, J. M. (**1958**). "Masking of speech by noise at high sound levels," J. Acoust. Soc. Am. **30**, 127–130.

Preves, D. A., Fortune, T. W., Woodruff, B., and Newton, J. (**1991**). "Strategies for enhancing the consonant to vowel intensity ratio with in the ear hearing aids," Ear Hear. **12**(6), 139S–153S.

Reed, C. M., Desloge, J. G., Braida, L. D., Perez, Z. D., and Léger, A. C. (**2016**). "Level variations in speech: Effect on masking release in hearing-impaired listeners," J. Acoust. Soc. Am. **140**, 102–113.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.

Saripella, R., Loizou, P. C., Thibodeau, L., and Alford, J. A. (**2011**). "The effects of selective consonant amplification on sentence recognition in noise by hearing-impaired listeners," J. Acoust. Soc. Am. **130**, 3028–3037.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.

Skowronski, M. D., and Harris, J. G. (**2006**). "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noise environments," Speech Commun. **48**, 549–558.

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2012**). "Notionally steady background noise acts primarily as a modulation masker of speech," J. Acoust. Soc. Am. **132**, 317–326.

Stone, M. A., and Moore, B. C. J. (**2014**). "On the near non-existence of 'pure' energetic masking release for speech," J. Acoust. Soc. Am. **135**, 1967–1977.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (**2009**). "Speech intelligibility in background noise with ideal binary time-frequency masking," J. Acoust. Soc. Am. **125**, 2336–2347.

Welker, D. P., Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (**1997**). "Microphone-array hearing aids with binaural output. II. A two-microphone adaptive system," IEEE Trans. Speech Audio Process. **5**, 543–551.

Zurek, P. M., and Delhorne, L. A. (**1987**). "Consonant reception in noise by listeners with mild and moderate hearing impairment," J. Acoust. Soc. Am. **82**, 1548–1559.