



MIT Open Access Articles

Effect of Energy Equalization on the Intelligibility of Speech in Fluctuating Background Interference for Listeners With Hearing Impairment

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

As Published	10.1177/2331216517710354
Publisher	SAGE Publications
Version	Final published version
Citable link	https://hdl.handle.net/1721.1/135709
Terms of Use	Creative Commons Attribution NonCommercial License 4.0
Detailed Terms	https://creativecommons.org/licenses/by-nc/4.0/

Effect of Energy Equalization on the Intelligibility of Speech in Fluctuating Background Interference for Listeners With Hearing Impairment

Trends in Hearing
2017, Vol. 21: 1–21
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216517710354
journals.sagepub.com/home/tia



Laura A. D'Aquila¹, Joseph G. Desloge¹, Charlotte M. Reed¹,
and Louis D. Braida¹

Abstract

The masking release (MR; i.e., better speech recognition in fluctuating compared with continuous noise backgrounds) that is evident for listeners with normal hearing (NH) is generally reduced or absent for listeners with sensorineural hearing impairment (HI). In this study, a real-time signal-processing technique was developed to improve MR in listeners with HI and offer insight into the mechanisms influencing the size of MR. This technique compares short-term and long-term estimates of energy, increases the level of short-term segments whose energy is below the average energy, and normalizes the overall energy of the processed signal to be equivalent to that of the original long-term estimate. This signal-processing algorithm was used to create two types of energy-equalized (EEQ) signals: EEQ1, which operated on the wideband speech plus noise signal, and EEQ4, which operated independently on each of four bands with equal logarithmic width. Consonant identification was tested in backgrounds of continuous and various types of fluctuating speech-shaped Gaussian noise including those with both regularly and irregularly spaced temporal fluctuations. Listeners with HI achieved similar scores for EEQ and the original (unprocessed) stimuli in continuous-noise backgrounds, while superior performance was obtained for the EEQ signals in fluctuating background noises that had regular temporal gaps but not for those with irregularly spaced fluctuations. Thus, in noise backgrounds with regularly spaced temporal fluctuations, the energy-normalized signals led to larger values of MR and higher intelligibility than obtained with unprocessed signals.

Keywords

sensorineural hearing loss, consonant reception in noise, masking release, energy equalization

Date received: 27 October 2016; revised: 20 March 2017; accepted: 21 April 2017

Introduction

Listeners with sensorineural hearing impairment (HI) who are able to understand speech in quiet environments require a higher speech-to-noise ratio (SNR) to achieve criterion performance when listening in background interference than do listeners with normal hearing (NH; Festen & Plomp, 1990). This is the case regardless of whether the noise is temporally fluctuating, such as interfering voices, or is steady, such as that caused by a fan or motor. For listeners with NH, better speech reception is observed in fluctuating-noise backgrounds than in continuous noise of the same long-term root-mean-square (RMS) level, and they are said to experience a release from masking. This may arise from the ability

to perceive glimpses of the target speech during dips in the fluctuating noise (Cooke, 2006), and it aids in the ability to converse normally in noisy social situations.

Studies conducted with listeners with HI have shown reduced (or even absent) release from masking compared with that obtained with listeners with NH (e.g., see Bernstein & Grant, 2009; Desloge, Reed, Braida, Perez, & Delhorne, 2010; Festen & Plomp, 1990). Desloge et al.

¹Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding author:

Charlotte M. Reed, Room 36-751, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
Email: cmreed@mit.edu



(2010), for example, measured the SNR required for 50%-correct reception of sentences in continuous and 10-Hz square-wave interrupted noise. For unprocessed speech presented in 80 dB SPL noise backgrounds, a mean difference of 13.9 dB between continuous and fluctuating noise was observed for listeners with NH while it was only 5.3 dB across a group of 10 listeners with HI. Desloge et al. (2010) explained this on the basis of the effects of reduced audibility or speech-weighted sensation level in listeners with HI, who are less likely to be able to receive the target speech in the noise gaps. Other factors may contribute to the reduced masking release (MR) of listeners with HI, including reductions in cochlear compression and frequency selectivity (Moore, Peters, & Stone, 1999; Oxenham & Kreft, 2014), a reduced ability to process the temporal fine structure of speech (Hopkins & Moore, 2009; Hopkins, Moore, & Stone, 2008; Lorenzi, Debruille, Garnier, Fleuriot, & Moore, 2009; Lorenzi, Gilbert, Carn, Garnier, & Moore 2006; Moore, 2014), and reduced sensitivity to the random amplitude fluctuations in noise (Oxenham & Kreft, 2014; Stone, Füllgrabe, & Moore, 2012). The current research, however, is concerned with further investigation of the role of audibility in the reception of speech in temporally fluctuating background interference.

Previous studies of Léger, Reed, Desloge, Swaminathan, and Braidà (2015) and Reed, Desloge, Braidà, Léger, and Perez (2016) suggest that MR for listeners with HI may be enhanced by processing methods which lead to a reduction of the normal level variations present in speech. In these two studies, however, the improvements in MR arose primarily from decreased performance in continuous background noise relative to that in interrupted noise, due to distortions introduced by the signal processing. To address these issues, Desloge, Reed, Braidà, Perez, and D'Aquila (2017) developed a signal-processing technique, energy-equalization (EEQ), designed to overcome these limitations without suffering a loss in intelligibility in continuous background noise. Using non-real-time processing of the broadband signal, the technique compares the short-term and long-term estimates of energy in the input and increases the level of short-term segments when their energy is below average. It then renormalizes the signal to maintain an output energy equal to that of the input. When combined with the linear amplification needed to overcome loss of sensitivity, lower level segments are amplified more than higher level segments. In this way, EEQ processing resembles traditional compression amplification systems. However, there are some important differences. The aim of EEQ is different from that of amplitude compression: EEQ is designed to elevate dips in short-term energy to match the long-term average energy while compression amplification attempts to match the range of speech levels into the

reduced dynamic range of a listener with sensorineural hearing loss. This allows EEQ to be implemented using very simple, minimally constrained processing without detailed knowledge of the hearing loss characteristics. Compression amplification, on the other hand, requires some knowledge of the level-dependent characteristics of the hearing loss. Additionally, the gain applied by EEQ is based on the relative short- and long-term signal energies as opposed to the absolute signal energy used by the majority of compression amplification techniques. The use of relative energies results in EEQ being a homogeneous form of processing (i.e., if an input x processed with EEQ yields an output y , then a scaled input Kx yields an identically scaled output Ky). This is in direct contrast to a compression amplifier operating on absolute energies, where the output is $U[x]$ and where $U[]$ is a nonlinear (and, more specifically, nonhomogeneous) function. The action of such a compressor is to reduce the variation in level over at least part of the range of input levels. This can only be done with a system that is nonlinear: specifically, lower level components receive greater boost than higher level components (Braidà, Durlach, De Gennaro, Peterson, & Bustamante, 1982; De Gennaro, Braidà, & Durlach, 1986; Lippmann, Braidà, & Durlach, 1981).

Desloge et al. (2017) evaluated MR for consonant identification using a non-real-time version of EEQ. Results were compared with an unprocessed condition. Listeners with HI achieved similar scores for processed and unprocessed stimuli in quiet and in continuous-noise backgrounds, while superior performance was obtained for the processed speech in fluctuating background noises. Thus, the energy-normalized signals led to greater MR than obtained with unprocessed signals. The current study extends the work of Desloge et al. (2017) through (a) implementation and evaluation of a real-time version of EEQ, (b) use of a multiband system (EEQ4) in addition to the wide-band system (EEQ1) studied previously, and (c) evaluation of EEQ using a broader range of types of background interference, including a more realistic set of temporally fluctuating noises derived from real speech signals.

A real-time signal-processing algorithm for EEQ was implemented as a wideband (EEQ1) and a multichannel (EEQ4) system. Consonant identification was measured for listeners with NH and HI in a variety of background noises with and without EEQ processing. The experiments focused on consonants due to their high informational load in running speech. Consonant segments were presented in a fixed vowel-consonant-vowel context to minimize the effects of different linguistic and memory abilities across listeners that contribute to the reception of connected speech (e.g., Boothroyd & Nitttrouer, 1988). The noises were selected to examine the effects of periodic versus non-periodic temporal fluctuations and to

include noises characteristic of real speech maskers without the introduction of informational masking (Phatak & Grant, 2014; Rosen, Souza, Ekelund, & Majeed, 2013). The results of the consonant tests were summarized using a normalized measure of masking release (NMR) which permitted comparisons across listeners with different hearing losses and baseline speech-reception abilities. The results of the experiments were considered in light of previous research on compression amplification, in terms of differences in performance between EEQ1 and EEQ4, and through a model of opportunities for glimpsing in the various types of noise.

Signal-Processing Algorithm

The short-term signal energy of speech varies at a syllabic rate from more intense (usually during vowels), less intense (usually during consonants), and silent (De Gennaro et al., 1986; Rosen, 1992). The long-term signal level remains relatively constant and reflects the overall effort at which a speaker is talking. These overall properties of speech persist when background noise is added to the signal. The less intense portions of the speech signal present the most difficulty to listeners with HI and lead to reduced speech comprehension. Energy equalization is a way of combatting this difficulty by amplifying those parts of the speech signal that occur during the dips in the background noise. This technique makes speech content present during the dips in background noise more audible and may improve speech understanding.

The EEQ processing operates blindly and without introducing excessive distortion. The following is a general description of the steps for real-time EEQ. This processing is based on the non-real-time EEQ processing described previously by Desloge et al. (2017) but has been modified for real-time application to a speech plus noise (S+N) signal $x(t)$. The steps of EEQ processing are described later.

- The processing begins by computing running short-term and long-term moving averages of the signal energy, $E_{short}(t)$ and $E_{long}(t)$:

$$E_{short}(t) = \text{AVG}_{short}[x^2(t)] \quad \text{and}$$

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

where AVG is a moving-average operator that utilizes specified short and long time constants to provide estimates of signal energy. In this implementation, the AVG operators were single-pole low-pass filters applied to the instantaneous signal energy, $x^2(t)$, with time constants of 5 ms and 200 ms for the short and long averages, respectively. To minimize onset effects, the single-pole

“memory” terms of these averages were preinitialized using the RMS level of the entire stimulus.

- A scale factor, $SC(t)$, is computed as:

$$SC(t) = \sqrt{E_{long}(t)/E_{short}(t)}$$

To prevent (a) attenuation of stronger signal components and (b) overamplification of the noise floor, $SC(t)$ was restricted to a range of 1 to 10 (corresponding to 0–20 dB). When $E_{short}(t)$ was equal to 0, $SC(t)$ was set to the maximum gain of 10.

- The scale factor is applied to the original signal:

$$y(t) = SC(t) x(t)$$

- The output $z(t)$ is formed by normalizing $y(t)$ to have the same energy as $x(t)$:

$$z(t) = K(t)y(t)$$

where $K(t)$ is chosen such that:

$$\text{AVG}_{long}[z^2(t)] = \text{AVG}_{long}[x^2(t)]$$

The processing described earlier can be applied either to a broadband signal or independently to bandpass filtered components. The current implementation operated on both the broadband signal (EEQ1) and a signal divided into four contiguous frequency bands (EEQ4). The multiband processing scheme was employed based on the hypothesis that separate processing within different spectral bands may be beneficial to listeners with frequency-dependent hearing losses. Both methods are described in further detail in Methods section.

Methods

The experimental protocol for testing human subjects was approved by the internal review board of the Massachusetts Institute of Technology (protocol 0403000005). 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.

Listeners

Six male (M) and three female (F) listeners with HI with bilateral, symmetric, mild-to-severe sensorineural hearing loss participated. They were all native speakers of American English and ranged in age from 20 to

69 years with an average age of 37 years. Six of the listeners were younger (33 years or less) and three were older (58–69 years). Five of the listeners had sloping high-frequency losses (HI-1, HI-2, HI-4, HI-5, and HI-7), three had relatively flat losses (HI-6, HI-8, and HI-9), and one had a “cookie-bite” loss (HI-3). Seven of the listeners (all but HI-1 and HI-3) were regular users of bilateral hearing aids. The five-frequency (0.25, 0.5, 1, 2, and 4 kHz) audiometric pure-tone average (PTA) ranged from 27 dB HL to 75 dB HL across listeners with an average of 45 dB HL.

The test ear, sex, age, and five-frequency PTA for each HI listener are listed in Table 1 along with the speech levels and SNRs employed in the experiment. The pure-tone thresholds of the listeners with HI (in dB SPL at the eardrum) are shown in Figure 1. The pure-tone threshold 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 for further details).

Four listeners with NH (defined as having pure-tone thresholds of 15 dB HL or better at octave frequencies between 250 and 8000 Hz) were also participated. They were native speakers of American English, included three M and one F and ranged in age from 19 to 54 years, with an average age of 30 years. A test ear was selected for each listener (two left ears and two right ears). The mean adaptive thresholds (measured as described earlier for the listeners with HI) across test ears of the listeners with NH are provided in the first panel of Figure 1.

Table 1. Test Ear, Sex, Age, and 5-Frequency PTA for Each Listener with HI.

Listener	Test ear	Sex	Age	5-Frequency PTA (dB HL)	Speech level (dB SPL)	SNR (dB)
HI-1	R	M	33	27	68	−8
HI-2	R	F	58	28	65	−2
HI-3	L	F	21	30	65	−6
HI-4	L	M	23	36	65	−2
HI-5	L	M	20	45	65	−4
HI-6	L	M	69	53	70	0
HI-7	L	M	59	56	68	0
HI-8	R	F	26	58	70	−2
HI-9	L	M	21	75	71	−2

Note. PTA = pure tone average; HI = hearing impairment; SNR = speech-to-noise ratio. The final two columns provide the comfortable speech presentation levels chosen by each listener prior to NAL amplification and the SNR used in testing all speech conditions. The SNR was chosen to yield 50% correct in continuous noise.

Speech Stimuli

The speech materials were vowel-consonant-vowel (VCV) stimuli, with a consonant set of /p t k b d g f s ʃ v z dʒ m n r l/ and the vowel /a/, taken from the corpus of Shannon, Jensvold, Padilla, Robert, and Wang (1999). The set used for testing consisted of 64 VCV tokens (1 utterance of each of the 16 disyllables by two M and two F speakers). The mean VCV duration was 945 ms with a range of 688 to 1339 ms across the 64 VCVs in the test set. The recordings were digitized with 16-bit precision at a sampling rate of 32 kHz and filtered between 80 and 8020 Hz.

Background Types

Seven different digitally generated noises from two broad categories of maskers were used. Four backgrounds, referred to as non-speech-derived noises, were generated from randomly generated speech-shaped Gaussian noise, while the remaining three backgrounds, referred to as speech-derived noises, were generated from actual speech samples.

The spectrum of non-speech-derived noises was speech-shaped and resulted from the average of the spectra across 128 VCV stimuli taken from the Shannon et al. (1999) corpus (in addition to the 64 test stimuli, another 64 stimuli from an additional two M and two F talkers were used in this average). These included the following: (a) a baseline noise consisting of continuous speech-shaped noise at 30 dB SPL (BAS), (b) BAS plus additional continuous noise (CON); (3) BAS plus square-wave interrupted noise (SQW) using 10-Hz interruption with 50% duty cycle and 100% modulation depth, and (4) BAS plus sinusoidally amplitude modulated (SAM) noise using 10-Hz modulation and 100% modulation depth.

The BAS was used in order to mask recording noise and to provide a common noise floor for the stimuli. Note that the BAS component of the SQW and SAM backgrounds meant that these modulated backgrounds never reached full modulation depth. A 10-Hz modulation rate was selected on the basis of previous studies indicating maximal MR for consonant identification in the vicinity 8 to 16 Hz (Füllgrabe, Berthommier, & Lorenzi, 2006).

The remaining three backgrounds were derived from actual speech samples. These maskers were designed to have temporal fluctuations characteristic of real speech while minimizing the effects of informational masking present with real speech maskers (see Phatak & Grant, 2014; Rosen et al., 2013) through the use of vocoding and time-reversed playback. These three vocoded (VOC) maskers were derived from speech recorded from either one (VOC-1), two (VOC-2), or four (VOC-4) talkers.

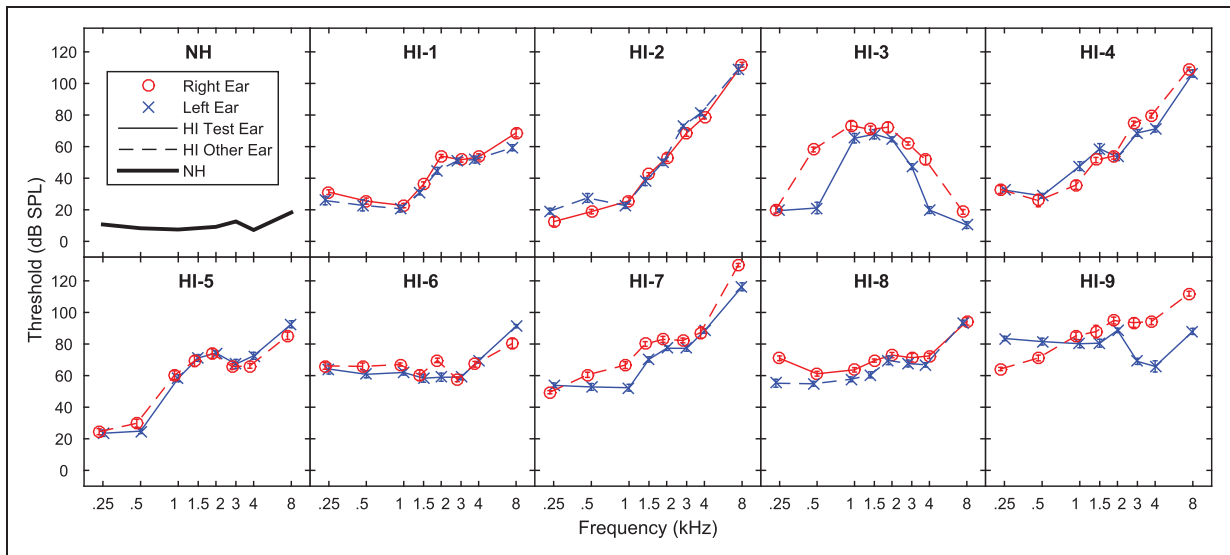


Figure 1. Pure-tone detection thresholds in dB SPL measured for 500-ms tones in a three-alternative forced-choice adaptive procedure. A thick black line representing the average thresholds of the test ears of the listeners with NH is shown in the upper left panel, and the thresholds for the listeners with HI are shown in the remaining panels. For the listeners with HI, thresholds are shown for the right ear (red circles) and left ear (blue x's), with the points of the test ear connected using a solid line and the points of the non-test ear connected using a dashed line.

The VOC-4 masker was chosen as a rough approximation to a continuous randomly generated noise (on the basis of results reported by Rosen et al., 2013 indicating only small changes in performance as the number of talkers in a vocoded noise masker increased above four).

The VOC tokens were derived through the application of a method described by Phatak and Grant (2014). Concatenated sentences from eight individual talkers (four M and four F) were equalized to the same RMS level. These sentences included recordings of IEEE sentences (IEEE, 1969), CUNY sentences (Boothroyd, Hanin, & Hnath, 1985), and the Rainbow Passage (Fairbanks, 1960). VOC-1 tokens were produced by filtering individual input speech samples and randomly generated noise of equal duration into six logarithmically equal bands in the range 80 to 8020 Hz via forward and time-reverse application of sixth order Butterworth filters (yielding 72 dB/octave aggregate roll-off). The speech-signal envelope in each of the six bands was derived via forward and time-reverse low-pass filtering of the rectified signal at 64 Hz with third order Butterworth filters (yielding 36 dB/octave aggregate roll-off). These envelopes were then used to modulate the corresponding bands of filtered noise, and the results were scaled to equalize the vocoded band energies with those of the original speech bands. Finally, the modulated and scaled bands were summed, and the output was time-reversed to yield a vocoded, time-reversed waveform corresponding to the original speech token from which it was derived. VOC-2 and VOC-4 tokens were generated by adding two or four VOC-1 tokens derived

from different speakers of the same gender. Finally, the tokens were scaled to the desired presentation level. For each trial, the gender from which the masker was derived was selected at random (i.e., the gender of the masker did not necessarily match that of the VCV stimulus presented on that trial).

Speech in Noise Processing

Three conditions were used: Unprocessed (UNP), EEQ1, and EEQ4. In the 1-band EEQ condition (EEQ1), EEQ processing was applied to the broadband S+N signal over the range of 80 to 8020 Hz. In the 4-band EEQ condition (EEQ4), the input was separated into four logarithmically equal bands in the range 80 to 8020 Hz (sixth order Butterworth with 36 dB/octave roll-off). EEQ processing was applied to each band independently, and the processed bands were summed to yield the final output. For the listeners with HI, NAL-RP amplification (Dillon, 2001, p. 241) based on each individual's hearing loss was applied to the UNP, EEQ1, and EEQ4 signals. Frequency-dependent NAL-RP amplification was achieved through the use of a 513-point finite impulse response digital filter.

Speech Plus Noise Stimuli

Examples of the effects of EEQ processing on S+N stimuli are provided in Figures 2 and 3. In these figures, the speech signal is a male utterance of the syllable /apa/ at a level of 65 dB SPL in background noise at an SNR of

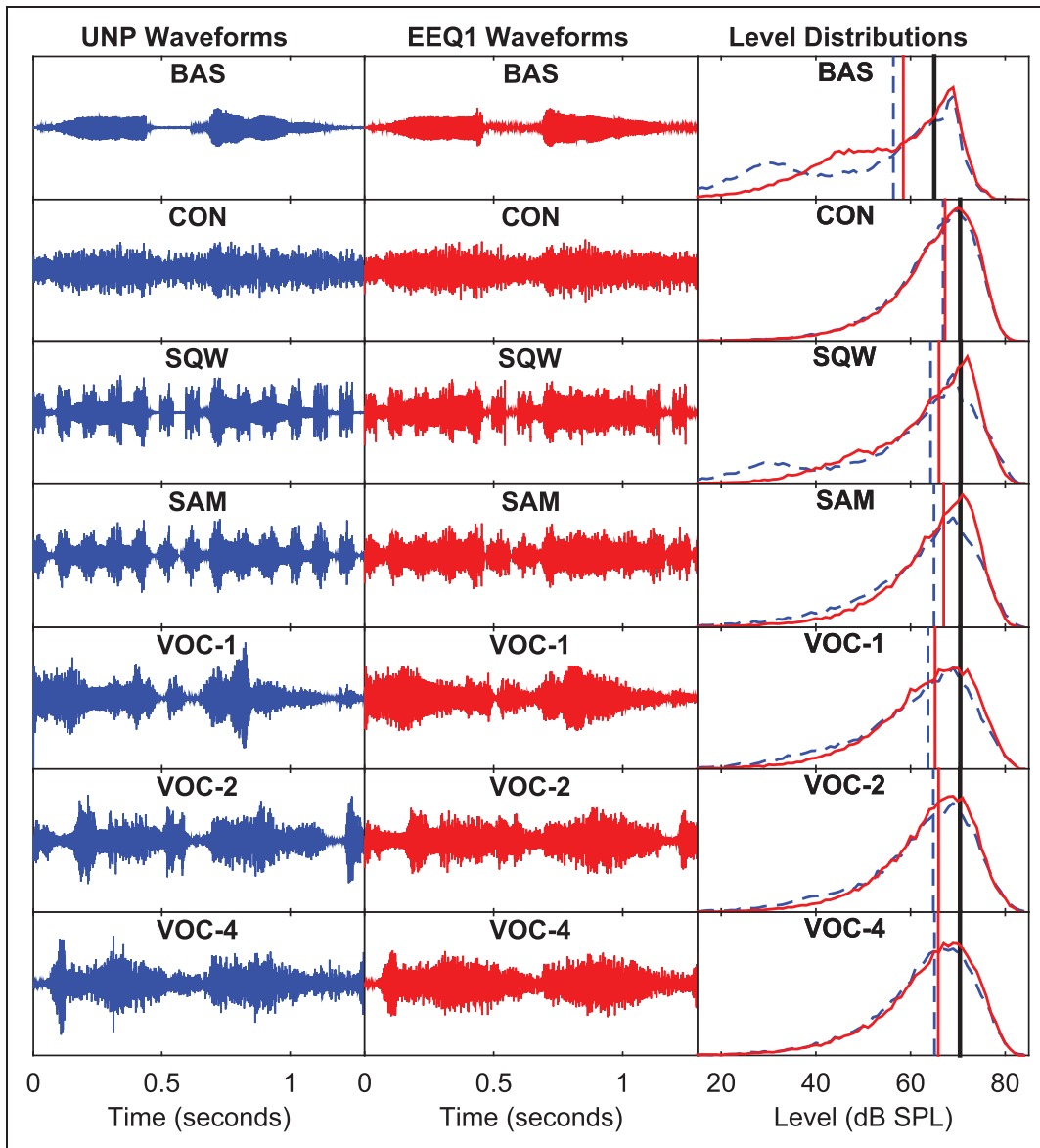


Figure 2. Waveforms and level distribution plots for the VCV token /apa/ presented with the seven different kinds of background interference (BAS, CON, SQW, SAM, VOC-1, VOC-2, and VOC-4) with UNP and EEQ1 for speech at 65 dB SPL and SNR of -4 dB (except for BAS where SNR is $+35$ dB). In the level distribution plots, the dashed blue line represents UNP, and the solid red line represents EEQ1. The thick black vertical line (rightmost in all plots) represents the RMS value (which is identical for both types of processing); the dashed blue vertical line and the solid red vertical line are the medians of the UNP and EEQ1 level distributions, respectively.

-4 dB (except for the BAS noise condition). Figure 2 provides a comparison of UNP and EEQ1 processing in each of the seven background noise conditions. S+N waveforms are shown for UNP and EEQ1 stimuli in the first two columns, respectively, with each row representing one of the seven noise conditions. The third column shows the distribution of the amplitude of the S+N signal in dB SPL for both types of processing. These distributions were generated by sampling at a rate of 44.1 kHz, converting the raw sample values of the S+N

stimuli into dB levels, grouping them into 1-dB bins, normalizing the distribution by the total number of samples and plotting the results over the range 15 to 85 dB SPL. The RMS value of the S+N signal in dB SPL is shown by the thick vertical line (about 65 dB SPL for BAS and 70 dB SPL for the remaining noises). The medians of the UNP and EEQ1 distributions are shown by the dashed and solid vertical lines, respectively.

For the UNP waveforms, differences among the noise conditions are evident. For example, the 10-Hz

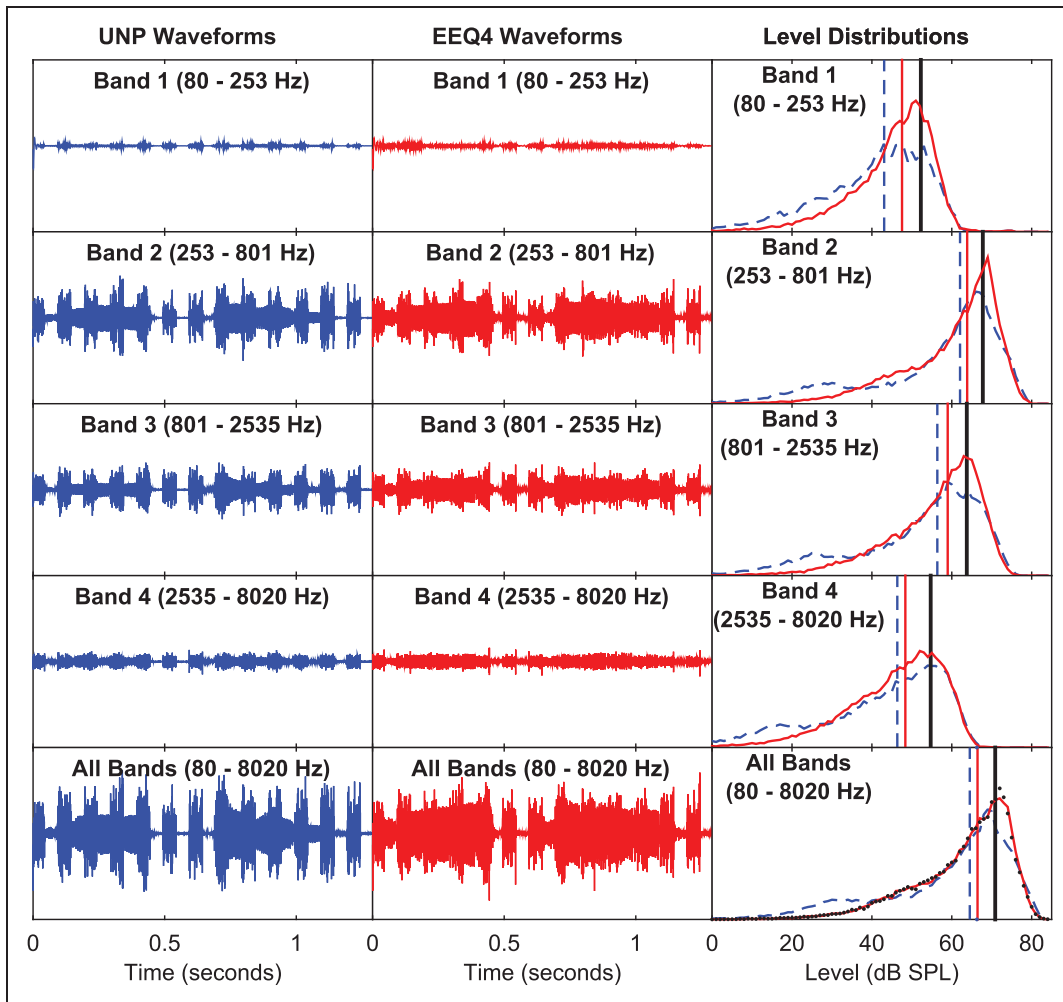


Figure 3. Waveforms and level distribution plots for the VCV token /apa/ in SQW noise with EEQ4 for speech at a level of 65 dB SPL and SNR of -4 dB. Each of the first four rows in the figure corresponds to a different logarithmically equal band in the range of 80 Hz to 8020 Hz, and the final row shows the sum of the four bands. In the level distribution plots, the dashed blue line represents UNP and the solid red line represents EEQ4. The thick black vertical line (rightmost in all plots) represents the RMS value (which is identical for both types of processing); the dashed blue vertical line and the solid red vertical line are the medians of the UNP and EEQ4 level distributions, respectively. For convenience, the last row plots the level distribution of the SQW noise for EEQ1 (given by the dotted curve) from Figure 2.

fluctuations of SQW and SAM are visible when compared with CON. Similarly, greater variations in level are visible for VOC-1 than for VOC-4 (with VOC-2 intermediate between these two). In the UNP amplitude distributions, the BAS noise is visible in the form of a peak at 30 dB SPL for BAS, SQW, and VOC-1 (but is not visible for the remaining noises, which dominate BAS throughout most of the signal duration). The effect of EEQ1 processing is most evident in the SQW and SAM conditions, where the lower energy speech components that are present during the low-level periods of the fluctuating interference are greater in energy in the EEQ1 signals. The reduction in amplitude variation is reflected in the rightward shift in the distributions for EEQ1 compared with UNP, despite the RMS value

remaining constant between UNP and EEQ. For the distributions shown in Figure 2, the differences in median levels between EEQ1 and UNP were 1.8 dB for SQW and 2.0 dB for SAM. Smaller differences were observed for the remaining noises: VOC-1 (1.5 dB), VOC-2 (1.1 dB), VOC-4 (0.8 dB), and CON (0.4 dB).

Figure 3 compares the UNP and EEQ4 waveforms and amplitude distributions for SQW noise on a band-by-band basis. As in Figure 2, results for a speech level of 65 dB SPL and an SNR of -4 dB are shown. The first four rows show plots for the individual bands which are logarithmically equal in width with ranges of 80 to 253 Hz (Band 1), 253 to 801 Hz (Band 2), 801 to 2535 Hz (Band 3), and 2535 to 8020 Hz (Band 4). Band 2 has the largest RMS value, followed by, in decreasing

order, Band 3, Band 4, and Band 1. The bottom row depicts the sum of the four bands. The amplitude distribution for EEQ1 (which has been added to the rightmost panel of the bottom row for comparison purposes) is similar to that for the wideband EEQ4. The only minor difference between these two plots is a slightly higher peak value for EEQ1.

Test Procedure

Experiments were controlled by a desktop PC using MatlabTM software. The digitized S+N stimuli were played through a 24-bit PCI sound card (E-MU 0404 by Creative Professional, Scotts Valley, CA) and then passed through a programmable attenuator (Tucker-Davis PA4, TDT, Alachua, FL) and a headphone buffer (Tucker-Davis HB6, TDT, Alachua, FL) before being presented monaurally to the listener in a soundproof booth via a pair of headphones (Sennheiser HD580, Old Lyme, CT). A monitor, keyboard, and mouse were located within the soundproof booth.

Consonant identification was tested using a 1-interval, 16-alternative, forced-choice procedure without correct-answer feedback. On each 64-trial run, 1 of the 64 tokens from the test set was selected randomly without replacement. A randomly selected noise equal in duration to that of the speech token was scaled to achieve the desired SNR and then added to the speech token. The resulting stimulus was either presented unprocessed (for the UNP conditions) or processed according to EEQ1 or EEQ4 before being presented to the listener for identification. The listener's task was to identify the medial consonant of the VCV token that had been presented by selecting a response (using a computer mouse) from a 4 × 4 visual array of orthographic representations associated with the consonants. No time limit was imposed on the listeners' responses. Each run typically lasted 3 to 5 min. Chance performance was 6.25% correct.

Experiment 1. Listeners with NH were tested using a speech level of 60 dB SPL. An SNR of -10 dB (selected to yield roughly 50%-correct performance for UNP stimuli in CON noise) was used for all noise conditions (except for BAS). Each HI listener selected a comfortable speech level when listening to UNP speech in the BAS condition. For these listeners, the SNR was selected to yield roughly 50%-correct performance for UNP speech in CON noise, based on the results of several preliminary runs. The speech levels and SNRs for each HI listener are listed in Table 1.

The UNP condition was always tested first, based on the assumption that listeners with HI would have prior real-world familiarity with these signals. The test order of the EEQ1 and EEQ4 conditions was then randomized for each listener. The seven noises were tested in order

BAS first, followed by a randomized order of the remaining six noises (CON, SQW, SAM, VOC-1, VOC-2, and VOC-4). Five 64-trial runs were presented for each of the 21 conditions (3 Processing Types × 7 Noises). The first run was considered as practice and discarded. The final four test runs were used to calculate the percent-correct score for each condition.

Experiment 2. Supplemental data were collected for four of the listeners with HI (HI-2, HI-4, HI-5, and HI-7) to examine the effects of EEQ as a function of SNR and to compare the resulting psychometric functions to those obtained with unprocessed materials. Each of these four listeners was tested at two additional SNRs after completing Experiment 1. One of the additional SNRs was 4 dB lower than that employed in Experiment 1 and the other was 4 dB higher. This testing was conducted with UNP and EEQ1 using six types of noise: CON, SQW, SAM, VOC-1, VOC-2, and VOC-4. The test order for UNP and EEQ1 was selected randomly for each listener. For each processing type, the two additional values of SNR were presented in random order. Within each SNR, the test order of the six types of noises was selected at random. Five 64-trial runs were presented for 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 for each of the 24 additional conditions (2 Processing Types × 6 Noises × 2 SNRs). Other than the SNR, all experimental parameters remained the same as for Experiment 1.

Data Analysis

For each condition and listener, percent-correct scores were averaged over the final four runs (consisting of 4 × 64 = 256 trials). A NMR (as defined by Léger et al., 2015) was calculated as:

$$\text{NMR} = (\text{FN} - \text{CN}) / (\text{BN} - \text{CN})$$

where FN is the score in fluctuating noise, CN is the score in continuous noise, and BN is the score in baseline noise.

In this metric, MR (which is defined as the numerator in the equation above) is represented as a fraction of the total possible amount of improvement (given by the denominator). NMR is useful for comparing performance among listeners with HI whose baseline scores are different and who require different test SNRs to achieve the target performance of 50% correct in continuous noise. By using baseline performance as a reference, NMR emphasizes the differences in performance with interrupted and continuous noise for an individual listener as opposed to the differences due to factors such as the severity of the hearing loss of the listener or the

distorting effects of the processing on the speech itself. NMR is preferable to the use of MR due to the influence of SNR on the size of MR. Previous studies (e.g., Bernstein & Grant, 2009; Oxenham & Simonson, 2009) have shown a tendency for an increase in MR with a decrease in SNR. The NMR measure takes into account the SNR at which a given listener was testing (as reflected in the continuous-noise score) and thus allows for better comparisons among listeners tested at different values of SNR. Furthermore, the use of NMR (as opposed to MR) discounts situations where MR arises as the result of a decrease in performance in continuous noise versus an increase in performance in fluctuating noise: the denominator in the former case is larger than that in the latter case, which results in a lower NMR.

Statistical tests, including ANOVAs (at a confidence level of .01) and post hoc Tukey-Kramer multiple comparisons (at a confidence level of .05), were conducted on NMR results.

The NMR calculations for the non-speech-derived SQW and SAM noises used CON noise as the continuous noise, and the NMR calculations for the speech-

derived VOC-1 and VOC-2 noises used VOC-4 noise as the continuous noise. These NMR formulas are listed here:

$$NMR_{SQW \text{ or } SAM} = \frac{(\text{SQW or SAM Score}) - \text{CON Score}}{\text{BAS Score} - \text{CON Score}}$$

$$NMR_{VOC-1 \text{ or } VOC-2} = \frac{(\text{VOC-1 or VOC-2 Score}) - \text{VOC-4 Score}}{\text{BAS Score} - \text{VOC-4 Score}}$$

Results

Experiment 1

The results of Experiment 1 are summarized in Figure 4 in terms of mean percent-correct scores across the four listeners in the NH group and across the nine listeners in the HI group. Individual results are also shown for each HI listener. For each of the seven types of background interference, scores are plotted for each of the three types of processing (UNP, EEQ1, and EEQ4).

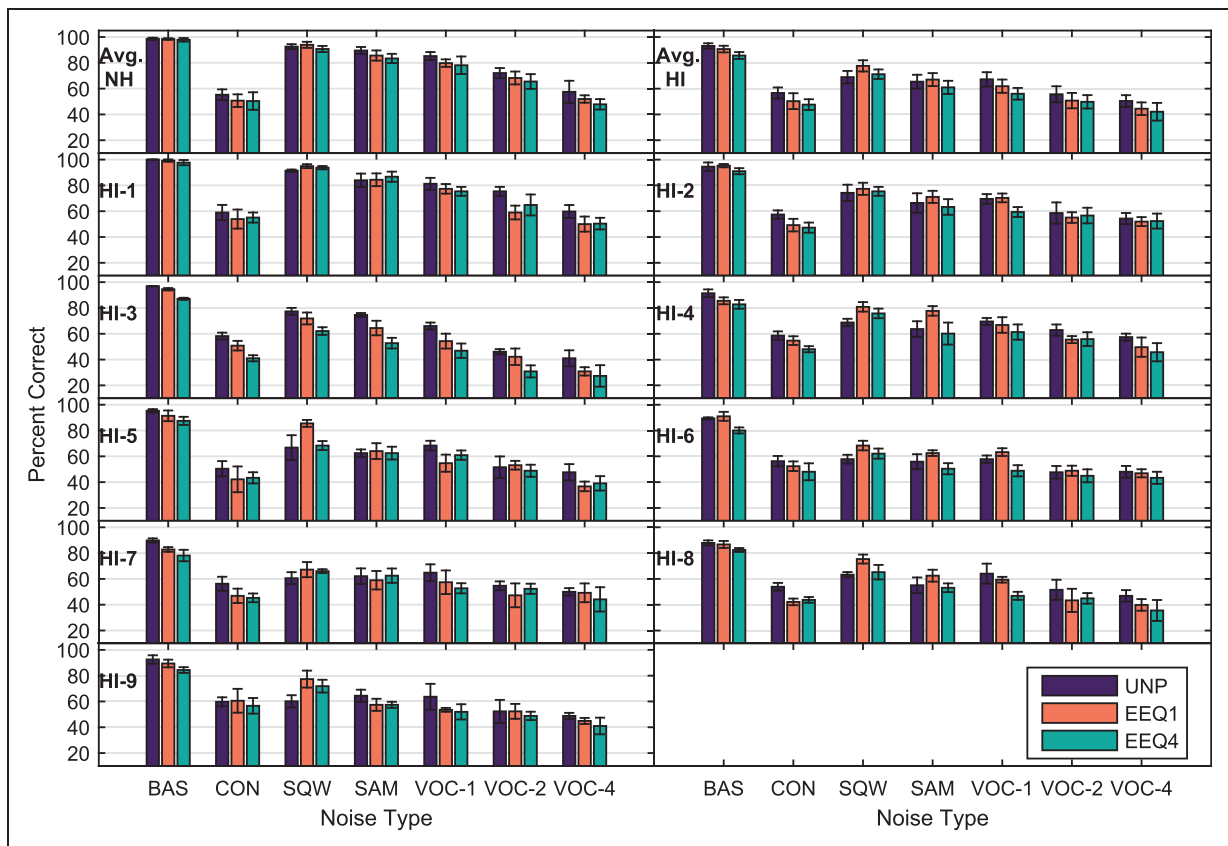


Figure 4. Mean percent-correct scores across the four listeners with NH (upper left panel) and the nine listeners with HI (upper right panel) in Experiment 1. The remaining nine panels provide individual percent-correct scores of the individual listeners with HI. The scores were measured with UNP (purple bars), EEQ1 (orange bars), and EEQ4 (green bars) with BAS, CON, SQW, SAM, VOC-1, VOC-2, and VOC-4 backgrounds. The error bars show ± 1 SD.

indicate ± 1 SD around the means. Listeners with HI exhibited slightly more variability in their results than did those with NH: the mean standard deviations across listeners (computed as the average of the mean standard deviations for each of the seven noises¹) in percentage points were 3.6 for UNP, 3.2 for EEQ1, and 4.4 for EEQ4 for listeners with NH and 4.7 for UNP, 4.9 for EEQ1, and 4.6 for EEQ4 for listeners with HI. In nearly all processing and noise conditions, with the exception of CON (which by design yielded scores of roughly 50% correct for both groups of listeners), the performance of the NH group exceeded that of HI group. Both groups, however, showed a decrease in performance for processed compared with unprocessed stimuli: averaged across noise types, the mean NH scores were 79% in UNP, 76% in EEQ1, and 73% in EEQ4, and the mean HI scores were 65% in UNP, 63% in EEQ1, and 59% in EEQ4. Because the UNP condition was always tested first, it is possible that its advantages over EEQ1 and EEQ4 are minimized here. Within each processing type, the performance for both groups was highest in

the BAS condition, lowest in CON and VOC-4 (the latter of which was derived from samples of enough speakers to behave similarly to continuous noise, as shown previously by Rosen et al., 2013; Simpson & Cooke, 2005), and intermediate between BAS and CON for the remaining noises. Averaged across the different listeners and processing types, the NH scores were 98% in BAS, 52% in CON, 92% in SQW, 86% in SAM, 81% in VOC-1, 69% in VOC-2, and 52% in VOC-4, and the HI scores were 90% in BAS, 52% in CON, 72% in SQW, 64% in SAM, 62% in VOC-1, 52% in VOC-2, and 46% in VOC-4.

The NMR results calculated from the scores for Experiment 1 are summarized in Figure 5. The two panels on the left show results for the non-speech-derived SQW (upper) and SAM (lower) noises, and the two panels on the right show results for the speech-derived VOC-1 (upper) and VOC-2 (lower) noises. Within each panel, NMR is plotted for each of the three types of processing for mean scores across listeners with NH, mean scores across listeners with HI, and individual listeners with HI.

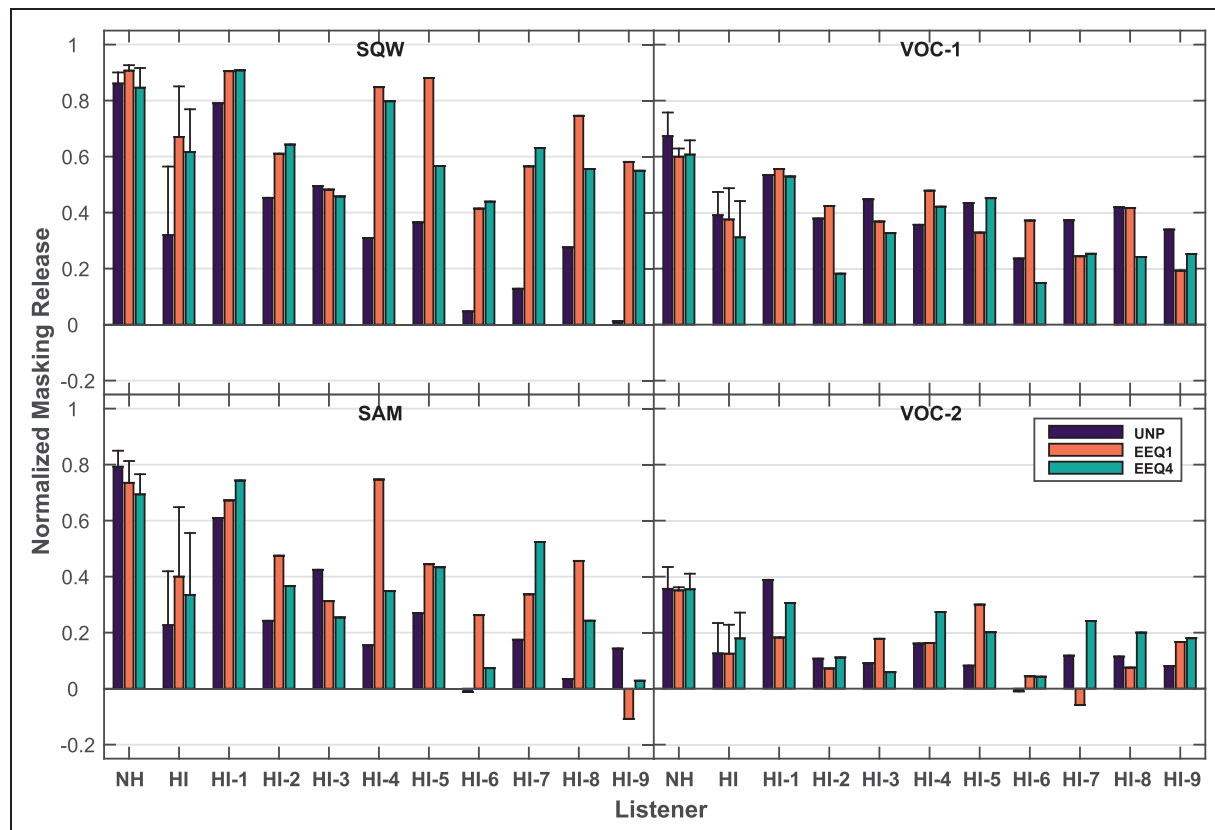


Figure 5. Mean NMR for the listeners with NH (first group of bars), mean NMR for the listeners with HI (second group of bars), and individual NMR for each of the listeners with HI (remaining nine groups of bars) with UNP (purple bars), EEQ1 (orange bars), and EEQ4 (green bars). Error bars show 1 SD. The NMR for the SQW (upper left panel) and SAM (lower left panel) noises was calculated relative to the CON condition, whereas the NMR for the VOC-1 (upper right panel) and VOC-2 (lower right panel) noises was calculated relative to the VOC-4 noise condition.

For the listeners with NH, values of NMR decreased in the order of SQW (0.87), SAM (0.74), VOC-1 (0.63), and VOC-2 (0.35) but were generally similar across the three types of processing. A repeated-measures ANOVA with main effects of noise condition and processing type was conducted on the NMR values of the four listeners with NH. This gave a significant effect of noise condition, $F(3, 9) = 227.24$, $p < .0001$, but not of processing type, $F(2, 6) = 2.23$, $p = .19$. There was no the interaction of processing by noise type, $F(6, 18) = 1.68$, $p = .18$. A post hoc Tukey-Kramer comparison of the noise effect indicated significant differences between all possible pairs of noises.

Among the individual listeners with HI, an effect of processing type was observed primarily with the SQW and SAM noises. Higher values of NMR were obtained for EEQ1 and EEQ4 than for UNP for SQW noise for all listeners with HI except HI-3 and for SAM noise for all except HI-3 and HI-9. The average NMR for SQW noise increased from 0.32 for UNP to 0.64 and 0.62 for EEQ1 and EEQ4, respectively. For SAM noise, the HI NMR means increased from 0.23 for UNP to 0.40 and 0.34 for EEQ1 and EEQ4, respectively. For the two speech-derived noises (VOC-1 and VOC-2), however, NMR tended to be similar across the three types of processing within individual listeners with HI. Averaged across listeners with HI, NMR values for UNP, EEQ1, and EEQ4, respectively, were 0.39, 0.38, and 0.31 for VOC-1 noise and 0.13, 0.12, and 0.18 for VOC-2. A repeated-measures ANOVA conducted on the NMR values of the listeners with HI showed significant effects of processing type, $F(2, 16) = 8.15$, $p = .004$, and noise type, $F(3, 24) = 48.06$, $p < .0001$, and a processing by noise interaction, $F(6, 48) = 8.53$, $p < .0001$. Post hoc Tukey-Kramer comparisons of the processing effect indicated that the NMR values obtained with EEQ1 and EEQ4 processing were significantly greater than for UNP. For the noise effect, post hoc comparisons indicated that the NMR obtained with SQW noise was greater than that obtained with the other three noises, and that the NMR with SAM and VOC-1 (not significantly different from each other) led to significantly higher NMR than with VOC-2 noise. Finally, a post hoc analysis of the processing by noise interaction indicated the following: for SQW noise, NMR for both EEQ1 and EEQ4 (not significantly different from each other) was greater than for UNP; for SAM noise, EEQ1 resulted in higher NMR than UNP; and for both VOC-1 and VOC-2, no significant differences were obtained among the three types of processing.

Experiment 2

The results of Experiment 2 are plotted in Figure 6 for each of the four listeners with HI. Data for the

non-speech-derived noises (SQW, SAM, and CON) are shown in the panels on the left and those for the speech-derived noises (VOC-1, VOC-2, and VOC-4) are shown on the right. Percent-correct scores are plotted as a function of SNR for UNP and EEQ1 processing for each type of noise along with sigmoidal fits to each of these psychometric functions. The sigmoidal fits assumed a lower bound corresponding to chance performance on the consonant-identification task (6.25% correct) and an upper bound corresponding to a given listener's score on the BAS condition for UNP or EEQ. The fitting process estimated the slope and midpoint values of a logistic function that minimized the error between the fit and the data points as summarized in Table 2.

The results of the fits are summarized in Table 2 in terms of their midpoints in dB and slopes around the midpoint (in percentage points per dB). A repeated-measures ANOVA on the midpoint values with factors processing type and noise type gave significant effects of noise, $F(5, 15) = 40.55$, $p < .0001$, and the interaction of processing by noise, $F(5, 15) = 16.13$, $p < .0001$, but not of processing, $F(1, 3) = 18.75$, $p = 0.03$. A post hoc Tukey-Kramer comparison of the main effect of noise condition indicated that the midpoint for SQW noise (-22.5 dB) was significantly lower than for all other noises, that the midpoints for CON, VOC-2, and VOC-4 noises were not significantly different from each other (mean of -3.8 dB) but were significantly lower than those for the other three types of noise, and that the midpoints for SAM and VOC-1 noises (mean of -10.8 dB) were not different from each other but were significantly different from those for the remaining noises. The interaction effect was due to the higher midpoints for EEQ1 than for UNP for SQW and SAM noises but similar midpoint values for the two types of processing for the remaining noises. A repeated-measures ANOVA conducted on the slopes of the psychometric functions gave a significant effect of noise condition, $F(5, 15) = 14.33$, $p < .0001$, but not of processing type, $F(1, 3) = 2.7$, $p = .20$, or of processing by noise interaction, $F(5, 15) = 1.86$, $p = .16$. A post hoc comparison of the noise effect indicated that the slope for the SQW noise (1.5%/dB) was significantly shallower than for the other noises; that the slopes for the CON and VOC-4 noises (not different from each other and averaging 4.6%/dB) were significantly steeper than for the other noises; and that the slopes for SAM, VOC-1, and VOC-2 (not different from each other and averaging 3.0%/dB) were significantly different from those for the three remaining noises.

NMR values for SQW, SAM, VOC-1, and VOC-2 noises were calculated from the percent-correct scores for obtained for each SNR and processing type for each listener with HI. In Figure 7, NMR for EEQ1 is plotted as a function of NMR for UNP for the

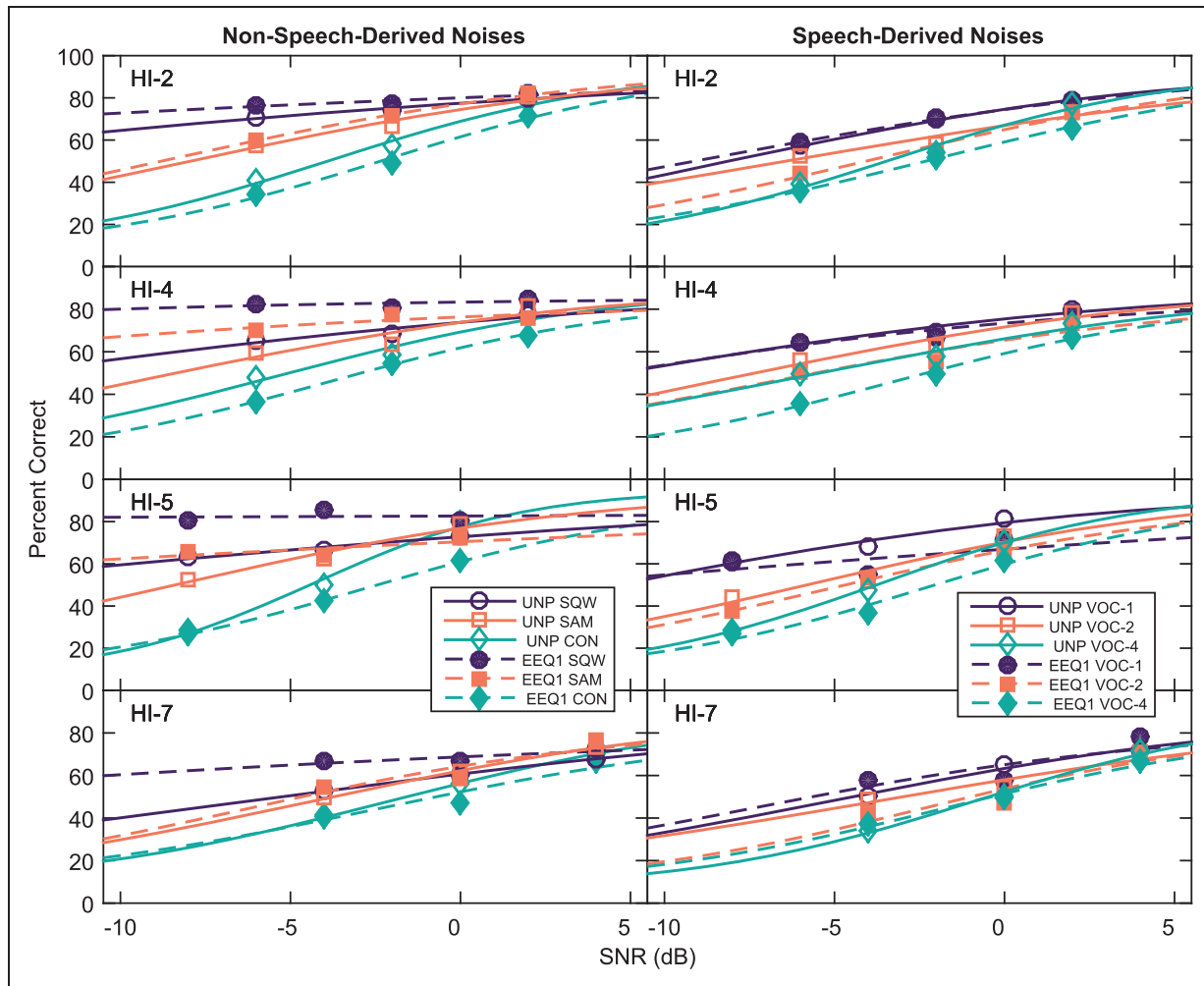


Figure 6. Percent-correct scores plotted as a function of SNR in dB for UNP (unfilled symbols) and EEQ1 (filled symbols). Each row shows results for one of the four listeners with HI. Data for the non-speech-derived noises (SQW noise in purple circles, SAM noise in orange squares, and CON noise in green diamonds) are shown in the panels on the left, and data for the speech-derived noises (VOC-1 noise in purple circles, VOC-2 noise in orange squares, and VOC-4 noise in green diamonds) on the right. Sigmoidal fits to each of these functions are shown with data points connected by continuous lines for UNP conditions and dashed lines for EEQ1 conditions.

individual listeners with HI in SQW and SAM noise at the various SNRs (in the panels on the left) and for VOC-1 and VOC-2 noise (on the right). In the panels on the left, it can be seen that every NMR data point lies above the 45-degree reference line, showing a strong tendency for larger NMR with EEQ1 for non-speech-derived noises at all SNRs tested. Additionally, NMR was greater with SQW interference than with SAM interference. For SQW noise, NMR averaged across the four listeners with HI at the low, mid, and high SNRs was 0.43, 0.31, and -0.10 , respectively, for UNP and 0.76, 0.66, and 0.56, respectively, for EEQ1. For SAM noise, these values were 0.28, 0.21, and 0.14, respectively, for UNP and 0.50, 0.50, and 0.47, respectively, for EEQ1. As shown in the right column of Figure 7, there was a smaller difference in NMR for UNP and EEQ1 for the

speech-derived noises than for the non-speech-derived noises. Averaged over listeners and SNRs, NMR was greater with VOC-1 than with VOC-2 noise for both types of processing. For UNP and EEQ1, respectively, NMR was 0.33 and 0.46 for VOC-1 and 0.14 and 0.17 for VOC-2.

A repeated measures ANOVA with factors SNR, processing type, and noise condition was conducted on the NMR values shown in Figure 7. Although a tendency was observed for a decrease in NMR with increasing SNR, this effect did not reach significance, $F(2, 6) = 5.22$, $p = .0486$. Significant effects were found for the other two main factors, Processing: $F(1, 3) = 44.01$, $p = .007$; Noise: $F(3, 9) = 36.55$, $p < .0001$, and for the 3 two-way interactions. The source of the significant interaction between SNR and processing type, $F(2, 57) = 7.87$,

Table 2. Midpoints (M) of SNR in dB and Slopes (S) Around the Midpoints in Percentage Points per dB of Sigmoidal Fits to the Data of the Four Individual Listeners with HI Shown in Figure 7.

	CON		SQW		SAM		VOC-1		VOC-2		VOC-4	
	M	S	M	S	M	S	M	S	M	S	M	S
Processing type: UNP												
HI-2	-3.8	5.1	-18.7	1.7	-7.8	3.5	-8.0	3.4	-6.3	2.8	-3.4	5.2
HI-4	-5.3	4.2	-13.9	2.1	-8.7	3.3	-11.8	2.7	-7.7	3.3	-5.8	3.2
HI-5	-4.3	7.2	-15.8	1.5	-8.2	3.7	-11.1	3.1	-5.6	3.7	-3.6	5.7
HI-7	-2.1	4.1	-6.2	2.1	-4.3	3.4	-5.1	3.2	-3.7	2.7	-0.9	5.0
Means	-3.9	5.2	-13.6	1.9	-7.3	3.5	-9.0	3.1	-5.8	3.1	-3.4	4.8
Processing type: EEQI												
HI-2	-2.2	5.0	-32.1	1.1	-8.6	3.5	-8.9	3.0	-3.9	3.8	-2.1	4.0
HI-4	-3.9	4.4	-37.3	1.9	-24.4	1.7	-13.1	2.6	-6.9	3.1	-3.3	4.2
HI-5	-2.7	4.7	<i>-300.1</i>	0.2	-24.2	1.0	-14.5	1.3	-5.0	3.7	-2.3	4.9
HI-7	-2.4	3.3	-24.4	1.2	-6.2	3.5	-7.4	3.1	-2.4	3.9	-1.8	3.9
Means	-2.8	4.3	-98.5	1.1	-15.9	2.4	-11.0	2.5	-4.5	3.6	-2.4	4.3

Note. HI = hearing impairment; CON = BAS plus additional continuous noise; SQW = BAS plus square-wave interrupted noise; SAM = BAS plus sinusoidally amplitude modulated noise; VOC-1 = one-talker vocoded noise; VOC-2 = two-talker vocoded noise; VOC-4 = four-talker vocoded noise. M and S are given for UNP and EEQI with six noise backgrounds: CON, SQW, SAM, VOC-1, VOC-2, and VOC-4. Means across listeners are provided in the final row. Note that the midpoint of HI-5 for UNP speech in SQW noise (-300.1 dB, shown in italics) was highly deviant relative to the remaining three HI listeners (whose midpoints ranged from -24.4 to -37.3 dB), due to a very small change in scores across the three values of SNR tested for HI-5 in this condition. In conducting the ANOVA on midpoint values, the value of -300.1 dB was replaced with the average of the midpoints of the other three listeners (i.e., -31.3 dB).

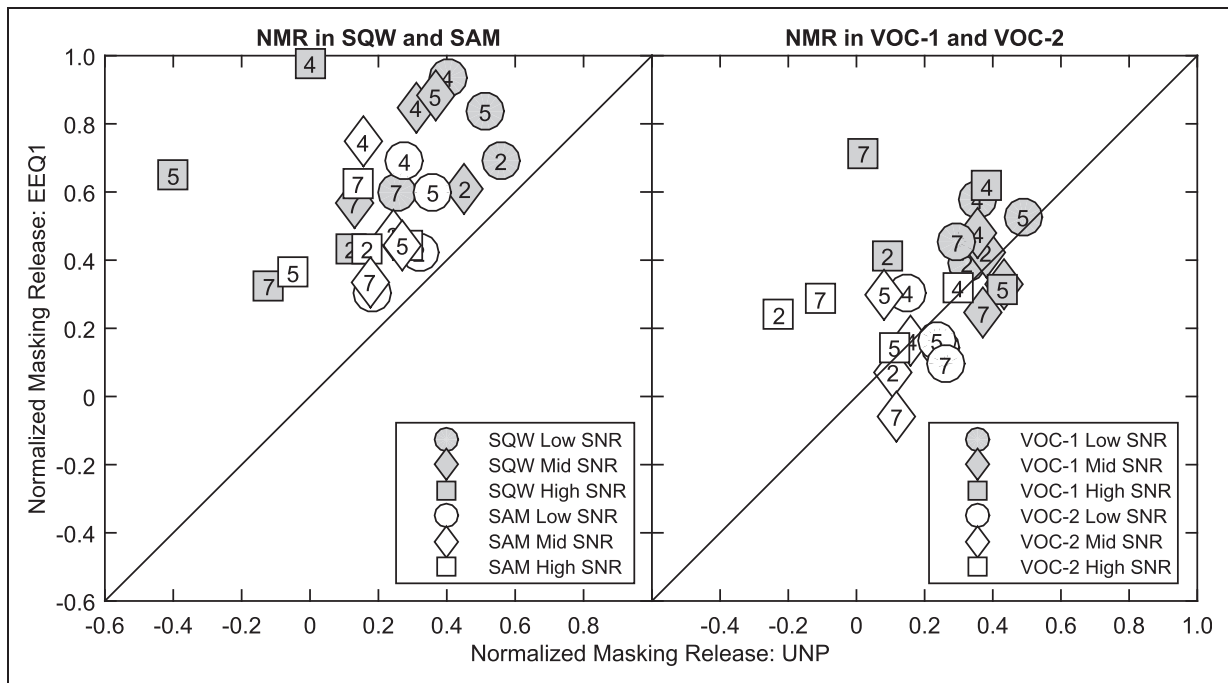


Figure 7. Normalized masking release (NMR) for EEQI plotted as a function of NMR for UNP for the four listeners with HI (HI-1, 2, 5, and 7, and depicted by numbers within the symbols). The data for the nonspeech-derived noises (SQW with filled symbols and SAM with unfilled symbols) are plotted on the left, and the data for the speech-derived noises (VOC-1 with filled symbols and VOC-2 with unfilled symbols) are plotted on the right. 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.

$p = .001$, lies in the decrease in NMR with increasing SNR for UNP but no significant change in NMR as a function of SNR for EEQ1 (confirmed by Tukey–Kramer post hoc comparisons). The significant interaction between SNR and noise condition, $F(6, 57) = 3.6$, $p = .0042$, was based on a different pattern of results for SQW noise relative to the other three noises (based on Tukey–Kramer post hoc comparisons). For SQW interference, larger NMR values were observed at low and mid SNRs than at a high SNR; for the other three noises, there were no significant differences in NMR across SNR. Finally, the significant processing by noise interaction, $F(3, 57) = 12.85$, $p < .0001$, arose from the observation that for SQW and SAM noises, NMR was significantly larger for EEQ1 than for UNP, whereas there was no difference between NMR for EEQ1 and UNP for the VOC-1 and VOC-2 noises (as confirmed by Tukey–Kramer post hoc comparisons).

Discussion

Effects of EEQ Processing on Speech Reception in Noise

The primary goal of EEQ was to increase the performance of listeners with HI in fluctuating interference while maintaining performance in the BAS and CON noise conditions. Compared with UNP, improvements in fluctuating noise with EEQ were observed for the non-speech-derived noises but not for the speech-derived noises. Among the listeners with HI, measures of NMR indicated benefits for EEQ1 for the SQW and SAM noises across a wide range of SNRs. A different pattern of results was observed for the speech-derived fluctuating noises (VOC-1 and VOC-2), however, where NMR was similar for UNP and EEQ. The psychometric functions shown in Figure 6 indicate that EEQ was successful in maintaining CON scores similar to those obtained for UNP across a wide range of SNRs, while leading to improved performance for the SQW and SAM noises at lower SNRs. Furthermore, performance levels for the BAS condition were maintained with EEQ1 (with HI scores averaging 91% for EEQ1 compared with 93% correct for UNP). Thus, the improvements to NMR arose primarily from improvements in scores for the SAM and SQW noises. The pattern of results obtained in Experiment 1 (where UNP was tested before EEQ) was highly similar to that obtained in Experiment 2 (where the presentation order of UNP and EEQ was randomized). This suggests that the benefits observed for EEQ on the SQW and SAM noises in Experiment 1 were not dependent on an order effect. Overall scores and benefits in fluctuating noises were lower for the multiband processing of EEQ4 than for the wideband processing of EEQ1.

Some insight into the improved performance with SQW and SAM noise for EEQ can be obtained from further examination of the amplitude distribution plots shown in Figure 2. Despite the RMS values being the same within a type of interference, the median amplitudes were greater for EEQ1 than for UNP. The upward shift in the median amplitudes with EEQ, which occurred due to the amplification of the lower energy S+N components, may be regarded as a measure of the impact of EEQ. These numbers indicate that the effect of EEQ is greatest for SQW and SAM, lowest for CON and VOC-4, and intermediate for VOC-1 and VOC-2. The movement of the tail of the amplitude distribution toward the center of the histogram corresponds to the reduction in amplitude variation in the processed stimuli.

The effects of EEQ across a range of hearing losses are shown in Figure 8, where NMR is plotted as a function of the 5-frequency PTA of each of the nine listeners with HI. For UNP, NMR decreased with increasing PTA; however, with EEQ1 and EEQ4, NMR varied much less with PTA, which highlights the benefits provided to listeners with HI by making the speech component of the signal more audible in the lower levels of the SQW noise (i.e., dips). Additionally, as shown in Figure 7, the increase in NMR with EEQ1 relative to UNP for SQW and SAM held at various SNRs: with UNP in these types of interference, NMR became close to zero or even negative at the high SNRs, whereas with EEQ1, NMR was always positive.

EEQ was not effective in improving NMR in the speech-derived noises. One explanation for this may lie in the fact that many listeners with HI demonstrated a greater NMR for VOC-1 and VOC-2 than for SQW and SAM noise for the UNP condition. As shown by Figure 5, the listeners with HI with the most severe hearing losses (HI-6, HI-7, HI-8, and HI-9) showed almost no NMR in the UNP condition with SQW (averaging 0.09) but did show a non-zero NMR (averaging 0.38) for VOC-1. In fact, in the UNP condition, the NMR for VOC-1 interference showed much less variability across listeners with HI (with a total range of 0.24 to 0.53) than was the case for SQW noise (with a range of 0.01 to 0.79). Thus, there was less room for NMR improvement with EEQ and the speech-derived noises.

Comparison With Compression Amplification

The performance with EEQ in modulated background noise can be compared with studies of compression-amplification aids for listeners with HI. Both methods of processing result in greater amplification of lower level sounds compared with higher level sounds. However, as described earlier, EEQ differs from compression amplification: the homogeneity of EEQ and its

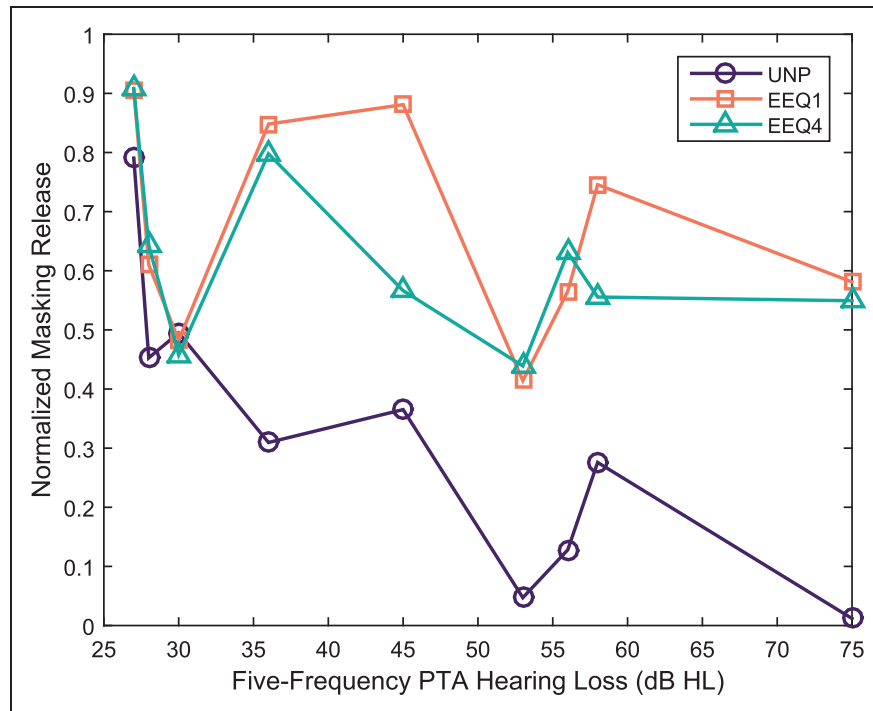


Figure 8. The NMR for Experiment I in the SQW condition attained by each of the listeners with HI with UNP (purple circles), EEQ1 (orange squares), and EEQ4 (green triangles) plotted as a function of PTA in dB HL.

operation on relative signal levels without detailed knowledge of the level-dependent characteristics of the hearing loss contrasts with the non-linearity of amplitude compression and its mapping of absolute signal levels to the specific hearing loss.

Of particular interest here are comparisons of EEQ results with studies of compression amplification that include fast attack or release times and evaluations in various types of modulated Gaussian background noise (e.g., Brennan et al., 2016; Houben, 2006; Moore, Peters, & Stone, 1999; Moore, Stainsby, Alcantara, & Kuhnel, 2004; Nordqvist & Leijon, 2004; Olsen, Olofsson, & Hagerman, 2004; Souza, Boike, Witherell, & Tremblay, 2007; Stone, Moore, Alcantara, & Glasberg, 1999). Moore et al. (1999) examined wideband and multiband compression algorithms using a compression ratio of 2 and effective attack and release times of 7 ms. For measurements of speech-reception thresholds (SRTs) for sentences in a background of a noise that was modulated by the temporal envelope of a single talker, only the one-channel system led to an improvement over a linear-gain system (by 2 dB). Stone et al. (1999) implemented four different compression algorithms in a wearable digital hearing aid and measured SRTs of listeners with HI for sentences presented in speech-shaped noise or in speech-shaped noise whose amplitude was modulated by the envelope of a single talker. Two of the systems employed fast-acting compression with the possibility of

amplifying low-level portions of the signal present during the gaps in the modulated noise. The results indicated that SRTs in modulated noise were lower (i.e., better) than for unmodulated noise (by roughly 2 dB on average) and that the best performance in modulated noise was obtained with the DUAL-HI fast-acting compression systems. Nordqvist and Leijon (2004) implemented an automatic gain-control system with four components of spectral shape (Bustamante & Braida, 1987). Their system consisted of an underlying target gain which was modified (to adapt to changes in environments, talker, speech/pause, etc.) using relative signal levels. However, the underlying target gain itself is dependent on absolute sound levels. When compared with the DUAL-HI version of the Stone et al. system (1999) on the intelligibility of sentences (Hagerman, 1982), the two systems were found to be nearly identical both for unmodulated and modulated noises. Olsen et al. (2004) measured the SRT for sentences in a modulated noise background through a linear-gain system and five versions of a three-channel, fast-acting compressor. On average, the listeners with HI performed similarly across the five compression systems and slightly worse than for the linear system. Moore et al. (2004) studied consonant reception with VCV syllables for amplitude-compression systems presented through a programmable hearing aid with a range of attack or release times. With one-talker interference, performance for a linear-gain

system was superior to that of any of the compression algorithms. Houben (2006) examined sentence reception in continuous noise and in a speech-derived fluctuating background noise using a wide range of compression settings (including variations in compression ratio, number of channels, and attack or release times). None of the compression systems yielded performance better than that obtained with linear-gain amplification in either continuous or fluctuating noise. Souza et al. (2007) measured word-recognition scores in sentences for a single-channel wide dynamic range compression system (with a compression ratio averaging 2 and attack or release times of 5/50 ms) in backgrounds of steady noise and a modulated noise derived from the temporal envelope of a 12-talker babble. On average, scores were several percentage points higher for linear than for compression amplification and for steady than for amplitude-modulated noise. Finally, Brennan et al. (2016) measured recognition of words in sentences for two compression systems that differed in their attack and release times (one fast and one slow) in backgrounds of continuous noise and modulated noise that was derived from a two-talker temporal envelope. MR was on the order of 1 dB for both compression speeds.

The results obtained here with EEQ and the speech-derived noises are generally consistent with those obtained in previous studies employing compression amplification and similar types of modulated backgrounds. In those studies that included a linear-amplification condition for comparison with compression amplification for speech reception in modulated noise backgrounds (Houben, 2006; Moore et al., 1999, 2004; Olsen et al., 2004; Souza et al., 2007), there was little evidence for improved performance with compressive systems (with the exception of the 2-dB improvement observed by Moore et al., 1999). In studies that included steady noise in addition to a modulated noise background, MR was either absent (Houben, 2006; Souza et al., 2007) or on the order of 1 to 2 dB (Brennan et al., 2016; Stone et al., 1999) for compression amplification. For those studies which employed speech-derived modulation maskers and for which MR was measured for both linear and compression amplification (Houben, 2006; Souza et al., 2007), there was no evidence of an increase in MR with compression. Thus, for the most part, these results (including those with fast attack/release times) are similar to those observed for EEQ with the VOC-1 and VOC-2 maskers which indicate similar performance for UNP and EEQ.

Note that, unlike EEQ processing, all of the compression amplification systems discussed here are customized to the particular hearing loss and operate on absolute sound level. Given its independence on hearing loss

and absolute level, it is possible that EEQ could be used as a preprocessor to one of these more traditional compression amplification techniques.

Comparison of EEQ1 and EEQ4

Despite the initial hypothesis that multiband EEQ might be beneficial for listeners with frequency-dependent hearing losses, overall performance was in fact worse for EEQ4 than for EEQ1. The amplitude distributions of the broadband EEQ1 and EEQ4 signals were similar (see Figure 3, bottom row, rightmost panel). In decibels, the absolute values of the differences in mean levels between EEQ1 and EEQ4 were 0.2 for BAS, 0.3 for CON, 0.3 for SQW, 0.3 for SAM, 0.2 for VOC-1, 0.7 for VOC-2, and 0.6 for VOC-4. However, by applying different scale factors to different frequency bands, the independent-band processing may have interfered with the spectral shape, resulting in decreased effectiveness.

To see if this might be the case, outputs of the three processing schemes were examined for each of the four bands used for EEQ4. In Figure 9, median levels for UNP, EEQ1, and EEQ4 within each of the four bands used in EEQ4 are plotted as a function of SNR. For UNP, the median levels generally decrease linearly with increasing SNR, whereas the slopes for the EEQ1 and EEQ4 functions level off at around 0 dB SNR. This is consistent with EEQ amplifying the low-energy speech components. The effect of EEQ on spectral shape is reflected in the relative median levels of the four bands. In general, EEQ1 preserves the relative median levels for UNP, which indicates that it has little impact upon the spectral shape of the stimulus. EEQ4, however, produces changes in relative levels. Specifically, at low SNRs, band 1 is amplified relative to band 4 (indicating a shift in energy toward lower frequencies), while at high SNRs, band 1 is attenuated relative to band 4 (indicating a shift in energy toward higher frequencies). These changes in spectral shape may explain the decreased performance for EEQ4 relative to EEQ1, and other metrics might reveal a larger difference in spectral shape between the two processing schemes. It is also possible that the additional processing involved in the multiband scheme introduced additional distortions to the signal, which led to the observed decreases in performance with EEQ4 compared with EEQ1.

Glimpse Analysis of Vcoded and Non-Vcoded Noises

A glimpse analysis, derived from work of Cooke (2006), was conducted to determine the role of opportunities for glimpses in the various noises in explaining the greater effectiveness of EEQ for improving performance in the non-speech-derived compared with the speech-derived noises. Cooke (2006) defined a glimpse as a connected

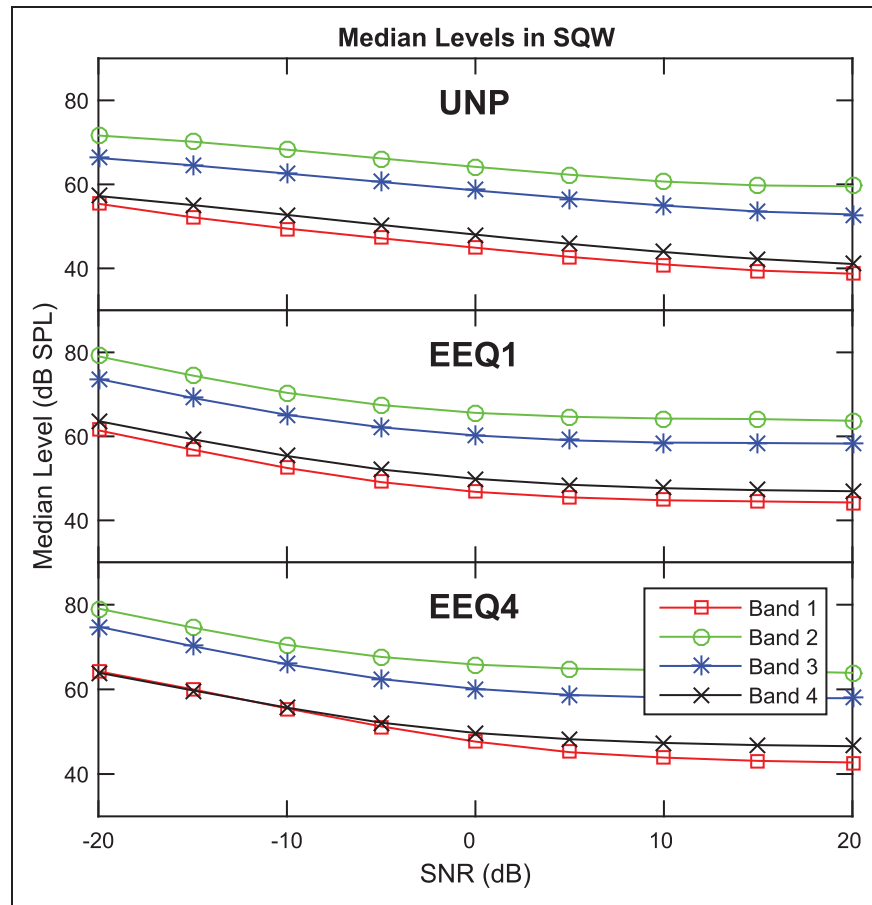


Figure 9. The median level of the syllable /apa/ presented at 70 dB SPL in SQW interference with UNP, EEQ1, and EEQ4. The values in four logarithmically spaced frequency bands in the range of 80 to 8020 Hz are plotted as a function of SNR: Band 1, red squares; Band 2, green circles; Band 3, blue asterisks; and Band 4, black x's.

region of the spectrotemporal excitation pattern in which the energy of the speech token exceeded that of the background by at least 3 dB in each time-frequency element. Unlike Cooke's analysis, the current analysis measured glimpses derived from the envelopes of the noises alone.

A noise glimpse was defined here as a section of the noise where the envelope dropped more than 3 dB below the RMS value of the noise for at least 10 ms (based on estimates of the effective duration of the auditory temporal window from Moore, Glasberg, Plack, & Biswas, 1988; Oxenham & Moore, 1994). An example of the method used to calculate glimpses is shown in Figure 10 for VOC-1, VOC-2, and VOC-4 noise samples, depicting their waveforms (top row of panels) and envelopes (bottom row). The envelope plots include the RMS value, the threshold value (defined as RMS - 3 dB), and the location and duration of glimpses. The envelope was computed by filtering the signal to a bandwidth of 80 to 8020 Hz and passing the absolute value of its Hilbert transform through a low-pass filter with a cutoff of 32 Hz. An interval was considered to contain a noise gap if the envelope level was below the threshold line

for at least 10 ms. As can be seen in Figure 10, as the number of speakers who made up the speech-derived noises increased from one to two to four, the envelope hovered closer to the RMS value, and the duration of the glimpses decreased.

The analysis was conducted for six of the noises used in the experiment (eliminating only the BAS noise). Five hundred samples of each of the noise types were generated to have a duration equal to an arbitrarily chosen VCV token of 1.29 s. For each noise sample, the occurrences of glimpses using the above definition were determined. This information was used to calculate the percentage of time that the glimpses were present, the number of glimpses per second, and the average length of the glimpses.

The average fraction of time spent in a glimpse in decreasing order was 0.43 for SQW, 0.40 for SAM, 0.37 for VOC-1, 0.32 for VOC-2, 0.24 for VOC-4, and 0.00 for CON. VOC-1 was therefore similar to SQW and SAM in terms of fraction of the time spent in a gap, whereas VOC-4 had more gaps than CON using the current metric. The number of glimpses per second was

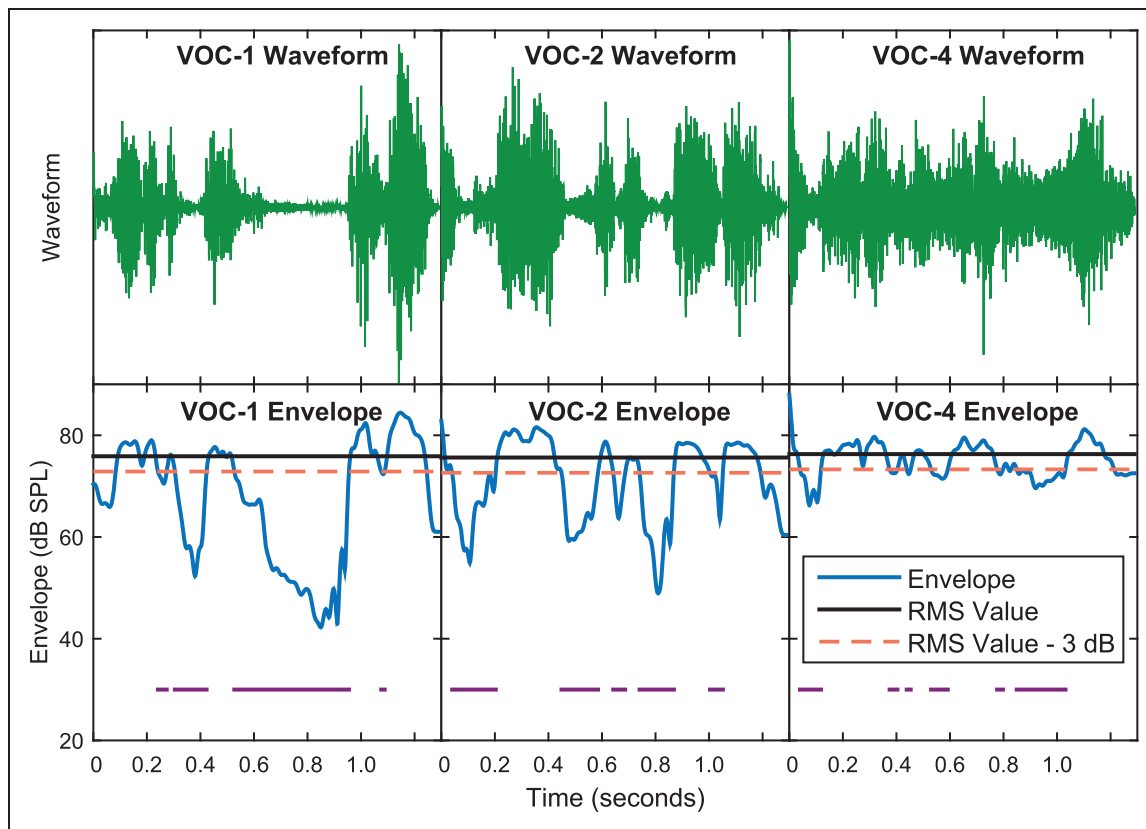


Figure 10. The waveforms (upper row) and envelopes (lower row) of sample VOC-1, VOC-2, and VOC-4 noises. Shown together with the blue envelope trace (in dB SPL) are the RMS value (shown by the solid black line) and the RMS value -3 dB (shown by the dashed orange line). The purple horizontal lines correspond to the location and duration of the glimpses in the noise.

greater for SQW (9.09) and SAM (9.03) than for the speech-derived noises of VOC-1 (2.97), VOC-2 (4.16), and VOC-4 (4.32), and fell to 0.05 for CON. Finally, the average glimpse durations for the speech-derived noises showed greater variability and were longer than for the non-speech-derived noises. These ranges were between 0 and 15 ms for CON, 45 and 49 ms for SQW, and 40 and 47 ms for SAM, compared with ranges of 0 to 856 ms for VC-1, 0 to 405 ms for VC-2, and 10 to 300 ms for VC-4.

Taken together, these analyses offer insight into why EEQ performed better with the non-speech-derived noises than with the speech-derived noises. Although VOC-1, SQW, and SAM had similar amounts of total time spent in glimpses, these times were distributed over a greater number of glimpses for SQW and SAM. Miller and Licklider (1950) showed that recognition of words in backgrounds of interrupted noise with a given noise-fraction time was dependent on the rate of interruption but not on whether the interruptions occurred with regular or irregular spacing. An interruption rate of 10 Hz (as used for the SQW and SAM noises) led to maximal word scores while performance dropped for slower rates (as in VOC-1 with an interruption rate of roughly 3 Hz)

as well as faster rates. With VOC-2 and VOC-4, there were both less total time spent in glimpses and fewer glimpses than with SQW and SAM. EEQ performed best with short, frequent glimpses, as this gave it the best opportunity to amplify speech during the gaps in the noise. With VOC-1, there were fewer but longer glimpses providing the listener with samples restricted to only a few temporal regions of the utterance rather than with shorter-duration samples throughout the entire length of the utterance. During the longer glimpses, the long-term average would be reduced, leading to smaller changes in gain in these sections. With fewer and longer glimpses (and therefore fewer and longer non-glimpses as well), it is also possible that the entirety of the low-energy consonant portion of the speech stimuli would be covered by noise. Thus, EEQ may have less of an opportunity to operate effectively on the portion of the speech where listeners with HI require the most amplification and could instead end up amplifying noise during these parts. On the other hand, the longer glimpses available in VOC-1 noise may explain the positive NMR seen for the listeners with HI with UNP signals. Specifically, the availability of longer speech segments could improve

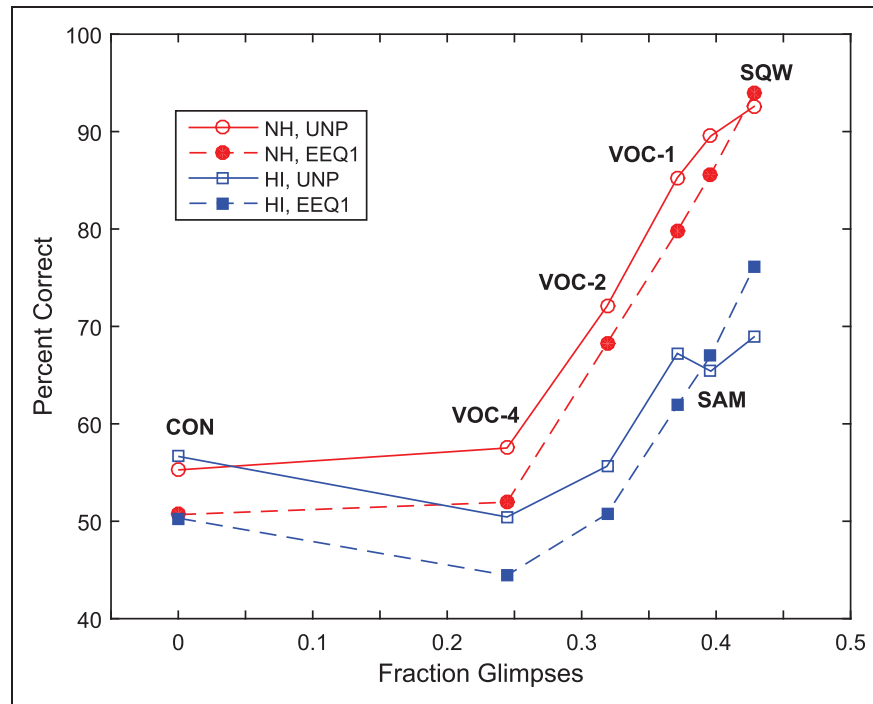


Figure 11. The scores averaged across the listeners with NH (red circles) and the listeners with HI (blue squares) for UNP (unfilled symbols connected by solid lines) and EEQ1 (filled symbols connected by dotted lines) plotted as a function of the fraction glimpses for CON (0.00), SQW (0.43), SAM (0.40), VOC-1 (0.37), VOC-2 (0.32), and VOC-4 (0.24).

speech intelligibility through a reduction of the effects of temporal masking (e.g., Dubno, Horwitz, & Ahlstrom, 2003), as compared with the case of SQW where the on and off segments of the noise alternated every 50 ms.

Previous studies have examined the role of glimpsing on speech intelligibility by manipulating various parameters associated with the temporal interruption of the speech signal, either in quiet or in backgrounds of continuous noise (e.g., Kidd & Humes, 2012; Li & Loizou, 2007; Miller & Licklider, 1950; Wang & Humes, 2010). In these studies, the proportion of glimpse time was found to have the most influence on intelligibility, compared with little or no effect of other factors, such as interruption rate, number of glimpses, or duration of glimpses. In Figure 11, the mean NH and HI scores for UNP and EEQ1 are plotted as a function of the average fraction of the noise spent in a glimpse. For both processing types and groups of listeners, scores increased with an increase in the fraction of glimpses once this measure exceeded approximately 0.25. Below this fraction, scores were roughly constant at the level observed for CON. For listeners with both NH and HI, the UNP curve lies above the EEQ1 curve for the smaller fractions of glimpses. However, as the fraction of glimpses increases, the difference between the curves gets smaller and even reverses at the highest fractions of glimpses. The greater effectiveness of EEQ for SQW and SAM noises may be attributed to the greater fraction of time

spent in glimpses, accumulated from periodic glimpses occurring at a rate of 10 Hz.

Conclusions

- Real-time EEQ was effective in improving NMR for listeners with HI for SQW and SAM interference. The EEQ effect on NMR was not apparent for VOC-1 and VOC-2 interference. These observations held across various SNRs.
- NMR improvements for EEQ resulted primarily from increased performance in fluctuating noise, especially in SQW interference. There was also a small decrease in performance in BAS and CON for EEQ.
- EEQ was more effective with regular and frequent gaps in the fluctuating noises, as for SQW and SAM. VOC-1 and VOC-2 had gaps that were more variable in length and limited the effectiveness of EEQ in using the short and long windows to normalize energy.
- EEQ1 was more effective than EEQ4. EEQ4 may have interfered with spectral-shape perception, resulting in decreased effectiveness.
- NMR decreased with increasing hearing loss for unprocessed stimuli but was roughly independent of degree of loss for EEQ1. This resulted in a large increase in NMR for the listeners with the most severe hearing losses.

Acknowledgments

This article is based on research conducted by Laura A. D'Aquila in conjunction with the M. Eng. degree granted by MIT in June, 2016. This research was carried out within the Sensory Communication Group of the Research Laboratory of Electronics under the supervision of Ms. D'Aquila's coauthors (J. G. Desloge, C. M. Reed, and L. D. Braida).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R01 DC000117.

Note

1. The mean standard deviation for a given noise and processing condition is calculated as $\sqrt{\sum_{i=1}^n \sigma_i^2 / n}$, where σ_i^2 is the variance of the four recorded runs on listener i and n is the number of listeners.

References

- Bernstein, J. G. W., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *125*(5), 3358–3372.
- Braida, L. D., Durlach, N. I., De Gennaro, S. V., Peterson, P. M., & Bustamante, D. K. (1982). Review of recent research on multi-band amplitude compression for the hearing impaired. In G. A. Studebaker, & F. H. Bess (Eds.), *The Vanderbilt hearing aid report: Monographs in contemporary audiology* (pp. 123–140). Upper Darby, PA: York Press.
- Brennan, M., McCreery, R., Kopun, J., Lewis, D., Alexander, J., & Stelmachowicz, P. (2016). Masking release in children and adults with hearing loss when using amplification. *Journal of Speech, Language, and Hearing Research*, *59*(1), 110–121.
- Boothroyd, A., Hanin, L., & Hnath, T. (1985). *A sentence test of speech reception: Reliability, set equivalence, and short term learning*. Speech and Hearing Sciences Report RC110. City University New York, New York.
- Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America*, *84*(1), 101–114.
- Bustamante, D. K., & Braida, L. D. (1987). Principal-component amplitude compression for the hearing-impaired. *Journal of the Acoustical Society of America*, *82*(4), 1227–1242.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, *119*(3), 1562–1573.
- De Gennaro, S., Braida, L. D., & Durlach, N. I. (1986). Multichannel syllabic compression for severely impaired listeners. *Journal of Rehabilitation Research and Development*, *23*(1), 17–24.
- Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., & D'Aquila, L. A. (2017). Technique to improve speech intelligibility in fluctuating background noise by normalizing signal energy over time. *Journal of the Acoustical Society of America*, (accepted).
- Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., & Delhorne, L. A. (2010). Speech reception by listeners with real and simulated hearing impairment: Effects of continuous and interrupted noise. *Journal of the Acoustical Society of America*, *128*(1), 342–359.
- Dillon, H. (2001). *Hearing Aids* (pp. 239–247). New York, NY: Thieme.
- Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2003). Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing. *Journal of the Acoustical Society of America*, *113*(4), 2084–2094.
- Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed., pp. 124–139). New York, NY: Harper.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, *88*(4), 1725–1736.
- Füllgrabe, C., Berthommier, F., & Lorenzi, C. (2006). Masking release for consonant features in temporally fluctuating background noise. *Hearing Research*, *211*(1–2), 74–84.
- Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology*, *11*(2), 79–87.
- Hopkins, K., & Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *Journal of the Acoustical Society of America*, *125*(1), 442–446.
- Hopkins, K., Moore, B. C. J., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *Journal of the Acoustical Society of America*, *123*(2), 1140–1153.
- 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.
- IEEE. (1969). *IEEE recommended practice for speech quality measurements*. New York, NY: Institute of Electrical and Electronics Engineers, Inc.
- Kidd, G. R., & Humes, L. E. (2012). Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. *Journal of the Acoustical Society of America*, *131*(2), 1434–1448.
- Léger, A. C., Reed, C. M., Desloge, J. G., Swaminathan, J., & 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. *Journal of the Acoustical Society of America*, *138*(1), 389–403.

- Li, N., & Loizou, P. C. (2007). Factors influencing glimpsing of speech in noise. *Journal of the Acoustical Society of America*, 122(2), 1165–1172.
- Lippman, R. P., Braida, L. D., & Durlach, N. I. (1981). Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. *Journal of the Acoustical Society of America*, 69(2), 524–534.
- Lorenzi, C., Debrulle, L., Garnier, S., Fleuriot, P., & Moore, B. C. J. (2009). Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *Journal of the Acoustical Society of America*, 125(1), 27–30.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Science USA*, 103(49), 18866–18869.
- Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 22(2), 167–173.
- Moore, B. C. J. (2014). *Auditory processing of temporal fine structure: Effects of age and hearing loss* (pp. 81–102). Hackensack, NJ: World Scientific.
- Moore, B. C. J., Glasberg, B. R., Plack, C. J., & Biswas, A. K. (1988). The shape of the ear's temporal window. *Journal of the Acoustical Society of America*, 83(3), 1102–1116.
- Moore, B. C. J., Peters, R. W., & Stone, M. A. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *Journal of the Acoustical Society of America*, 105(1), 400–411.
- Moore, B. C. J., Stainsby, T. H., Alcantara, J. I., & Kuhnel, V. (2004). The effect on speech intelligibility of varying compression time constants in a digital hearing aid. *International Journal of Audiology*, 43(7), 399–409.
- Nordqvist, P., & Leijon, A. (2004). Hearing-aid automatic gain control adapting to two sound sources in the environment, using three time constants. *Journal of the Acoustical Society of America*, 116(5), 3152–3155.
- Olsen, H. L., Olofsson, A., & Hagerman, B. (2004). The effect of presentation level and compression characteristics on sentence recognition in modulated noise. *International Journal of Audiology*, 43(5), 283–294.
- Oxenham, A. J., & Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends in Hearing*, 18(1), 1–14.
- Oxenham, A. J., & Moore, B. C. J. (1994). Modeling the additivity of nonsimultaneous masking. *Hearing Research*, 80(1), 105–118.
- Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *Journal of the Acoustical Society of America*, 125(1), 457–468.
- Phatak, S., & Grant, K. W. (2014). Phoneme recognition in vocoded maskers by normal-hearing and aided hearing-impaired listeners. *Journal of the Acoustical Society of America*, 136(2), 859–866.
- Reed, C. M., Desloge, J. G., Braida, L. D., Léger, A. C., & Perez, Z. D. (2016). Level variations in speech: Effect on masking release in hearing-impaired listeners. *Journal of the Acoustical Society of America*, 140(1), 102–113.
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London Series B*, 336(1278), 367–373.
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *Journal of the Acoustical Society of America*, 133(4), 2431–2443.
- Shannon, R. V., Jansvold, A., Padilla, M., Robert, M. E., & Wang, X. (1999). Consonant recordings for speech testing. *Journal of the Acoustical Society of America*, 106(6), L71–L74.
- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N -talker babble is a nonmonotonic function of N (L). *Journal of the Acoustical Society of America*, 118(5), 2775–2778.
- Souza, P. E., Boike, K. T., Witherell, K., & Tremblay, K. (2007). Prediction of speech recognition from audibility in older listeners with hearing loss: Effects of age, amplification, and background noise. *Journal of the American Academy of Audiology*, 18(1), 54–65.
- Stone, M. A., Füllgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *Journal of the Acoustical Society of America*, 132(1), 317–326.
- Stone, M. A., Moore, B. C. J., Alcantara, J. I., & Glasberg, B. R. (1999). Comparison of different forms of compression using wearable digital hearing aids. *Journal of the Acoustical Society of America*, 106(6), 3603–3619.
- Wang, X., & Humes, L. E. (2010). Factors influencing recognition of interrupted speech. *Journal of the Acoustical Society of America*, 128(4), 2100–2111.