Original Article

Trends in Hearing
Volume 22: 1–12

DOI: 10.1177/2331216518790903
journals.sagepub.com/home/tia
SAGE

Signal-to-Noise-Ratio-Aware Dynamic Range Compression in Hearing Aids

Tobias May[1], Borys Kowalewski[1], and Torsten Dau[1]

Abstract

Fast-acting dynamic range compression is a level-dependent amplification scheme which aims to restore audibility for hearing-impaired listeners. However, when being applied to noisy speech at positive signal-to-noise ratios (SNRs), the gain function typically changes rapidly over time as it is driven by the short-term fluctuations of the speech signal. This leads to an amplification of the noise components in the speech gaps, which reduces the output SNR and distorts the acoustic properties of the background noise. An adaptive compression scheme is proposed here which utilizes information about the SNR in different frequency channels to adaptively change the characteristics of the compressor. Specifically, fast-acting compression is applied to speech-dominated time-frequency (T-F) units where the SNR is high, while slow-acting compression is used to effectively linearize the processing for noise-dominated T-F units where the SNR is low. A systematic evaluation of this SNR-aware compression scheme showed that the effective compression of speech components embedded in noise was similar to that of a conventional fast-acting system, whereas natural fluctuations in the background noise were preserved in a similar way as when a slow-acting compressor was applied.

Keywords

wide dynamic range compression, signal-to-noise ratio, hearing-aid signal processing

Date received: 13 March 2018; revised: 27 June 2018; accepted: 28 June 2018

Introduction

One of the primary tasks of a hearing aid is to improve speech recognition through restored audibility (e.g., Jenstad & Souza, 2007; Souza, Boike, Witherell, & Tremblay, 2007; Souza & Turner, 1999). Wide dynamic range compression (WDRC) provides level-dependent amplification. It is therefore capable of improving the audibility of soft speech components while avoiding excessive amplification of high-intensity inputs and the loudness discomfort that would result from it otherwise (e.g., Alexander & Rallapalli, 2017; Villchur, 1973). WDRC is characterized by a number of parameters, such as the attack and release times, compression ratio (CR), compression threshold (CT), and the number of frequency channels. The attack time is usually very short (below 10 ms) such that the compressor can react to a rapid increase in the intensity of the input signal (Alexander & Rallapalli, 2017; Jenstad & Souza, 2005). A compressor is typically classified as fast-acting, with release times shorter than 200 ms, or slow-acting, with release times longer than 200 ms (for a review, see Souza, 2002).

For a maximum audibility benefit, the compression system must be able to follow changes in the speech amplitude on timescales corresponding to the duration of a syllable or even a phoneme. This requires a *very-fast-acting* system with a release time below about 60 ms (Edwards, 2004). If a longer release time is used, the gain might lag behind the dynamic changes in the speech envelope, leaving low-intensity components underamplified (Jerlvall & Lindblad, 1978; Kuk, 1996). As demonstrated by Braida et al. (1982) and Stone and Moore (1992), the effective compression ratios (ECRs) decrease to only a fraction of the *nominal* ratios when the release time is too long compared with the rate of the envelope fluctuations in the signal.

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

Corresponding author:
Tobias May, Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.
Email: tobmay@elektro.dtu.dk

Several studies have demonstrated a benefit of fast-acting compression for speech recognition in quiet (Souza & Turner, 1998, 1999; Villchur, 1973). In contrast, Davies-Venn, Souza, Brennan, & Stecker, (2009) found that when audibility was adjusted with linear versus level-dependent amplification using WDRC, the latter was found to be detrimental for speech recognition. This was probably caused by altered level differences between phonemes, distortions of the temporal envelope, or a reduction of the modulation depth of the speech signal (Alexander & Rallapalli, 2017; Gallun & Souza, 2008; Jenstad & Souza, 2005, 2007; Plomp, 1988; Rosen, 1992; Souza & Gallun, 2010; Souza & Turner, 1996; Stone & Moore, 2003, 2004, 2007, 2008; van Buuren, Festen, & Houtgast, 1999; Walaszek, 2008). Such distortions are typically more pronounced for shorter release times and higher CRs (Alexander & Rallapalli, 2017; Jenstad & Souza, 2005, 2007).

The relative benefit of WDRC versus linear amplification depends on the acoustic condition. When noise is present, the amount of the *effective* compression and the distortions of the speech envelope seem to be less pronounced compared with the processing of speech in quiet (Rhebergen, Versfeld, & Dreschler, 2009; Souza, Jenstad, & Boike, 2006). Yund and Buckles (1995) studied the impact of multichannel compression on speech recognition in the presence of a fixed-level stationary background noise and found an increased benefit as the signal-to-noise ratio (SNR) decreased. Moreover, Gatehouse, Naylor, and Elberling (2003, 2006) suggested that if the noise is fluctuating with distinct temporal dips, fast-acting compression would provide *differential amplification* by applying more gain to the low-intensity glimpses of the speech than to the noise peaks, potentially leading to improved intelligibility. This prediction is consistent with recent results from Rhebergen, Maalderink, and Dreschler (2017) and Desloge, Reed, Braida, Perez, and D'Aquila (2017), who established a link between increased speech audibility and improved speech intelligibility when applying fast-acting compression to speech in the presence of fluctuating background noise. On the contrary, compression can negatively affect the output SNR by reducing the speech level and overamplifying portions of the noise occurring in the speech gaps (Alexander & Masterson, 2014; Hagerman & Olofsson, 2004; Naylor & Johannesson, 2009; Rhebergen et al., 2017; Souza et al., 2006). As recently shown by Rhebergen et al. (2017), the reduction of the output SNR can be detrimental to speech recognition. Apart from a reduced output SNR, fast-acting compression of mixed sources (e.g., competing talkers or speech in noise) introduces across-signal modulations. Stone and Moore (2007, 2008) demonstrated that this distortion might be detrimental to speech intelligibility, at least when primarily envelope cues are available. Even if the effect on recognition can be small, other perceptual attributes might be affected, such as the perceived noisiness of the sound (e.g., Kuk, 1996; Neuman, Bakke, Mackersie, Hellman, & Levitt, 1998), leading to a perception of reduced overall quality. Therefore, it has been suggested that the compression parameters should be adjusted according to the environment (Kates, 2010; Yund, Simon, & Efron, 1987) to reach the *balance point*, at which the positive and negative acoustic effects optimally offset each other (Souza, Hoover, & Gallun, 2012).

The hypothesis of the current study was that an *optimal* hearing-aid compensation strategy should (a) amplify low-level portions of speech, (b) reduce the dynamic range of speech to avoid excessive loudness, (c) avoid amplifying the noise in speech gaps (so-called *pumping*), and (d) maintain the natural fluctuations in the background noise. To achieve this, an adaptive amplification scheme would be required that selectively changes the characteristics of the compressor in a given time-frequency (T-F) unit depending on whether speech or noise components are dominating. In earlier approaches, such as the K-amp strategy (Killion, Teder, Johnson, & Hanke, 1992) and the dual front-end automatic gain control system (Moore & Glasberg, 1988; Stone, Moore, Alcántara, & Glasberg, 1999), the release time varied according to how long the compression circuit had been activated, which can help to reduce the pumping artifacts. Similar principles have been applied in guided level estimators (Neumann, 2008; Simonsen & Behrens, 2009). Moreover, Lai, Li, Tsai, Chu, and Young (2013) proposed an adaptive WDRC system that adjusted the CR in individual frequency channels depending on the estimated short-term dynamic range. These systems, however, are only sensitive to changes in the overall signal level but do not utilize information related to the presence of the target signal versus the background noise. In the context of binaural WDRC, an adaptive amplification scheme was proposed by Hassager, May, Wiinberg, and Dau (2017), where knowledge about the acoustic scene in terms of the direct-to-reverberant energy ratio (DRR) was utilized to selectively apply fast-acting compression only to T-F units that are dominated by the direct sound. This direct sound–driven compression scheme, in conjunction with a binaural link, was demonstrated to improve sound source localization and externalization compared with conventional fast-acting compression (Hassager et al., 2017).

In this study, the idea of such a *scene-aware* amplification scheme was studied for acoustic scenes where speech and background noise were presented simultaneously. Specifically, an adaptive amplification system was considered that applied fast-acting compression only to speech-dominated T-F units, while the processing

of noise-dominated T-F units was linearized through a longer release time. The resulting amplification scheme, termed SNR-aware dynamic range compression, was compared with conventional fast- and slow-acting compression systems using three objective metrics based on the ECR as well as relative changes in the modulation spectrum and the broadband SNR.

System

The block diagram of the SNR-aware dynamic range compression algorithm is shown in Figure 1. First, the input signal was analyzed by a short-time discrete Fourier transform (STFT). In the acoustic scene analysis stage, a binary decision about speech activity was obtained by applying a threshold criterion to the estimated short-term SNRs in individual frequency channels. This decision was then utilized in the dynamic range compression stage to adaptively adjust the release time of the compressor. Specifically, a short release time was selected if a particular T-F unit was dominated by speech (high SNR), whereas a long release time was used for noise-dominated T-F units (low SNR). Then, a gain function was calculated and applied to the STFT representation of the noisy speech signal. Finally, the output signal was reconstructed using the STFT synthesis stage. All of the individual building blocks are described in detail in the following subsections.

STFT Analysis

The input signal was sampled at a rate of 20 kHz and segmented into overlapping frames of 10 ms duration with a shift of 2.5 ms. Each frame was Hann-windowed and zero-padded to a length of 512 samples and a 512-point discrete Fourier transform (DFT) was computed, producing an STFT representation of the input signal (Allen, 1977).

Speech Detection

Based on the STFT representation of noisy speech, a binary decision about speech activity was performed for each individual T-F unit. Therefore, the speech power spectral density (PSD) was first obtained in individual DFT bins using the minimum mean-square error estimator by Erkelens, Hendriks, Heusdens, and Jensen (2007). This method relies on an estimate of the noise PSD, which was derived from noisy speech using the algorithm proposed by Hendriks, Heusdens, and Jensen (2010). Both the noisy speech power and the estimated speech PSD were then integrated into seven octave–wide bands, by applying the filterbank described below, and subsequently used to estimate the short-term SNR (Eaton, Brookes, & Naylor, 2013; May, Kowalewski, Fereczkowski, & MacDonald, 2017). Finally, speech activity was detected by applying a threshold to the estimated SNRs[1] in individual T-F units. These thresholds were determined by a training procedure described in the *Parameters* subsection.

Filterbank

The dynamic range compressor operated separately in seven octave–wide bands with center frequencies ranging from 125 Hz to 8 kHz. The octave bands were designed to have rectangular filter weights that were applied to each DFT bin. Given the DFT resolution, the *effective* filter shape of the individual octave bands was as rectangular as possible. For each octave band, the power of the respective DFT bins was integrated and the magnitude of individual T-F units was returned.

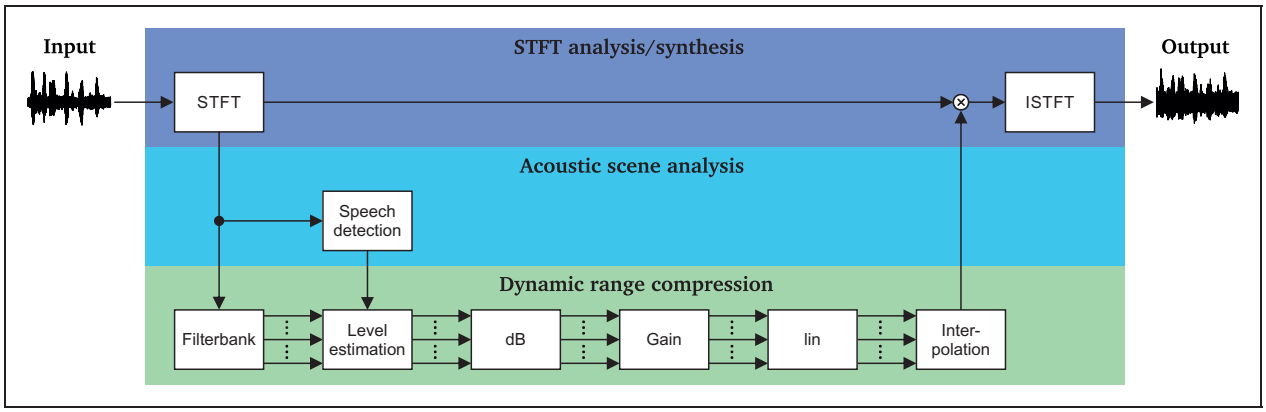

Figure 1. Block diagram of the SNR-aware compressor consisting of three processing layers: (a) STFT-based analysis and synthesis, (b) acoustic scene analysis, and (c) dynamic range compression. See System section for more details regarding the individual processing steps. ISTFT = inverse short-time discrete Fourier transform; SNR = signal-to-noise ratio; STFT = short-time discrete Fourier transform.

Level Estimation

The magnitude of the individual T-F units was smoothed by a first-order infinite impulse response filter with different time constants associated with attack and release. Given the binary decision about speech activity, two different sets of attack and release time constants were defined for speech-dominated and noise-dominated T-F units: (a) a short attack time of 5 ms and a short release time of 40 ms were used for the speech-dominated T-F units with a high SNR, and (b) a short attack time of 5 ms and a long release time of 2,000 ms were used for the noise-dominated T-F units where the SNR was low. In both cases, a short attack time was chosen to maintain the responsiveness of the compressor to rapid intensity changes, irrespective of whether the dominant signal was speech or noise.

Gain Calculation

Given the smoothed level estimation in decibels (dB), a broken-stick gain function was used to derive the respective gains in the individual T-F units. The broken-stick gain function provided a linear gain below the CT and a constant CR above the CT. This gain function was based on the NAL-NL2 prescription (Keidser, Dillon, Flax, Ching, & Brewer, 2011) fitted to the N_4 standard audiogram corresponding to a flat and moderately sloping hearing loss (Bisgaard, Vlaming, & Dahlquist, 2010) using the settings *slow* and *unilateral*. The CTs were derived by measuring the output level of the individual frequency channels in response to stationary speech-shaped noise. The speech-shaped noise had the same long-term average spectrum (LTAS) as the Danish hearing in noise test (HINT) speech material and was normalized to a root mean square–level of 50 dB. The resulting CRs and CTs for the seven octave bands are summarized in Table 1.

Interpolation of Gain Values

The linear gains were interpolated from the channel center frequencies to the DFT frequency axis using a piecewise cubic interpolation to avoid aliasing artifacts. These interpolated gains were subsequently applied to the STFT representation of noisy speech.

Table 1. CTs in Decibels and CRs for Individual Channel Center Frequencies.

	Channel center frequency (Hz)						
	125	250	500	1000	2000	4000	8000
CT (dB)	43	43	41	41	37	31	28
CR	2.2:1	2.2:1	2.2:1	3.0:1	3.5:1	3.3:1	2.5:1

Note. CT = compression threshold; CR = compression ratio.

STFT Synthesis

After multiplying the gains with the STFT representation of noisy speech, the processed time domain signal was reconstructed by applying an inverse short-time discrete Fourier transform (ISTFT). Specifically, an inverse discrete Fourier transform produced individual time segments that were combined by a weighted overlap-add method (Crochiere, 1980). The weighted overlap-add approach extends the original overlap-add method proposed by Allen (1977) with a synthesis window. A 512-sample tapered cosine window with 39-sample ramps was used as a synthesis window (Grimm, Herzke, Berg, & Hohmann, 2006) to smooth discontinuities at the frame boundaries, which can occur because of temporal aliasing.

Evaluation

Stimuli

Noisy speech was created by mixing clean speech from the Danish HINT (Nielsen & Dau, 2011) with four different types of background noise at seven SNRs (–6, –3, 0, 3, 6, 9, and 12 dB). The following noise types were used: Stationary International Collegium of Rehabilitative Audiology (ICRA)-1 noise and nonstationary ICRA-7 noise representing a six-talker speech babble (Dreschler, Verschuure, Ludvigsen, & Westermann, 2001) as well as car noise and factory noise from the NOISEX database (Varga & Steeneken, 1993). The noise signals were split into two halves of equal size to ensure that there was no overlap between the noise segments used for training the speech detection stage (see *Parameters* subsection) and evaluation. Following Naylor and Johannesson (2009), the LTAS of all noise types measured in 1/3 octave bands was adjusted to match the LTAS of the Danish HINT corpus.

Each noisy speech mixture consisted of 10 randomly selected HINT sentences from the test lists that were concatenated and mixed with a random noise segment. The noise was normalized to a root mean square–level corresponding to 50 dB while the level of the speech signal was adjusted to yield a predefined SNR. An initial noise-only segment of 250-ms duration was incorporated to ensure that the noise PSD estimator (see *Speech Detection* subsection) was properly initialized. After processing, this noise-only segment was removed and did not bias the analysis of the objective metrics. For each of the four noise types and seven SNRs, 20 noisy speech mixtures with an average length of 15.5 s were created,

resulting in a set of $4 \times 7 \times 20 = 560$ noisy speech mixtures used for evaluation.

Parameters

The binary decision of speech activity was obtained by thresholding the estimated SNRs in individual T-F units (see *Speech Detection* subsection). These thresholds were found by maximizing the hit rate minus false alarm rate (H − FA) between the estimated and the *true* speech activity using a small training set. For this purpose, 10 randomly selected HINT sentences from the training lists were mixed with ICRA-1 and IRCA-7 noise at −5, 0, and 5 dB SNR, producing a training set of $10 \times 2 \times 3 = 60$ noisy speech mixtures. The true speech activity was obtained by applying a threshold criterion of 0 dB to the *a priori* SNR, which was calculated from the individual speech and noise signals.

The noise PSD estimator by Hendriks et al. (2010) was used with the default parameter set and initialized for each noisy speech mixture by averaging the PSD across the initial noise-only segment of 250 ms. The speech PSD estimator from Erkelens et al. (2007) was configured with the two generalized gamma parameters $\gamma = 1$ and $\nu = 0.6$. Moreover, the smoothing factor α employed by the decision-directed approach corresponded to a time constant of 0.792 s.

Objective Metrics

Shadow-filtering (Fredelake, Holube, Schlueter, & Hansen, 2012; Gustafsson, Martin, & Vary, 1996) was employed to investigate the impact of compression on speech, noise, and noisy speech separately. The compressor gain was always estimated based on the noisy speech mixture and then subsequently applied to speech alone, noise alone, and noisy speech (in the STFT domain). The following three objective metrics were computed for a range of input SNRs:

- The ECR was calculated based on the estimated dynamic range before and after compression (Souza et al., 2006). The dynamic range was derived by calculating the level difference between the 99th and the 50th percentile in the different frequency channels.
- The relative change in the modulation spectrum (ΔMS) was computed before and after processing. The modulation spectrum reveals perceptual distortions introduced by compression (Alexander & Rallapalli, 2017; Gallun & Souza, 2008; Souza & Gallun, 2010). The modulation spectrum was computed based on the broadband envelope which was extracted by half-wave rectification and low-pass filtering with a cut-off frequency of 100 Hz. Subsequently, the power in seven octave–spaced modulation filters (0.5, 1, 2, 4, 8, 16, and 32 Hz) was calculated and normalized by the direct current component of the envelope.
- The input/output SNR was computed based on the broadband signals before and after processing (Naylor & Johannesson, 2009; Rhebergen et al., 2017; Souza et al., 2006).

Compression Systems

The following four compression systems were evaluated which all operated in seven octave bands: fast-acting, slow-acting, SNR-aware compression as well as ideal SNR-aware compression based on the *a priori* SNR. An overview of the respective parameters is given in Table 2. While the conventional fast- and slow-acting compression systems were characterized by the attack and release times, the SNR-aware approach adaptively switched between two sets of attack and release times for speech- and noise-dominated T-F units. The ideal SNR-aware compression system used the true speech activity based on the *a priori* SNR (see *Parameters* subsection), rather than the speech activity estimator described in the *Speech Detection* subsection.

The processing principle of the four different compression schemes is illustrated in Figure 2 for a speech signal mixed with ICRA-1 noise at 6 dB SNR. Given the noisy speech signal, the respective gain functions are shown for a channel center frequency of 2 kHz. The fast-acting system is able to follow rapid intensity changes of the noisy speech signal, while inherent fluctuations in the noise-only segments also result in fast changes in the gain function. In contrast, the slow-acting system only responds to strong onsets and only slowly recovers following the offset of the dominant signal (speech, in this case). Because of the prolonged recovery, the gain remains relatively low after higher intensity segments, leaving other low-level speech components underamplified. The SNR-aware system adaptively switches between fast and slow processing depending on the estimated speech activity. Thus, in speech-active time segments,

Table 2. Configuration of the Four Tested Compression Schemes.

Compressor	Attack (ms)	Release (ms)	Speech detection	Estimator
Fast	5	40	Off	–
Slow	5	2,000	Off	–
SNR-aware	5/5	40/2,000	On	Estimated SNR
SNR-aware ideal	5/5	40/2,000	On	*a priori* SNR

Note. SNR = signal-to-noise ratio.

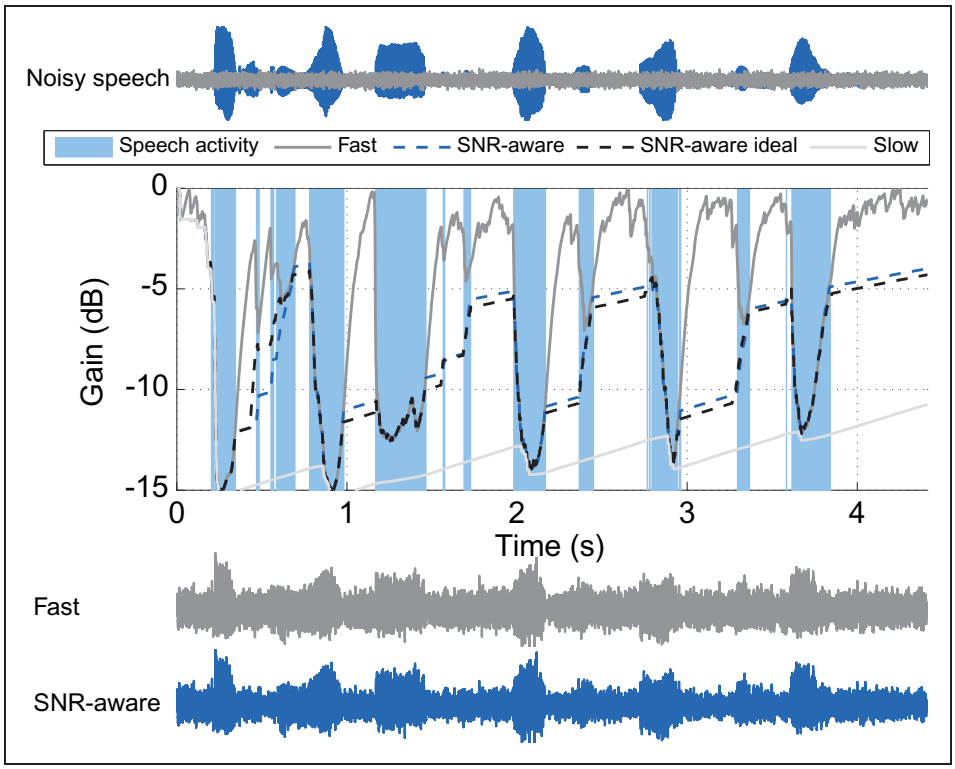

Figure 2. Speech mixed with ICRA-1 noise at 6 dB SNR (top panel) along with the estimated speech activity and gain functions of four compression systems (fast-acting, slow-acting, SNR-aware, and ideal SNR-aware compression) for a frequency channel centered at 2 kHz. The lowest two panels show the output of the fast-acting and the SNR-aware compressor, respectively. ICRA = International Collegium of Rehabilitative Audiology; SNR = signal-to-noise ratio.

the SNR-aware system is able to follow rapid intensity changes caused by the short release time, while the use of a long release time for noise-dominant time segments effectively linearizes the processing, which avoids rapid fluctuations in the gain in response to noise-only segments.

Results

The ECRs are shown in Figure 3 as a function of the input SNR and the channel center frequency. Each of the four rows represents a different compression scheme, that is, fast-acting (first row), slow-acting (second row), SNR-aware (third row), and ideal SNR-aware compression (fourth row). The left, middle, and right columns show results for the three different signal categories, that is, shadow-filtered speech, shadow-filtered noise, and noisy speech.

As expected, the fast-acting compression system provided the highest ECRs for all three signal categories. For noisy speech (right column), a maximum ECR of up to 2.0 was measured for high frequencies. When using shadow-filtering to analyze the impact of compression on speech and noise separately (left and middle columns), it can be seen that both speech and noise components were compressed, with ECRs of up to 1.6 and 1.3, respectively. The slow-acting compression system did not compress the noise components (with ECRs of 1 and lower) and also provided no compression to the speech components, where the ECR was 1.1 for the entire range of input SNRs. The ECRs of the SNR-aware compressor for the speech components were in a similar range (up to 1.4) as for the fast-acting compressor, while the ECR associated with the noise components was close to 1 (± 0.1) for a wide range of input SNRs. Finally, the ECR contours of the SNR-aware and the ideal SNR-aware compressor were very similar to each other for all three signal categories.

Figure 4 shows the relative change in the modulation spectrum (ΔMS) as a function of modulation frequency (ranging from 0.5 to 32 Hz) and the input SNR. Negative values indicate a reduction in modulation depth, while positive values reflect an increase in modulation depth caused by the level-dependent amplification (compression). Again, the four rows represent the different compression schemes (fast-acting, slow-acting, SNR-aware, and ideal SNR-aware compression) and the three columns show results for shadow-filtered speech, shadow-filtered noise, and the noisy speech mixture, respectively.

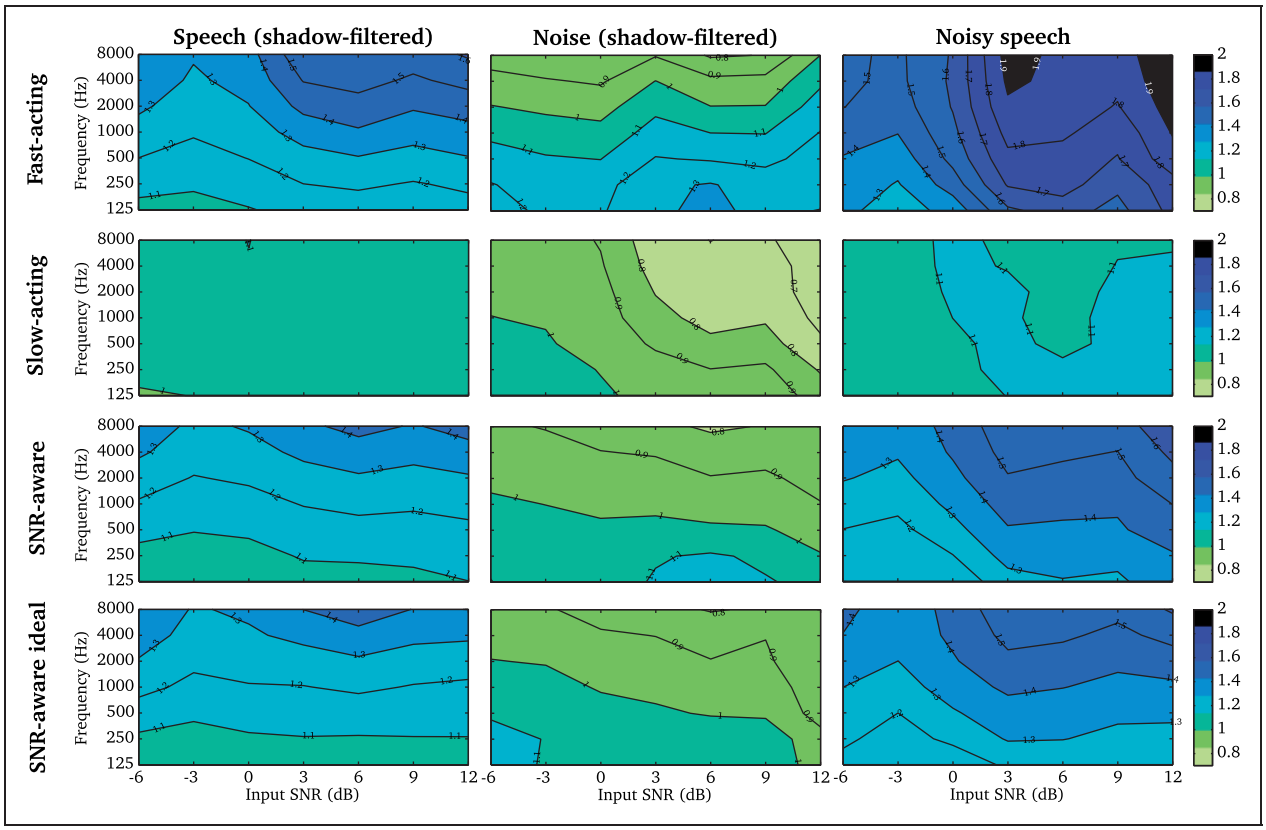

Figure 3. Contours of ECRs for the fast-acting (first row), slow-acting (second row), SNR-aware (third row), and ideal SNR-aware compressor (fourth row) as a function of the input SNR and the channel center frequency. Results were averaged across all four noise types. The left, middle, and right columns show results for shadow-filtered speech, shadow-filtered noise, and noisy speech, respectively. ECR = effective compression ratio; SNR = signal-to-noise ratio.

Fast-acting compression reduced the modulation depth of the shadow-filtered speech signal for modulation frequencies between 0.5 and 8 Hz and this effect increased with increasing SNR. At the same time, the modulation depth of the shadow-filtered noise signal was enhanced with a clear peak around 4 Hz for higher input SNRs. Slow compression did not markedly affect the modulation spectra of the shadow-filtered speech and noise signals. While ΔMS was positive in the range between 0.5 and 8 Hz for the shadow-filtered noise, the individual functions obtained for the different SNRs were fairly flat and did not show any pronounced peak. This coincided with a decreased ECR as already observed in Figure 3. Both SNR-aware systems resembled the conventional fast-acting compressor in terms of ΔMS for shadow-filtered speech. Although modulations around 4 Hz were to some extent enhanced in the shadow-filtered noise, the individual functions were much flatter compared with the fast-acting system and the respective magnitudes were closer to those obtained with the slow-acting compression system.

Finally, the input/output SNR analysis for the four compression schemes and a linear reference condition (dashed line) is shown in Figure 5. All tested compression systems led to a reduction in the output SNR, which was most pronounced at higher input SNRs. The fast-acting compressor reduced the output SNR by up to 4.8 dB, while the slow-acting system was closest to the linear reference condition. The SNR-aware compressor produced a consistently higher output SNR than the fast-acting system over the complete range of input SNRs. This benefit was about 2 dB at higher input SNRs and was very similar for the SNR-aware and the ideal SNR-aware compressors.

In general, the objective metrics computed for the SNR-aware and the ideal SNR-aware compressor were very similar, suggesting that the accuracy of the SNR estimator was sufficiently high. The performance of the speech detection algorithm is summarized in Table 3 in terms of the hit rate (H), the false alarm rate (FA), and the H − FA for different frequency channels. While the H − FA was not higher than 34.7 % for the lowest two frequency channels, performance increased up to 59.0 % at higher center frequencies.

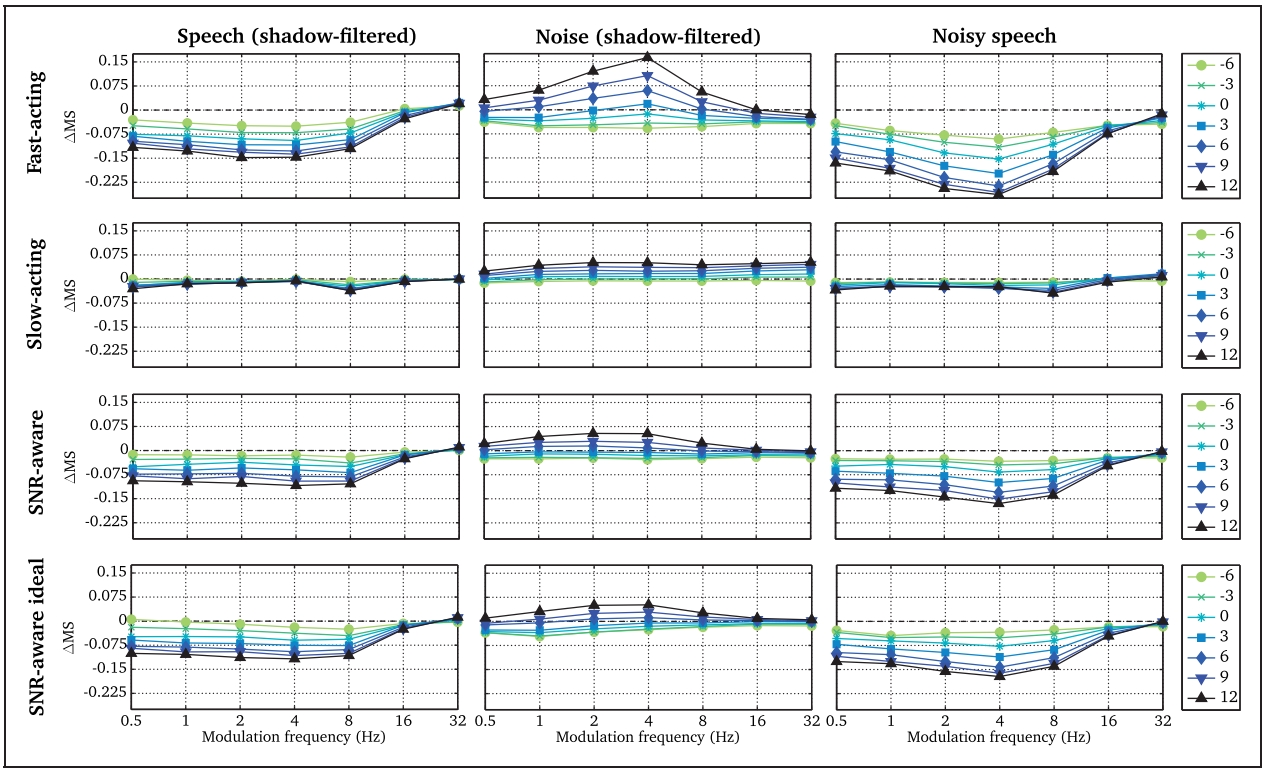

Figure 4. Relative change in modulation spectra (ΔMS) caused by fast-acting (first row), slow-acting (second row), SNR-aware (third row), and ideal SNR-aware compression (fourth row) as a function of the modulation frequency and the input SNR. Results were averaged across all four noise types. The black dashed line indicates the zero line while the left, middle, and right columns show results for shadow-filtered speech, shadow-filtered noise, and noisy speech, respectively. SNR = signal-to-noise ratio.

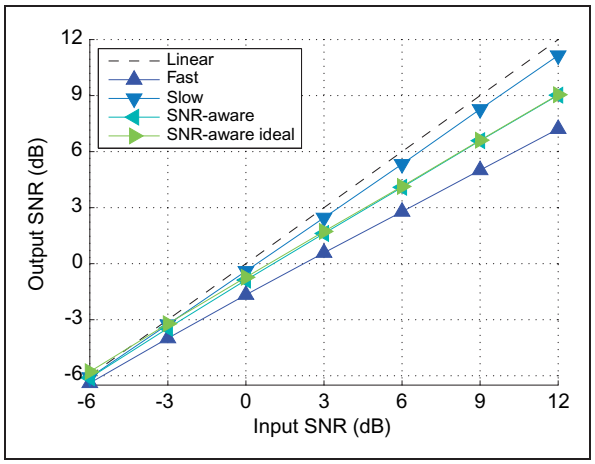

Figure 5. Input/output SNR analysis for the four different compression schemes and a linear system averaged across all four noise types. SNR = signal-to-noise ratio.

Table 3. Performance Analysis of the Binary Speech Detection Algorithm in Terms of H, FA, and H − FA in Percentage as a Function of the Channel Center Frequency Averaged Across All Noise Types and SNRs.

	Channel center frequency (Hz)						
Rates (%)	125	250	500	1000	2000	4000	8000
H	53.1	55.2	67.5	72.0	74.0	73.2	81.3
FA	18.9	20.5	13.5	15.9	18.3	21.8	22.3
H − FA	34.2	34.7	54.0	56.1	55.7	51.4	59.0

Note. H = hit rate; FA = false alarm rate; H − FA = hit rate minus false alarm rate; SNR = signal-to-noise ratio.

Discussion

The analysis of ΔMS indicated that distortions of the speech components are an inevitable consequence of fast-acting compression. A rapidly changing gain function reduces the temporal contrasts of the speech components which, in turn, reduces the modulation power. This is also reflected in the ECRs, which are highest for the fast-acting compression scheme. As pointed out by Villchur (1989), the reduction in modulation power is not necessarily detrimental, as long as it coincides with an improvement in speech audibility. At the same time, fast-acting compression increases the modulation depth of noise signal components at positive SNRs. As shown in Figure 4, the largest increase was

found around the 4-Hz region, which corresponds to the typical maximum in the speech modulation spectrum (e.g., Plomp, 1983; Souza & Gallun, 2010). This results from the compressor gain following short-term fluctuations in the intensity of the dominating speech signal, which disrupts the natural fluctuations in the background noise. As a consequence, the glimpses of noise that are cyclically amplified because of the increased gain during the speech pauses may lead to a sensation of *pumping* and increased overall noisiness (Neuman et al., 1998). Such processing thus is likely to decrease the SNR in the modulation domain, which has been proposed to be detrimental for speech intelligibility (Jørgensen & Dau, 2011; Jørgensen, Ewert, & Dau, 2013). Furthermore, the long-term level of the noise is increased at the output of the compressor, causing a reduced output SNR (Naylor & Johannesson, 2009). In contrast, slow-acting compression avoids the amplification of the noise components. As shown in Figure 2, the changes in the gain function of the slow-acting system do not follow the fluctuations of speech very closely. Therefore, distortions in the modulation spectrum of the noise components, as shown in Figure 4, are of much smaller magnitude. This leads to a more linear behavior in terms of the input/output SNR analysis. However, a slow-acting system does not provide any substantial compression to the speech signal components.

The SNR-aware compression scheme appears to combine the desired properties of the two conventional systems. The analysis of the ECR suggests that the effective compression of speech embedded in noise, as provided by the SNR-aware system, is very similar to the one obtained with conventional fast-acting compression. This behavior should be advantageous, as it is linked to improved audibility (Alexander & Rallapalli, 2017). At the same time, the fluctuations in the gain function become much slower when speech is absent, which avoids the amplification of noise-only segments and increases the output SNR relative to that obtained with fast-acting compression. This is also reflected in the ECRs associated with the noise components, which closely resemble the behavior of the slow-acting compressor. Thus, the SNR-aware compression scheme maintains the acoustic properties of the background noise similar to slow-acting compression while applying fast-acting compression to the speech signal components. Preserving the modulation fidelity of the background noise may facilitate the target-background segregation, improve the perceived quality of the acoustic scene, and aid speech recognition in adverse conditions.

The SNR-aware compression scheme utilizes an estimation of the short-term SNR to detect speech-dominated T-F units. The estimation accuracy of this speech detection stage, as reflected by the H − FA, was as high as 59% and generally in a similar range as the speech detector used in the DRR–aware compression scheme (Hassager et al., 2017). Instead of using the output of the speech detection stage directly for noise reduction, the binary classification of speech activity was used to adaptively select different time constants for speech and noise components. Thus, estimation errors in the speech detection stage do not introduce clearly audible artifacts, and only limit the effective compression of speech components. In a binaural setup with two hearing aids, the estimation of speech activity could be further improved by spatial cues (May, van de Par, & Kohlrausch, 2011), which would allow the application of fast-acting compression to speech-dominated T-F units corresponding to a target source at a specific spatial location.

Conclusion

This study presented a scene-aware amplification strategy that adaptively changes the characteristics of the compressor depending on the estimated speech activity in individual T-F units. Specifically, fast-acting compression was applied to speech-dominated T-F units where the SNR was high, while slow-acting compression was performed for noise-dominated T-F units with a low SNR. A systematic analysis using three technical metrics showed that this SNR-aware compression scheme achieved similar ECRs compared with conventional fast-acting compression, while the natural fluctuations in the background noise were preserved in a similar way as processing the noise components with a conventional slow-acting system. Future work will quantify the subjective benefit of the SNR-aware compression scheme by performing behavioral listening tests.

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 ofthis article: This research was supported by the Technical University of Denmark and funding from Sonova AG (Stäfa, Switzerland).

Note

1. The speech detection performance of the first two frequency channels was relatively poor, probably because of the high temporal resolution of 10 ms which limited the frequency resolution and caused the first two octave filters to be based on only up to five discrete Fourier transform bins. As a consequence, the estimated signal-to-noise ratio was less reliable, which limited the speech detection

performance. Thus, the signal-to-noise ratio estimate of the third channel (500 Hz) was used for the first two channels, for which nevertheless individual thresholds were found as described in the *Parameters* subsection.

References

Alexander, J. M., & Masterson, K. (2014). Effects of WDRC release time and number of channels on output SNR and speech recognition. *Ear and Hearing*, *36*(2), 1–15. doi: 10.1097/AUD.0000000000000115

Alexander, J. M., & Rallapalli, V. (2017). Acoustic and perceptual effects of amplitude and frequency compression on high-frequency speech. *The Journal of Acoustical Society of America*, *142*(2), 908–923. doi: 10.1121/1.4997938

Allen, J. B. (1977). Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *25*(3), 235–238. doi: 10.1109/TASSP.1977.1162950

Bisgaard, N., Vlaming, M. S., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, *14*(2), 113–120. doi: 10.1177/1084713810379609

Braida, L., Durlach, N., De Gennaro, S., Peterson, P., Bustamante, D., Studebaker, G., & Bess, F. (1982). Review of recent research on multiband amplitude compression for the hearing impaired. In G. A. Studebaker, & F. H. Bess (Eds.), *The Vanderbilt hearing aid report* (pp. 123–140). Upper Darby, PA: York Press.

Crochiere, R. E. (1980). A weighted overlap-add method of short-time Fourier analysis/synthesis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *28*(1), 99–102. doi: 10.1109/TASSP.1980.1163353

Davies-Venn, E., Souza, P., Brennan, M., & Stecker, G. C. (2009). Effects of audibility and multichannel wide dynamic range compression on consonant recognition for listeners with severe hearing loss. *Ear and Hearing*, *30*(5), 494–504. doi: 10.1097/AUD.0b013e3181aec5bc

Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., & D'Aquila, L. A. (2017). Masking release for hearing-impaired listeners: The effect of increased audibility through reduction of amplitude variability. *The Journal of Acoustical Society of America*, *141*(6), 4452–4465. doi: 10.1121/1.4985186

Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *International Journal of Audiology*, *40*(3), 148–157. doi: 10.3109/00206090109073110

Eaton, J., Brookes, M., & Naylor, P. A. (2013). A comparison of non-intrusive SNR estimation algorithms and the use of mapping functions. *Proceedings of the European Signal Processing Conference*, 1–5.

Edwards, B. (2004). Hearing aids and hearing impairment. In S. Greenberg, W. A. Ainsworth, & R. R. Fay (Eds.), *Speech processing in the auditory system* (Chapter 17, pp. 339–421). New York, NY: Springer.

Erkelens, J. S., Hendriks, R. C., Heusdens, R., & Jensen, J. (2007). Minimum mean-square error estimation of discrete Fourier coefficients with generalized gamma priors. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(6), 1741–1752. doi: 10.1109/TASL.2007.899233

Fredelake, S., Holube, I., Schlueter, A., & Hansen, M. (2012). Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms. *International Journal of Audiology*, *51*(4), 299–308. doi: 10.3109/14992027.2011.645075

Gallun, F., & Souza, P. (2008). Exploring the role of the modulation spectrum in phoneme recognition. *Ear and Hearing*, *29*(5), 800–813. doi: 10.1097/AUD.0b013e31817e73ef

Gatehouse, S., Naylor, G., & Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, *42*(Suppl. 1), 77–85. doi: 10.3109/14992020309074627

Gatehouse, S., Naylor, G., & Elberling, C. (2006). Linear and nonlinear hearing aid fittings—1. Patterns of benefit. *International Journal of Audiology*, *45*(3), 130–152. doi: 10.1080/14992020500429518

Grimm, G., Herzke, T., Berg, D., & Hohmann, V. (2006). The master hearing aid: A PC-based platform for algorithm development and evaluation. *Acta Acustica United with Acustica*, *92*(4), 618–628.

Gustafsson, S., Martin, R., & Vary, P. (1996). On the optimization of speech enhancement systems using instrumental measures. *Proceedings of the Workshop on Quality Assessment in Speech, Audio and Image Communication*, 36–40.

Hagerman, B., & Olofsson, Å. (2004). A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acustica United with Acustica*, *90*(2), 356–361.

Hassager, H. G., May, T., Wiinberg, A., & Dau, T. (2017). Preserving spatial perception in rooms using direct-sound driven dynamic range compression. *The Journal of Acoustical Society of America*, *141*(6), 4556–4566. doi: 10.1121/1.4984040

Hendriks, R. C., Heusdens, R., & Jensen, J. (2010). MMSE-based noise PSD tracking with low complexity. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4266–4269. doi: 10.1109/ICASSP.2010.5495680

Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language, and Hearing Research*, *48*(3), 651–667. doi: 10.1044/1092-4388(2005/045)

Jenstad, L. M., & Souza, P. E. (2007). Temporal envelope changes of compression and speech rate: Combined effects on recognition for older adults. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1123–1138. doi: 10.1044/1092-4388(2007/078)

Jerlvall, L., & Lindblad, A. (1978). The influence of attack time and release time on speech intelligibility: A study of the effects of AGC on normal hearing and hearing impaired subjects. *Scandinavian Audiology Supplementum*, (6), 341–353.

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*

Acoustical Society of America, *130*(3), 1475–1487. doi: 10.1121/1.3621502

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

Kates, J. M. (2010). Understanding compression: Modeling the effects of dynamic-range compression in hearing aids. *International Journal of Audiology*, *49*(6), 395–409. doi: 10.3109/14992020903426256

Keidser, G., Dillon, H., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, *1*(e24), 88–90. doi: 10.4081/audiores.2011.e24

Killion, M. C., Teder, H., Johnson, A. C., & Hanke, S. P. (1992). Variable recovery time circuit for use with wide dynamic range automatic gain control for hearing aid. *U.S. Patent No. 5,144,675*. Washington, DC: U.S. Patent and Trademark Office.

Kuk, F. K. (1996). Theoretical and practical considerations in compression hearing aids. *Trends in Amplification*, *1*(1), 5–39. doi: 10.1177/108471389600100102

Lai, Y. H., Li, P. C., Tsai, K. S., Chu, W. C., & Young, S. T. (2013). Measuring the long-term SNRs of static and adaptive compression amplification techniques for speech in noise. *Journal of the American Academy of Audiology*, *24*(8), 671–683. doi: 10.3766/jaaa.24.8.4

May, T., Kowalewski, B., Fereczkowski, M., & MacDonald, E. N. (2017). Assessment of broadband SNR estimation for hearing aid applications. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 231–235. doi: 10.1109/ICASSP.2017.7952152

May, T., van de Par, S., & Kohlrausch, A. (2011). A probabilistic model for robust localization based on a binaural auditory front-end. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(1), 1–13. doi: 10.1109/TASL.2010.2042128

Moore, B. C., & Glasberg, B. R. (1988). A comparison of four methods of implementing automatic gain control (AGC) in hearing aids. *British Journal of Audiology*, *22*(2), 93–104.

Naylor, G., & Johannesson, R. B. (2009). Long-term signal-to-noise ratio at the input and output of amplitude-compression systems. *Journal of the American Academy of Audiology*, *20*(3), 161–171. doi: 10.3766/jaaa.20.3.2

Neuman, A. C., Bakke, M. H., Mackersie, C., Hellman, S., & Levitt, H. (1998). The effect of compression ratio and release time on the categorical rating of sound quality. *The Journal of Acoustical Society of America*, *103*(5), 2273–2281. doi: 10.1121/1.422745

Neumann, J. (2008). Method for dynamic determination of time constants, method for level detection, method for compressing an electric audio signal and hearing aid, wherein the method for compression is used. *U.S. Patent No. 7,333,623*. Washington, DC: U.S. Patent and Trademark Office.

Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International Journal of Audiology*, *50*(3), 202–208. doi: 10.3109/14992027.2010.524254

Plomp, R. (1983). Hearing—Physiological bases and psychophysics. In R. Klinke, & R. Hartmann (Eds.), *The role of modulation in hearing* (pp. 270–276). Berlin, Germany: Springer.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of Acoustical Society of America*, *83*(6), 2322–2327. doi: 10.1121/1.396363

Rhebergen, K. S., Maalderink, T. H., & Dreschler, W. A. (2017). Characterizing speech intelligibility in noise after wide dynamic range compression. *Ear and Hearing*, *38*(2), 194–204. doi: 10.1097/AUD.0000000000000369

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2009). The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise. *The Journal of Acoustical Society of America*, *126*(6), 3236–3245. doi: 10.1121/1.3257225

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. doi: 10.1098/rstb.1992.0070

Simonsen, C., & Behrens, T. (2009). A new compression strategy based on a guided level estimator. *Hearing Review*, *16*(13), 26–31.

Souza, P., & Gallun, F. (2010). Amplification and consonant modulation spectra. *Ear and Hearing*, *31*(2), 268–276. doi: 10.1097/AUD.0b013e3181c9fb9c

Souza, P., Hoover, E., & Gallun, F. (2012). Application of the envelope difference index to spectrally sparse speech. *Journal of Speech, Language, and Hearing Research*, *55*(3), 824–837. doi: 10.1044/1092-4388(2011/10-0301)

Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends in Amplification*, *6*(4), 131–165. doi: 10.1177/108471380200600402

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. doi: 10.3766/jaaa.18.1.5

Souza, P. E., Jenstad, L. M., & Boike, K. T. (2006). Measuring the acoustic effects of compression amplification on speech in noise. *The Journal of Acoustical Society of America*, *119*(1), 41–44. doi: 10.1121/1.2108861

Souza, P. E., & Turner, C. W. (1996). Effect of single-channel compression on temporal speech information. *Journal of Speech, Language, and Hearing Research*, *39*(5), 901–911. doi: 10.1044/jshr.3905.901

Souza, P. E., & Turner, C. W. (1998). Multichannel compression, temporal cues, and audibility. *Journal of Speech, Language, and Hearing Research*, *41*(2), 315–326. doi: 10.1044/jslhr.4102.315

Souza, P. E., & Turner, C. W. (1999). Quantifying the contribution of audibility to recognition of compression-amplified speech. *Ear and Hearing*, *20*(1), 12–20.

Stone, M. A., & Moore, B. C. (1992). Syllabic compression: Effective compression ratios for signals modulated at different rates. *British Journal of Audiology*, *26*(6), 351–361. doi: 10.3109/03005369209076659

Stone, M. A., & Moore, B. C. (2003). Effect of the speed of a single-channel dynamic range compressor on intelligibility

in a competing speech task. *The Journal of Acoustical Society of America*, *114*(2), 1023–1034. doi: 10.1121/1.1592160

Stone, M. A., & Moore, B. C. (2004). Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. *The Journal of Acoustical Society of America*, *116*(4), 2311–2323. doi: 10.1121/1.1784447

Stone, M. A., & Moore, B. C. (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *The Journal of Acoustical Society of America*, *121*(3), 1654–1664. doi: 10.1121/1.2434754

Stone, M. A., & Moore, B. C. (2008). Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task. *The Journal of Acoustical Society of America*, *123*(2), 1063–1076. doi: 10.1121/1.2821969

Stone, M. A., Moore, B. C., Alcántara, J. I., & Glasberg, B. R. (1999). Comparison of different forms of compression using wearable digital hearing aids. *The Journal of the Acoustical Society of America*, *106*(6), 3603–3619. doi: 10.1121/1.428213

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

Varga, A. P., & Steeneken, H. J. M. (1993). Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication*, *12*(3), 247–251. doi: 10.1016/0167-6393(93)90095-3

Villchur, E. (1973). Signal processing to improve speech intelligibility in perceptive deafness. *The Journal of Acoustical Society of America*, *53*(6), 1646–1657. doi: 10.1121/1.1913514

Villchur, E. (1989). Comments on the negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function [*J. Acoust. Soc. Am. 83*, 2322–2327 (1988)]. *The Journal of Acoustical Society of America*, *86*(1), 425–427. doi: 10.1121/1.398306

Walaszek, J. (2008). *Effects of compression in hearing aids on the envelope of the speech signal, signal based measures of the side-effects of the compression and their relation to speech intelligibility* (Unpublished Master's thesis). Technical University of Denmark, Lyngby, Denmark.

Yund, E. W., & Buckles, K. M. (1995). Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. *The Journal of Acoustical Society of America*, *97*(2), 1224–1240. doi: 10.1121/1.412232

Yund, E. W., Simon, H. J., & Efron, R. (1987). Speech discrimination with an 8-channel compression hearing aid and conventional aids in background of speech-band noise. *Journal of Rehabilitation Research and Development*, *24*(4), 161–180.