Original Article

Effects of Expanding Envelope Fluctuations on Consonant Perception in Hearing-Impaired Listeners

Trends in Hearing
Volume 22: 1–12
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216518775293
journals.sagepub.com/home/tia
SAGE

Alan Wiinberg[1], Johannes Zaar[1], and Torsten Dau[1]

Abstract

This study examined the perceptual consequences of three speech enhancement schemes based on multiband nonlinear expansion of temporal envelope fluctuations between 10 and 20 Hz: (a) "idealized" envelope expansion of the speech before the addition of stationary background noise, (b) envelope expansion of the noisy speech, and (c) envelope expansion of only those time-frequency segments of the noisy speech that exhibited signal-to-noise ratios (SNRs) above −10 dB. Linear processing was considered as a reference condition. The performance was evaluated by measuring consonant recognition and consonant confusions in normal-hearing and hearing-impaired listeners using consonant-vowel nonsense syllables presented in background noise. Envelope expansion of the noisy speech showed no significant effect on the overall consonant recognition performance relative to linear processing. In contrast, SNR-based envelope expansion of the noisy speech improved the overall consonant recognition performance equivalent to a 1- to 2-dB improvement in SNR, mainly by improving the recognition of some of the stop consonants. The effect of the SNR-based envelope expansion was similar to the effect of envelope-expanding the clean speech before the addition of noise.

Keywords

consonant recognition, hearing impairment, hearing instruments, speech enhancement, temporal envelope

Date received: 10 August 2017; revised: 13 February 2018; accepted: 16 February 2018

Introduction

People with a sensorineural hearing impairment often complain about difficulties understanding speech in situations with several interfering talkers or background noise, particularly in reverberant environments. Some of these difficulties are considered to be caused by loudness recruitment, reflecting a reduced sensitivity to soft sounds and a steeper loudness growth function than observed in normal-hearing (NH) people (e.g., Fowler, 1936; Steinberg & Gardner, 1937). Modern hearing aids attempt to compensate for loudness recruitment by applying multiband dynamic-range compression (DRC) that provides level-dependent amplification in various frequency bands, such that soft sounds are amplified more than higher level portions of the sound. Apart from reduced audibility, cochlear hearing loss is often associated with a "distortion loss" that is considered to reflect suprathreshold processing deficits and assumed to be caused by inner hair-cell damage or loss of auditory-nerve fibers and synapses (e.g., Festen & Plomp, 1990; Plomp, 1978). One of the perceptual consequences of a distortion loss could be a reduced ability to capture and discriminate envelope fluctuations in a sound (e.g., Schlittenlacher & Moore, 2016; Wiinberg, Jepsen, Epp, & Dau, 2018). The course of the envelope of speech in different frequency bands has been shown to be crucial for speech intelligibility (e.g., Shannon, Zeng, & Kamath, 1995; Stone, Anton, & Moore, 2012; Stone, Füllgrabe, & Moore, 2008) and contains information related to voicing, manner, and place of articulation (Xu, Thompson, & Pfingst, 2005). In the case of a background noise, the modulation depth of the speech envelope becomes reduced because of the less varying noise envelope. This commonly deteriorates speech

[1] Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark

Corresponding author:
Alan Wiinberg, Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark.
Email: alwiin@elektro.dtu.dk

intelligibility, particularly in listeners with a hearing impairment (e.g., Stone et al., 2008, 2012).

It has been proposed that artificially increasing the modulation depth of the speech envelope may facilitate the extraction of speech cues and thereby improve speech intelligibility in noise (e.g., Plomp, 1988). Increasing the modulation depth of the speech envelope, without affecting the noise, would increase the signal-to-noise ratio (SNR) in the modulation domain, which has been shown to be related to speech intelligibility (Jørgensen, Decorsière, & Dau, 2015). Consistent with this idea, recent speech intelligibility models based on the SNR in the modulation domain have been able to account for the effects of a large range of interferers and distortion types on speech intelligibility in NH listeners (Chabot-Leclerc, Jørgensen, & Dau, 2014; Chabot-Leclerc, MacDonald, & Dau, 2016; Jørgensen & Dau, 2011; Jørgensen, Ewert, & Dau, 2013).

Different implementations of temporal envelope enhancement schemes have been investigated, with varying degree of success. Several studies found significant benefits from envelope expansion of the speech before the addition of noise both in NH and hearing-impaired (HI) listeners (e.g., Apoux, Tribut, Debruille, & Lorenzi, 2004; Langhans & Strube, 1982). However, the idealized processing of the speech before the addition of noise requires a priori knowledge of the clean speech signal, which cannot be assumed in practice (e.g., in hearing-aid signal processing schemes). If envelope expansion is instead applied to the noisy speech mixture, both the speech fluctuations and the intrinsic noise fluctuations are enhanced, such that no benefit in terms of the SNR in the modulation domain can be expected. In fact, consistent with this reasoning, several studies that applied envelope expansion to the noisy speech showed no benefit or even a decreased performance relative to linear processing (Freyman & Nerbonne, 1996; Van Buuren, Festen, & Houtgast, 1999) while others showed small benefits (e.g., Apoux et al., 2004; Clarkson & Bahgat, 1991). These results were typically consistent across NH and HI listeners when reduced audibility was compensated for by amplification. Part of the large variability regarding the benefit of envelope expansion across the different studies may have been caused by (a) differences in the details of the expansion schemes employed (e.g., the number of frequency bands, the range of modulation frequencies in which an expansion was applied, envelope thresholding, the amount of expansion, etc.), (b) differences in the modulation spectra of the (stationary vs. fluctuating noise) interferers and the speech material (e.g., sentences vs. consonant-vowel nonsense syllables [CVs]) as well as (c) differences in the tested stimulus SNRs.

In most studies, the envelope expansion was applied to the "entire" modulation-frequency range (e.g., between 0 and 500 Hz). The modulation power of long-term speech typically has a maximum around the syllabic rate, which is about 4 Hz for English, and decays thereafter with increasing modulation frequency (e.g., Plomp, 1988). Boosting modulation frequencies in this low-frequency range around the syllabic rate therefore enhances the overall dynamic range of the speech signal. Consequently, low-level speech segments are suppressed such that they may fall below the detection threshold while high-level speech segments may become uncomfortably loud, particularly for HI listeners with loudness recruitment. Therefore, audibility effects might contribute to the detrimental effects observed with some of the proposed expansion schemes. Using an alternative approach, Langhans and Strube (1982) applied expansion only at modulation frequencies *above* a lower cutoff modulation frequency of 2 Hz and provided DRC for slow envelope fluctuations (below 2 Hz). The idea behind this approach was that the DRC could compensate for loudness recruitment while the amplitude expansion could enhance speech envelope cues above 2 Hz. Langhans and Strube reported substantial benefits in NH listeners in terms of speech intelligibility when the processing was applied before the addition of noise. Even though this expansion scheme was successful in such idealized conditions, it might not be advantageous when applied at modulation frequencies as low as 2 Hz in the case of HI listeners with loudness recruitment as stimulus audibility might be affected. Alternatively, an enhancement of higher frequency modulations (e.g., above 10 Hz) may increase the robustness of stop consonants and vowel onsets without compromising audibility. For example, the intelligibility of /t/ utterances has been shown to be highly correlated with the detectability of the transient in the release burst when presented in noise (Li, Menon, & Allen, 2010; Régnier & Allen, 2008).

This study investigated the effects of expanding modulation frequencies (in the range from 10 to 20 Hz) on consonant recognition and consonant confusions in NH and HI listeners using consonant-vowel nonsense syllables (CVs) mixed with stationary Gaussian noise. It was hypothesized that the considered envelope expansion will improve the recognition of stop consonants and that detrimental effects caused by the enhancement of noise fluctuations will be minimized if the envelope expansion processing is only applied to those time-frequency segments that are dominated by speech. Three different envelope expansion methods were tested: (a) "Idealized" envelope expansion of the speech before the addition of noise, (b) envelope expansion of the noisy speech, and (c) envelope expansion of only those time-frequency segments of the noisy speech that exhibited SNRs above a certain limit. Linear processing was considered as the reference condition. Loss of audibility was compensated for by providing individual linear frequency-dependent amplification for the HI listeners. The experimental data were analyzed with respect to overall and consonant group–specific consonant recognition scores, as well as in terms of listener-specific consonant recognition scores.

Methods

Listeners

Two groups of listeners participated in the experiments, an NH group and a HI group. The NH group consisted of eight adults with a median age of 26 years and ages ranging from 21 to 61 years. All had absolute thresholds below 20 dB HL for the octave frequencies between 0.125 and 8 kHz. The HI group consisted of 12 adults with symmetrical mild- to moderately-severe sensorineural hearing losses. The median age was 72 years and the range was 50 to 80 years. The absolute thresholds for the test ear, measured using conventional audiometry, are shown in Figure 1. All listeners reported Danish as their first language, signed an informed consent document, and were reimbursed for their efforts. Approval for the study was granted by the Scientific Ethical Committee of the Capital Region in Denmark (De Videnskabsetiske Komitéer for Region Hovedstaden).

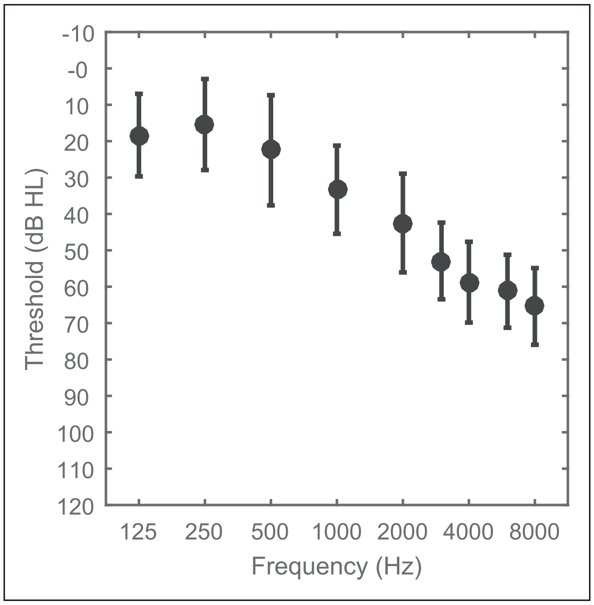

Figure 1. Mean absolute thresholds for the tested ear of the HI listeners, measured using conventional manual audiometry, and expressed in dB HL. Error bars represent ±1 standard deviation. HI = hearing impaired.

Stimuli

The CVs consisted of 15 consonants (/p, t, k, b, d, g, f, s, ʃ, v, j, l, h, m, n/) followed by the vowel /i/. Two tokens (one recording of a female talker and one of a male talker) were selected per CV from the Pitu Danish nonsense syllable speech material (Christiansen, 2011), amounting to 30 tokens overall (15 CVs × two talkers). The tokens represent a subset of the speech tokens used in a recent study on consonant perception in white noise (Zaar & Dau, 2015) which considered three recordings of each CV per talker. For each CV, the most intelligible recording of each talker was selected in this study. The levels of the tokens were equalized using VUSOFT, a software implementation of an analog VU-meter developed by Lobdell and Allen (2007), such that all CVs showed the same VUSOFT peak value. This equalization strategy is mainly based on the vowel levels, thus ensuring realistic relations between the levels of the individual consonants. After equalization, the reference speech level for the SNR calculation was defined as the overall root-mean-square (RMS) level averaged across all speech tokens.

SNR conditions of 12, 6, and 0 dB were generated by fixing the noise level and adjusting the level of the speech tokens based on the reference speech level according to the desired SNR. The speech tokens were mixed with stationary Gaussian noise such that the speech token onset was temporally positioned 400 ms after the noise onset. The stimulus duration was 1 s, including 50-ms raised-cosine onset and offset ramps for the noise. The sound pressure level (SPL) of the noise was set to 65 dB, while the overall stimulus level differed depending on the level of the speech, that is, on the SNR. Envelope expanded signals (clean speech or noisy speech) were equalized in RMS level with the corresponding signals obtained without expansion processing. For the HI listeners, the stimuli were linearly amplified according to the NAL-R(P) frequency-dependent prescription rule based on their individual audiometric thresholds (Byrne, Parkinson, & Newall, 1990). The frequency-dependent amplification was provided using a bank of seven octave–wide bandpass linear-phase, finite-impulse-response (FIR) filters with center frequencies between 0.125 and 8 kHz.

Envelope Expansion Processing

The proposed multiband envelope expansion algorithm, depicted in Figure 2, is similar to the algorithm described in Langhans and Strube (1982). The input signal was short-time Fourier transformed by Hann-windowing the signal in time frames of 256 samples with 75% overlap between frames using a sampling rate of 44100 Hz. Each of the windowed segments was padded with 128 zeros at the beginning and the end and transformed to

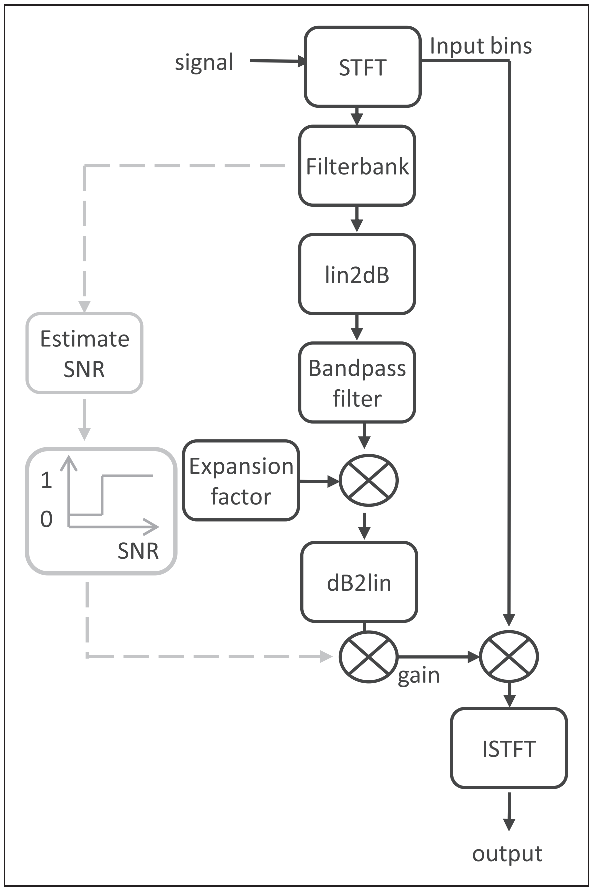

Figure 2. Block diagram of the proposed envelope expansion algorithm. First, the signal was windowed in time segments and transformed into the frequency domain by an STFT. The frequency bins in each time window were combined into 15 third-octave spaced frequency bands (Filterbank). The power in each band was converted to dB SPL (lin2dB) and bandpass filtered (Bandpass filter). The filtered temporal envelope was then multiplied by an expansion factor. The bandpass-filtered temporal envelope in each of the frequency bands was converted to linear units (dB2lin) and thereafter used as gain values for the input. For the SNR-based expansion scheme, indicated in gray, the gain was set to 0 dB for time-frequency bands with SNRs below a certain limit. Finally, an ISTFT was computed to generate the final expanded signal. ISTFT = inverse short-time Fourier transform; STFT = short-time Fourier transform.

Table 1. Overview of the Three Different Envelope Expansion Conditions.

Abbreviation	Processing	Expander mode
Exp_{mix}	Noisy speech	Envelope expansion
Exp_{SNR}	Noisy speech	SNR-based envelope expansion
$\text{Exp}_{\text{speech}}$	Clean speech	Envelope expansion

SNR = signal-to-noise ratio.

the spectral domain using a 512-point fast Fourier transform (FFT). The power spectral density of the resulting frequency bins was combined into 15 third-octave wide frequency bands with center frequencies between 0.323 and 8.192 kHz. The power in each band was converted to dB SPL, and the resulting logarithmic representation of the temporal envelope was bandpass filtered over time-frames using a zero-phase fourth-order Chebyshev Type II filter (−24 dB/octave roll-off) with 3-dB cutoff frequencies at 10 and 20 Hz. The bandwise expansion gains per timeframe were computed by multiplying the bandpass-filtered envelopes with a scaling factor of 1.3. Thus, a bandpass-filtered level of 1 dB resulted in an amplification of the output level by 1.3 dB. The value of the scaling factor was based on data from Wiinberg et al. (2018). The factor was chosen such that the expansion processing restored the average modulation-depth discrimination performance of the HI listener group to that of the NH listener group at a modulation frequency of 16 Hz. The bandwise gains were converted to linear units and smoothed in the frequency domain using a piecewise cubic interpolation to avoid aliasing artifacts. The frequency smoothed gains were applied to the bins of the short-time Fourier transformed input stimulus and an inverse FFT was applied to produce time segments of the envelope-expanded stimuli. These time segments were subsequently windowed with a Hann-window to avoid aliasing artifacts and combined using an overlap-add method to provide the processed temporal waveform.

For the SNR-based expansion scheme, a priori information about the speech and noise components of the noisy speech mixture was used. The power of both the speech and noise components was computed for each of the 15 frequency bands and the SNR was calculated in dB. For time-frequency segments with SNRs below −10 dB, the expansion gain was set to 0 dB. Otherwise, the expansion gain was not changed.

As listed in Table 1, three different envelope expansion settings were tested: Envelope expansion of the noisy speech (Exp_{mix}); envelope expansion applied to time-frequency segments with SNRs above −10 dB (Exp_{SNR}), and envelope expansion of the speech before the addition of noise ($\text{Exp}_{\text{speech}}$).

Figure 3 shows the temporal waveform of the male speech token \mi\ along with the waveforms obtained with the same speech token mixed with noise at 0-dB SNR for linear processing and the three envelope expansion conditions. For illustration purposes, only the results at the output of an auditory-inspired gammatone filter tuned to 500 Hz are shown. From the top, the panels show the temporal output of the gammatone filter for the clean speech, linear processing, Exp_{mix}, Exp_{SNR}, and $\text{Exp}_{\text{speech}}$ conditions, respectively. The illustration shows that $\text{Exp}_{\text{speech}}$ enhances only the speech

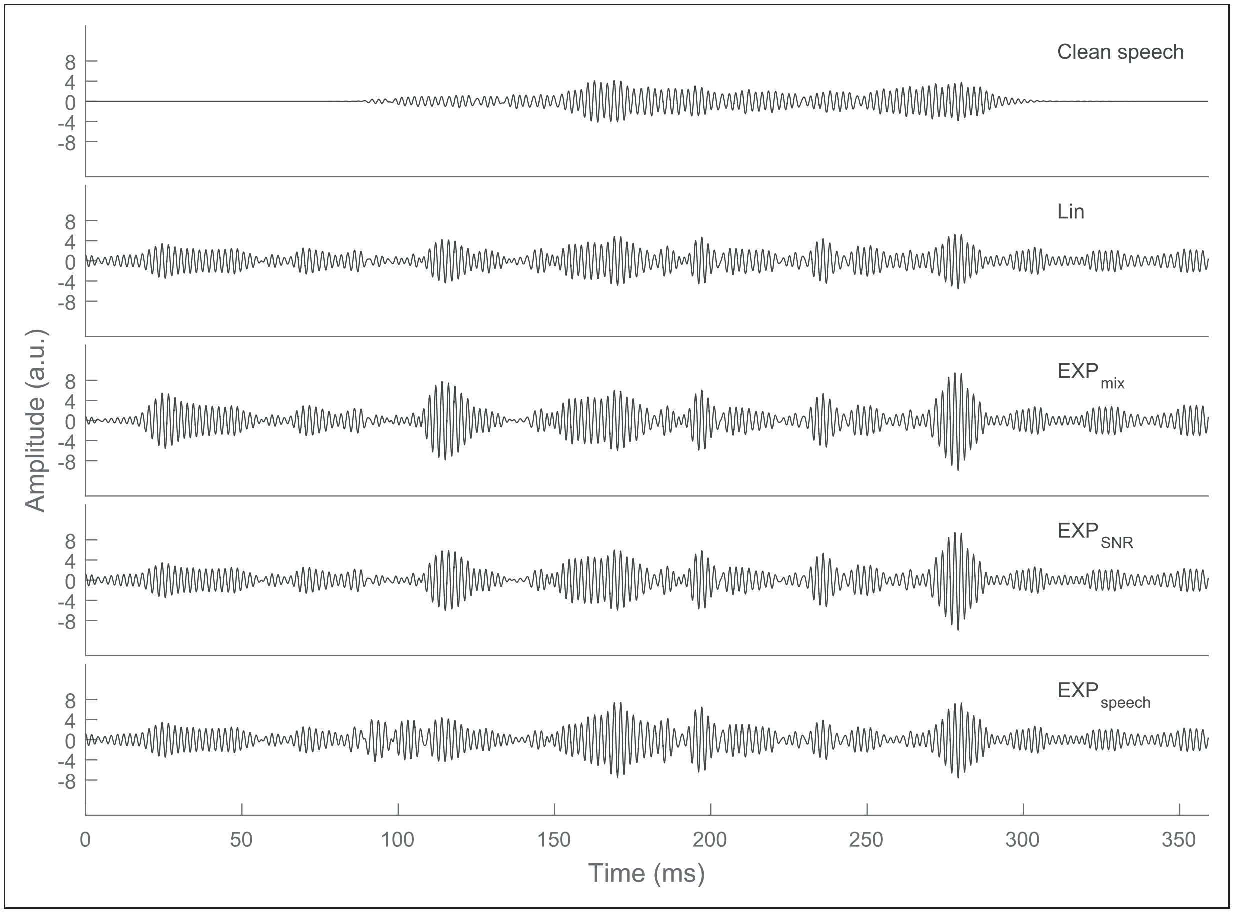

Figure 3. Waveforms of the male speech token /mi/ along with waveforms obtained with the same speech token mixed with noise at 0-dB SNR for linear processing and the three envelope expansion conditions. For illustration purposes, only the results at the output of an auditory-inspired gammatone filter tuned to 500 Hz are shown. From the top, the panels show the temporal output of the gammatone filter for the clean speech, linear processing, Exp$_{mix}$, Exp$_{SNR}$, and Exp$_{speech}$ conditions, respectively. The ordinate is the signal magnitude, expressed in arbitrary linear units. The abscissa is time, expressed in milliseconds.
SNR = signal-to-noise ratio.

modulations, Exp$_{SNR}$ enhances the modulations of the noisy-mixture portions with little noise contributions, and Exp$_{mix}$ "blindly" enhances the modulations in the entire noisy mixture, irrespective of whether a particular portion of the signal is dominated by noise or speech.

Experimental Design

A control condition with speech presented in quiet was defined as "Q65." This control condition was included to evaluate whether the CV tokens were sufficiently audible in quiet at the lowest speech level occurring in the SNR conditions. The clean speech (without envelope enhancement) was therefore presented at 65 dB SPL for the NH listeners and 65 dB SPL + NAL-R(P) amplification for the HI listeners, corresponding to the speech level in the 0-dB SNR condition. The experimental sessions were split into four consecutive blocks corresponding to the four signal-processing conditions (Lin, Exp$_{mix}$, Exp$_{speech}$, and Exp$_{SNR}$). For each of the listener groups, the order of presentation for the experimental blocks was counterbalanced using a Latin-square design to control for order effects. In order to get the listeners accustomed to the task, the "easy" control listening condition Q65 was presented first. Within each of the succeeding four experimental blocks, the three SNR conditions ranked from easy to difficult, that is, with SNR tested in the order 12, 6, and 0 dB. For each of the SNR conditions, the 30 CV tokens were presented in random order within each of five repetition blocks. This was done to facilitate the evaluation of potential learning effects.

Procedure and Apparatus

All signals were generated digitally in MATLAB (Version 2015b; The MathWorks, Inc., Natick, MA, United States) on a PC equipped with an RME UCX Fireface sound card at a sampling rate of 44.1 kHz and with a resolution of 16 bits per sample. The stimuli were presented in a sound-attenuating booth via Sennheiser HD 650 headphones to the better ear of the listeners, as derived from the average of the audiometric thresholds at 500 Hz, 1000 Hz, and 2000 Hz. The transfer function of each earpiece of the headphones was digitally equalized (101-point FIR filter) to produce a flat frequency response for frequencies between 0.100 and 10 kHz, measured with an ear simulator (B&K 4153) and a flat plate adaptor as specified in IEC 60318-1 (2009).

Statistical Analysis

An analysis of variance (ANOVA) was conducted on a mixed-effect model to evaluate whether hearing impairment, SNR, and processing condition had an effect on consonant recognition performance. In the mixed-effect model, listeners were nested within hearing status (NH vs. HI). Listeners and repetitions were treated as random block effects, while SNR, processing condition, and hearing status were treated as fixed effects. The random-listener effect accommodates the repeated-measures design by assuming that observations from the same listener are correlated. The assumptions underlying a parametric analysis were met without transforming the dependent variable. Tukey's Honestly Significant Difference corrected post hoc tests were conducted to test for main effects and interactions. A confidence level of 5% was considered to be statistically significant. The statistical analysis was performed using the lme4 and lsmeans packages in R (Bates, Mächler, Bolker, & Walker, 2015; Lenth, 2016).

Results

Consonant Recognition Scores of NH and HI Listeners

Figure 4 shows the consonant recognition scores obtained with the four different signal-processing conditions for the NH listeners (left panel) and the HI listeners (right panel) as a function of the SNR. The consonant recognition scores were calculated as the mean percentage correct across all consonants, talkers, repetitions, and listeners for both listener groups. For all SNRs and processing conditions, the consonant recognition scores were poorer for the HI than for the NH listeners. The scores generally increased with SNR and reached their maximum value for the quiet condition. The results showed that, for both listener groups, the two expansion conditions $\text{Exp}_{\text{speech}}$ (squares) and Exp_{SNR} (triangles) provided a small but consistent improvement relative to linear processing (asterisks) except for the $\text{Exp}_{\text{speech}}$ results for the NH listeners at 12 dB SNR where a slightly detrimental effect was found. In contrast, the condition Exp_{mix} (circles) provided a small improvement for the NH listeners but not for the HI listeners.

The outcomes of the ANOVA, summarized in Table 2, showed main effects of hearing impairment, SNR, and processing condition as well as an interaction between hearing impairment and SNR. The effects of the envelope expansion schemes were largely consistent

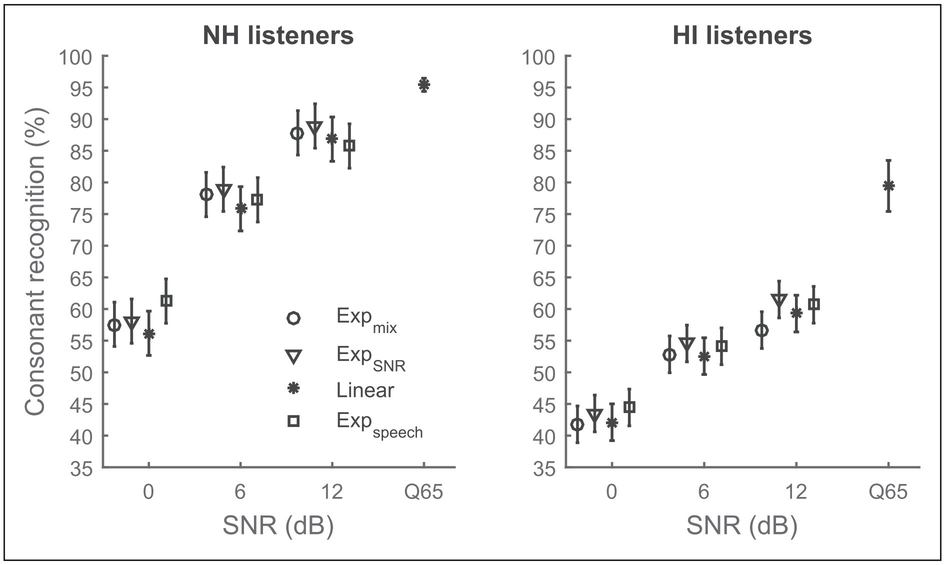

Figure 4. Overall consonant recognition scores for the NH listeners (left) and the HI listeners (right) as a function of the SNR for the four different signal-processing conditions. (Circles: $\text{Exp}_{\text{mix,}}$ triangles: Exp_{SNR}, asterisks: linear processing, squares: $\text{Exp}_{\text{speech}}$). The error bars represent ±1 standard errors of the mean. A slight horizontal jitter was added to the data for better readability.
HI = hearing impaired; NH = normal hearing.

Table 2. Summary of the ANOVA Outcomes for a Mixed-Effect Model Fitted to the Consonant Recognition Data With a Between-Listener Factor of Hearing Impairment, and Within-Listener Factors of SNR and Processing Condition.

	df	*F* ratio	Probability
Processing condition	(3, 1133)	7.68	<.001
SNR	(2, 36)	292.93	<.001
Hearing impairment	(1, 18)	26.04	<.001
Hearing Impairment × SNR	(2, 36)	22.38	<.001
Processing Condition × SNR	(6, 1127)	1.89	.08
Processing Condition × Hearing Impairment	(3, 1124)	2.08	.10
Hearing Impairment × Processing Condition × SNR	(6, 1118)	1.26	.27

ANOVA = analysis of variance; SNR = signal-to-noise ratio.

across NH and HI listeners. Post hoc comparisons confirmed that the consonant recognition performance was improved in the $\text{Exp}_{\text{speech}}$ condition (by 1.8 percentage points, $p = .008$) and the Exp_{SNR} condition (by 2.1 percentage points, $p = .001$), relative to the linear processing condition. The standard error was 0.5 percentage points in both cases. In contrast, the consonant recognition scores for the Exp_{mix} and linear processing conditions were not significantly different ($p = .99$). There were no significant differences between the consonant recognition scores for the $\text{Exp}_{\text{speech}}$ and Exp_{SNR} conditions ($p = .95$), but the scores in both of these conditions were significantly higher than in the Exp_{mix} condition ($p = .01$).

An alternative, more familiar, performance measure is the change in SNR corresponding to the improvement in recognition scores. The statistical analysis of the data (shown in Figure 4) indicates that the improvement in terms of percentage correct for the Exp_{SNR} and $\text{Exp}_{\text{speech}}$ conditions versus linear processing was roughly constant across the tested SNRs, as there was no interaction between processing condition and SNR. Psychometric functions fitted to the data points obtained with linear processing in Figure 4 revealed that the recognition-score improvement for the Exp_{SNR} and $\text{Exp}_{\text{speech}}$ conditions relative to linear processing was equivalent to a 1-dB change in SNR for the NH listeners. For the HI listeners, this improvement amounted to a 1.9-dB change in SNR. The difference in SNR improvement between the two listener groups, despite similar recognition-score improvements, was caused by differences in the slopes of the respective psychometric functions, which were shallower for the HI listeners than for the NH listeners.

Figure 5 compares the consonant recognition scores obtained in the linear reference condition to those obtained in the three expansion conditions. To evaluate how the individual expansion schemes affect different phonetic categories, the recognition scores were averaged within the categories /p,k,t/ (blue), /b,g,d/ (green), /f,s,ʃ,v/ (red), /n,m/ (black), and /h,j,l/ (cyan). The average recognition scores obtained with the three expansion schemes (left: Exp_{mix}, middle: Exp_{SNR}, right: $\text{Exp}_{\text{speech}}$) are shown as a function of the average recognition scores obtained with linear processing. The results for the NH listeners are shown in the top panels and the results for the HI listeners are shown in the bottom panels. None of the expansion schemes had a detrimental effect on the recognition scores in the NH listeners, as no points fall more than one percentage point below the diagonals in the top panels of Figure 5. As expected, the recognition scores were mainly increased for the stop consonants /p,k,t/ (blue). This improvement was largest for Exp_{SNR} (upper middle panel), slightly smaller for Exp_{mix} (upper right panel), and small for the "ideal" $\text{Exp}_{\text{speech}}$ (top right panel). In contrast to the NH listeners, the expansion schemes had a detrimental effect on the HI listeners for the consonant groups /n,m/ and /b,g,d/ (bottom panels of Figure 5). However, similar to the effects observed in the NH listeners, the recognition scores for /p,k,t/ (blue) were increased substantially in all expansion conditions. The effects of the expansion schemes varied strongly across the consonant groups. Interestingly, Exp_{mix} did not affect consonant recognition for the fricatives /f,s,ʃ,v/ (red dot, lower left panel), whereas Exp_{SNR} (red dot, lower middle panel) and $\text{Exp}_{\text{speech}}$ (red dot, lower right panel) provided a benefit of 5% and 4%, respectively. Overall, the SNR-based expansion Exp_{SNR} provided the largest benefits for /f,s,ʃ,v/ and the smallest detrimental effects (−2% for /n,m/ and −3% for /b,g,d/) in the HI listeners.

Individual Listener Analysis

The abovementioned analysis focused on group averages, showing moderate improvements of consonant recognition scores induced by the envelope expansion on a group level. However, the individual listeners may have experienced largely different benefits from the expansion processing. To analyze the individual differences in benefit, Figure 6 shows a scatter plot of the across-SNR average consonant recognition performance with linear processing on the abscissa and the across-SNR average performance with Exp_{SNR} on the ordinate. Each symbol in Figure 6 represents the result for an individual listener (circles: NH; triangles: HI). The scatter plot reveals that the improvement in the overall recognition performance for the Exp_{SNR} conditions was mainly driven by the six listeners (4 HI, 2 NH) for whom the expansion processing was most beneficial (on average 6.8% and 9.4%, respectively, for the Exp_{SNR} and $\text{Exp}_{\text{speech}}$ conditions). For the 14 other listeners, the expansion processing affected the consonant recognition performance by less

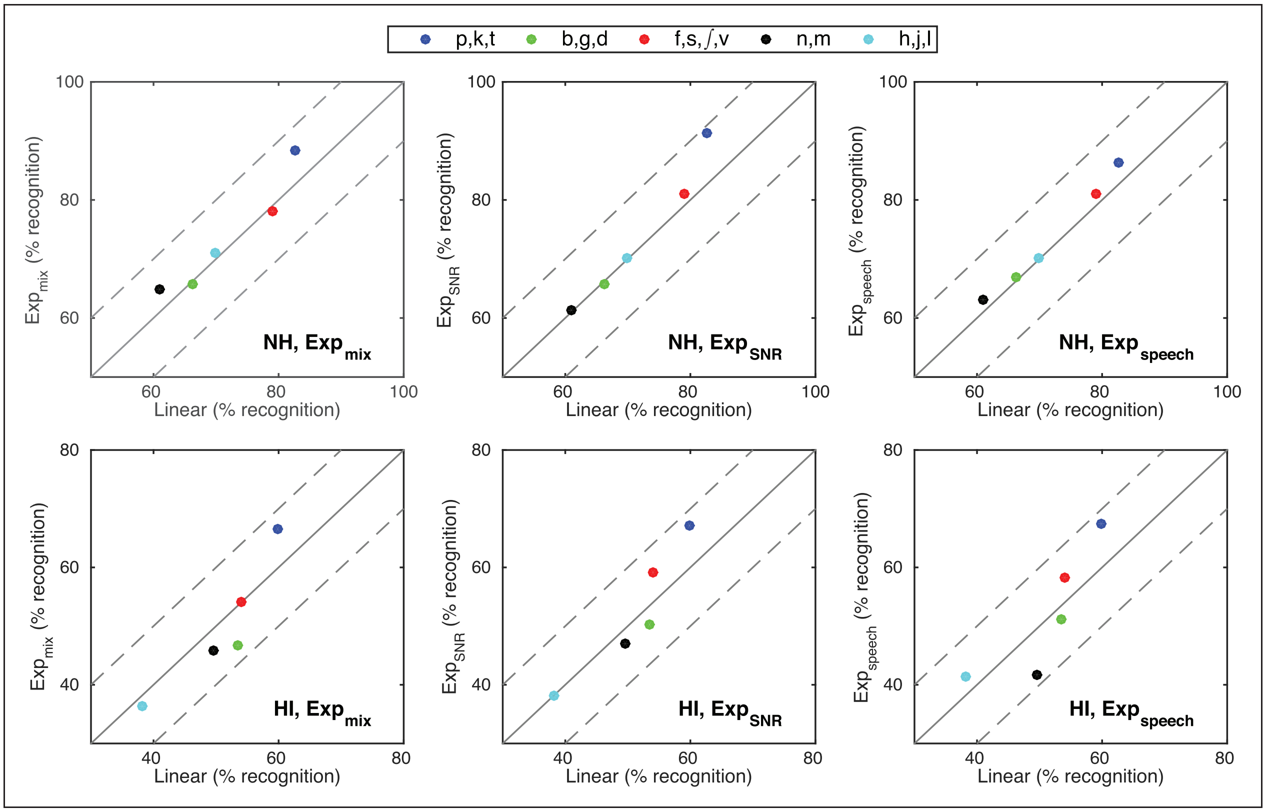

Figure 5. Scatter plot of consonant recognition in percentage measured with the linear condition versus the three expansion conditions (from left to right: Exp_{mix}, Exp_{SNR}, and $\text{Exp}_{\text{speech}}$) in NH (top) and HI (bottom) listeners. The consonant recognition scores were averaged within the phonetic categories shown in the legend. Within each panel, the solid gray line represents equal performance while the dashed lines represent ±10% differences induced by the respective expansion scheme.
HI = hearing impaired; NH = normal hearing.

than ±5 percentage points (0.1% and −1.0% on average for Exp_{SNR} and $\text{Exp}_{\text{speech}}$, respectively).

Discussion

Increasing the modulation depth of the speech envelope has been suggested to facilitate the extraction of speech cues and thereby improve speech intelligibility in noise. In this study, the effects of three different envelope expansion schemes that increase the depth of envelope fluctuations between 10 and 20 Hz were tested in a consonant identification task. Envelope expansion of the noisy speech showed no significant effect on the overall consonant recognition performance relative to linear processing, neither for the NH nor the HI listeners. This finding is consistent with results from earlier studies that investigated the effect of expanding the envelope of noisy speech (Apoux, Crouzet, & Lorenzi, 2001; Clarkson & Bahgat, 1991; Van Buuren et al., 1999). While the processing improved the intelligibility of some of the plosives most likely because of an enhancement of the detectability of the transient release bursts, this was accompanied by an increased proportion of consonant confusions for the other consonant categories. In contrast, SNR-based envelope expansion of the noisy speech, which confined the enhancement to the time-frequency segments in which the speech power was present, improved the overall consonant recognition performance both for the NH and the HI listeners. Interestingly, the effect of the SNR-based envelope expansion was found to be similar to the effect of envelope-expanding the clean speech before the addition of noise.

While the expansion benefit in terms of recognition-score improvement was substantial for some listeners (about 10 percentage points), the average effect for the entire population was relatively small (about two percentage points improvement of consonant recognition). Nevertheless, the observation of similar results obtained with the SNR-based processing and the clean-speech envelope expansion is promising, given that previous studies reported substantial improvements in speech perception with envelope expansion of clean speech (e.g., Apoux et al., 2004; Langhans & Strube, 1982). This suggests that the SNR-based envelope expansion scheme proposed in this study could provide larger improvements in speech perception when combined with

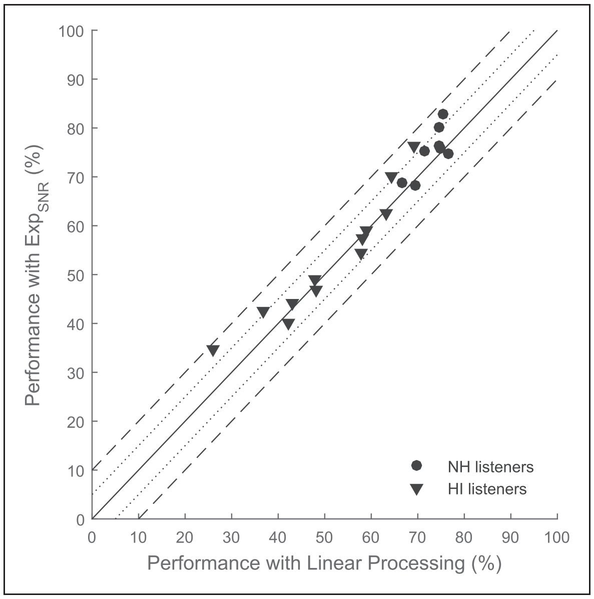

Figure 6. Scatter plot of consonant recognition performance with Exp$_{SNR}$ as a function of recognition performance with linear processing. Circles and triangles show results for NH and HI listeners, respectively. For visual clarity, the data were averaged across SNR conditions. The solid, dotted, and dashed lines represent equal performance, ±5 percentage-point improvements, and ±10 percentage-point improvements, respectively, obtained with the expansion processing relative to linear processing. HI = hearing impaired; NH = normal hearing; SNR = signal-to-noise ratio.

alternative parameter settings, such as those considered in the earlier studies. In contrast to the expansion of clean speech before the addition of noise, SNR-based envelope expansion is feasible in hearing-aid algorithms using blind SNR-estimation methods (e.g., Gerkmann & Hendriks, 2012; Martin, 2001) such that this approach might help improve speech perception in hearing-aid users.

When expressing the changes in consonant recognition performance as equivalent change in SNR, the net effect was an improvement (relative to the linear processing condition) that was about 1 dB larger for the HI listeners than for the NH listeners in the Exp$_{SNR}$ and Exp$_{speech}$ conditions. Thus, a larger increase in SNR is required for the HI listeners than for the NH listeners to obtain the same increase of the recognition score. The benefit achieved with the expansion processing may thus be larger for HI listeners than for NH listeners.

The relatively small (yet statistically significant) improvements in consonant recognition performance induced by the proposed envelope expansion processing indicate that the chosen parameter settings were suboptimal and should therefore be optimized to achieve benefits that justify a potential hearing-aid application. Consistent with the results from this study where modulation frequencies between 10 and 20 Hz were enhanced, envelope filtering studies have demonstrated that the contribution of envelope fluctuations above 12 Hz to phoneme intelligibility is small in quiet and in stationary background noise (Drullman, Festen, & Plomp, 1994; Xu et al., 2005; Xu & Zheng, 2007). However, ceiling effects were observed in those studies and it thus remained unclear whether this finding could be reproduced if the experiment was not confounded by such effects. In contrast, in terms of sentence intelligibility, the envelope expansion of (clean) speech has been shown to provide a greater benefit when applied to a wider range of modulation frequencies. For example, Apoux et al. (2004) showed that their expansion processing of modulation frequencies in the range 0 to 256 Hz was more effective than in the range 0 to 16 Hz. Therefore, it is possible that an additional enhancement of a wider range of modulation frequencies (below 10 Hz and above 20 Hz) would increase the benefit provided by the expansion processing. However, it should be taken into account that while expanding clean speech up to modulation frequencies in the range of the fundamental frequency (about 100–200 Hz) may yield more robust periodicity information, this approach might be detrimental when applied to a noisy signal where the SNR in the modulation domain typically decreases monotonically with increasing modulation frequency. In any case, this type of expansion would still require that modulation frequencies below the syllabic rate are not enhanced to avoid compromising audibility for the HI listeners. Furthermore, expansion of slow envelope fluctuations tends to decrease the consonant-vowel intensity ratio (CVR) as the processing enhances high-intensity vowels more than low-intensity consonants (Apoux et al., 2004) which, in turn, may affect consonant recognition performance (Freyman & Nerbonne, 1989). A possible solution may be to apply expansion processing at modulation frequencies between 4 and 256 Hz in combination with amplitude compression of the slow envelope fluctuations below 4 Hz, such that the CVR is increased as compared with the case where only expansion is applied.

The rationale for using stationary background noise rather than fluctuating background noise in this study was to maximize the benefit provided by the expansion processing in terms of consonant recognition. This expectation was based on the results from Apoux et al. (2004) who found larger benefits provided by their expansion processing in terms of word recognition scores in stationary noise than in fluctuating noise. Furthermore, supraprocessing deficits have been shown to provide stronger links to speech intelligibility in stationary noise than in fluctuating noise (e.g., Van Esch & Dreschler, 2015). Hence, the effect of the proposed expansion schemes may be smaller for fluctuating background noise maskers with a more similar modulation spectrum to the target speech. However, it should be noted that the bandpass filtering applied in the expansion algorithm corresponds to low-pass filtering in the modulation domain, such that the individual frequency bands of the noise considered for envelope expansion were in fact highly modulated. This make the distinction between stationary and fluctuating noise less prominent than in the case of a wideband envelope expansion scheme as used in the Apoux et al. (2004) study.

It has been demonstrated that listeners can learn to adapt to artificially produced, nonlinear changes of the natural auditory cues that are used for auditory perception. For example, frequency-lowering signal processing strategies have been implemented in hearing aids. Frequency lowering shifts acoustic cues from high-frequency regions to lower frequencies where audibility is typically better, thereby potentially improving the listener's access to the speech cues (for a review, see Simpson, 2009). Several studies have indicated that a period of acclimatization was necessary before frequency lowering provided benefits in speech recognition (Ellis & Munro, 2015; Glista, Scollie, & Sulkers, 2012; Wolfe et al., 2011). Thus, the benefit of nonlinear signal processing schemes, such as envelope expansion, may not be immediately apparent when assessed without a period of acclimatization.

The observed improvements in consonant recognition performance induced by the proposed envelope expansion schemes were found in a subgroup of the listeners, that is, only selected listeners benefited from this type of processing. Further research is needed to clarify why these differences in benefit occur and to establish to what extent they are related to intersubject variability caused by the experimental design and to what extent these differences can be accounted for by individual differences in psychoacoustic measures, such as temporal envelope detection and discrimination (e.g., Schlittenlacher & Moore, 2016; Wiinberg et al., 2018) or in terms of acclimatization to the processing.

Conclusion

This study investigated the effect of expanding envelope fluctuations between 10 and 20 Hz on consonant recognition performance in NH and HI listeners. Envelope expansion of noisy speech showed no significant effect on the overall consonant recognition performance relative to linear processing. In contrast, SNR-based envelope expansion of the noisy speech improved the overall consonant recognition performance by about two percentage points, mainly resulting from an improved recognition of some of the stop consonants. If the change in performance was expressed in terms of equivalent change in SNR, the net effect was an improvement (relative to the linear condition) of 1 dB and 1.9 dB for the NH and HI listeners, respectively. The effect of the SNR-based envelope expansion was comparable with the effect of ‘‘idealized’’ envelope expansion of the clean speech before the addition of noise. The size of the measured effects was relatively small compared with other related studies, indicating that extending the enhanced modulation-frequency range from 10–20 Hz to, for example, 4–20 Hz might yield larger benefits. Overall, the results support the hypothesis that the detrimental effect of enhancing the noise fluctuations in the different frequency bands on speech perception is effectively reduced by SNR-based envelope expansion. Furthermore, the results suggest that, because of its practical feasibility, the proposed SNR-based envelope expansion scheme may be interesting for speech-enhancement applications in hearing aids.

Acknowledgments

We would like to thank Morten Løve Jepsen and Christoph Scheidiger for helpful discussions. This project was carried at the Centre for Applied Hearing Research which is supported by Widex, Oticon, GN Hearing and the Technical University of Denmark.

Declaration of Conflicting Interests

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

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Alan Wiinberg http://orcid.org/0000-0001-5239-1486

References

Apoux, F., Crouzet, O., & Lorenzi, C. (2001). Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: Effects on identification performance and response times. *Hearing Research*, *153*(1–2), 123–131. doi: 10.1016/S0378-5955(00)00265-3

Apoux, F., Tribut, N., Debruille, X., & Lorenzi, C. (2004). Identification of envelope-expanded sentences in normal-hearing and hearing-impaired listeners. *Hearing Research*, *189*(1–2), 13–24. doi: 10.1016/S0378-5955(03)00397-6

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). Retrieved from https://cran.r-project.org/package=lmerTest. doi: 10.18637/jss.v067.i01

Byrne, D., Parkinson, A., & Newall, P. (1990). Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired. *Ear and Hearing*, *11*(1), 40–49. doi: 10.1097/00003446-199002000-00009

Chabot-Leclerc, A., Jørgensen, S., & Dau, T. (2014). The role of auditory spectro-temporal modulation filtering and the decision metric for speech intelligibility prediction. *The Journal of the Acoustical Society of America*, *135*(6), 3502–3512. doi: 10.1121/1.4873517

Chabot-Leclerc, A., MacDonald, E. N., & Dau, T. (2016). Predicting binaural speech intelligibility using the signal-to-noise ratio in the envelope power spectrum domain. *The Journal of the Acoustical Society of America*, *140*(1), 192–205. doi: 10.1121/1.4954254

Christiansen, T. U. (2011, June 26). *Objective evaluation of consonant–vowel pairs produced by native speakers of Danish*. Paper presented at the Proceedings of Forum Acusticum, Aalborg, Denmark.

Clarkson, P. M., & Bahgat, S. F. (1991). Envelope expansion methods for speech enhancement. *The Journal of the Acoustical Society of America*, *89*(3), 1378–1382. doi: 10.1121/1.400538

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, *95*(2), 1053–1064. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8132899

Ellis, R. J., & Munro, K. J. (2015). Benefit from, and acclimatization to, frequency compression hearing aids in experienced adult hearing-aid users. *International Journal of Audiology*, *54*(1), 37–47. doi: 10.3109/14992027.2014.948217

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736. doi: 10.1121/1.400247

Fowler, E. P. (1936). A method for the early detection of otosclerosis: A study of sounds well above threshold. *Archives of Otolaryngology*, *24*(6), 731–741.

Freyman, R. L., & Nerbonne, G. P. (1989). The importance of consonant–vowel intensity ratio in the intelligibility of voiceless consonants. *Journal of Speech Language and Hearing Research*, *32*(3), 524–535. doi: 10.1044/jshr.3203.524

Freyman, R. L., & Nerbonne, G. P. (1996). Consonant confusions in amplitude-expanded speech. *Journal of Speech Language and Hearing Research*, *39*(6), 1124–1137. doi: 10.1044/jshr.3906.1124

Gerkmann, T., & Hendriks, R. C. (2012). Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. *IEEE Transactions on Audio, Speech and Language Processing*, *20*(4), 1383–1393. doi: 10.1109/TASL.2011.2180896

Glista, D., Scollie, S., & Sulkers, J. (2012). Perceptual acclimatization post nonlinear frequency compression hearing aid fitting in older children. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1765–1787. doi: 10.1044/1092-4388

International Electrotechnical Commission. (2009). Electroacoustics - Simulators of human head and ear - Part 1: Ear simulator for the measurement of supra-aural and circumaural earphones. *IEC 60318-1-2009*, Geneva, Switzerland: IEC.

Jørgensen, S., & Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, *130*(3), 1475–1487. doi: 10.1121/1.3621502

Jørgensen, S., Decorsière, R., & Dau, T. (2015). Effects of manipulating the signal-to-noise envelope power ratio on speech intelligibility. *The Journal of the Acoustical Society of America*, *137*(3), 1401–1410. doi: 10.1121/1.4908240

Jørgensen, S., Ewert, S. D., & Dau, T. (2013). A multi-resolution envelope-power based model for speech intelligibility. *The Journal of the Acoustical Society of America*, *134*(1), 436–446. doi: 10.1121/1.4807563

Langhans, T., & Strube, H. (1982). Speech enhancement by nonlinear multiband envelope filtering. In C. Gueguen (Ed.), *ICASSP '82. IEEE international conference on acoustics, speech, and signal processing* (vol 7, pp. 156–159). New York, NY: Institute of Electrical and Electronics Engineers. doi: 10.1109/ICASSP.1982.1171715

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–43. doi: 10.18637/jss.v069.i01

Li, F., Menon, A., & Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *The Journal of the Acoustical Society of America*, *127*(4), 2599–2610. doi: 10.1121/1.3295689

Lobdell, B. E., & Allen, J. B. (2007). A model of the VU (volume-unit) meter, with speech applications. *The Journal of the Acoustical Society of America*, *121*(1), 279–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17297783

Martin, R. (2001). Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Signal Processing Letters*, *9*(5), 504–512.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America*, *63*(2), 533–549. doi: 10.1121/1.381753

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America*, *83*(6), 2322–2327.

Régnier, M. S., & Allen, J. B. (2008). A method to identify noise-robust perceptual features: application for consonant /t/. *Journal of the Acoustical Society of America*, *123*(5), 2801–2814. doi: 10.1121/1.2897915

Schlittenlacher, J., & Moore, B. C. J. (2016). Discrimination of amplitude-modulation depth by subjects with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *140*(5), 3487–3495. doi: 10.1121/1.4966117

Shannon, R., Zeng, F., & Kamath, V. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Simpson, A. (2009). Frequency-lowering devices for managing high-frequency hearing loss: A review. *Trends in Amplification*, *13*(2), 87–106. doi: 10.1177/1084713809336421

Steinberg, J., & Gardner, M. (1937). The dependence of hearing impairment on sound intensity. *The Journal of the Acoustical Society of America*, *9*, 11–23. doi: doi: /10.1121/1.1915905

Stone, M. A., Anton, K., & Moore, B. C. J. (2012). Use of high-rate envelope speech cues and their perceptually relevant dynamic range for the hearing impaired. *The Journal of the Acoustical Society of America*, *132*(2), 1141–1151.

Stone, M. A., Füllgrabe, C., & Moore, B. C. J. (2008). Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region. *The Journal of the Acoustical Society of America*, *124*(4), 2272–2282. doi: 10.1121/1.2968678

van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality. *The Journal of the Acoustical Society of America*, *105*(5), 2903–2913. doi: 10.1121/1.426943

Van Esch, T. E. M., & Dreschler, W. A. (2015). Relations between the intelligibility of speech in noise and psychophysical measures of hearing measured in four languages using the auditory profile test battery. *Trends in Hearing*, *19*(0), 1–12. doi: 10.1177/2331216515618902

Wiinberg, A., Jepsen, M. L., Epp, B., & Dau, T. (2018). Effects of Hearing Loss and Fast-Acting Compression on Amplitude Modulation Perception and Speech Intelligibility. *Ear and hearing*. Advance online publication. doi: 10.1097/AUD.0000000000000589

Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., Caraway, T., & Hudson, M. (2011). Long-term effects of non-linear frequency compression for children with moderate hearing loss. *International Journal of Audiology*, *50*(6), 396–404. doi: 10.3109/14992027.2010.551788

Xu, L., Thompson, C. S., & Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America*, *117*(5), 3255–3267. doi: 10.1121/1.1886405

Xu, L., & Zheng, Y. (2007). Spectral and temporal cues for phoneme recognition in noise. *The Journal of the Acoustical Society of America*, *122*(3), 1758–1764. doi: 10.1121/1.2767000

Zaar, J., & Dau, T. (2015). Sources of variability in consonant perception of normal-hearing listeners. *The Journal of the Acoustical Society of America*, *138*(3), 1253–1267. doi: 10.1121/1.4928142