

Effects of Slow- and Fast-Acting Compression on Hearing-Impaired Listeners' Consonant–Vowel Identification in Interrupted Noise

Trends in Hearing
Volume 22: 1–12
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216518800870
journals.sagepub.com/home/tia



Borys Kowalewski¹ , Johannes Zaar¹, Michal Fereczkowski¹,
Ewen N. MacDonald¹ , Olaf Strelcyk², Tobias May¹ ,
and Torsten Dau¹

Abstract

There is conflicting evidence about the relative benefit of slow- and fast-acting compression for speech intelligibility. It has been hypothesized that fast-acting compression improves audibility at low signal-to-noise ratios (SNRs) but may distort the speech envelope at higher SNRs. The present study investigated the effects of compression with a nearly instantaneous attack time but either fast (10 ms) or slow (500 ms) release times on consonant identification in hearing-impaired listeners. Consonant–vowel speech tokens were presented at a range of presentation levels in two conditions: in the presence of interrupted noise and in quiet (with the compressor “shadow-controlled” by the corresponding mixture of speech and noise). These conditions were chosen to disentangle the effects of consonant audibility and noise-induced forward masking on speech intelligibility. A small but systematic intelligibility benefit of fast-acting compression was found in both the quiet and the noisy conditions for the lower speech levels. No detrimental effects of fast-acting compression were observed when the speech level exceeded the level of the noise. These findings suggest that fast-acting compression provides an audibility benefit in fluctuating interferers when compared with slow-acting compression while not substantially affecting the perception of consonants at higher SNRs.

Keywords

hearing aids, hearing loss, amplification, speech

Date received: 1 February 2018; revised: 7 August 2018; accepted: 21 August 2018

Introduction

It is widely accepted that due to the limited dynamic range of levels perceived by hearing-impaired (HI) listeners, level-dependent amplification is required to compensate for hearing loss. The majority of modern hearing aids apply dynamic-range compression (see Edwards, 2004; Souza, 2002 for reviews). In such systems, the gain is determined by one or more level-estimation circuits, which are characterized by attack and release time constants. The most commonly used attack times have values below 10 ms (Jenstad & Souza, 2005) to quickly reduce the gain in response to loud sounds. However, the optimal speed of gain recovery, that is, the release time, is still a subject of discussion. Shorter release times allow more gain to be applied to the low-intensity speech components

(e.g., consonants) that follow other, high-intensity components (e.g., vowels) or noise bursts. This increased gain can potentially improve audibility and reduce the amount of forward masking, which in turn might lead to an improved speech recognition performance in HI listeners (Alexander & Rallapalli, 2017; Desloge, Reed, Braida, Perez, & D'Aquila, 2017; Desloge, Reed, Braida, Perez,

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

²Sonova U.S. Corporate Services, Warrenville, IL, USA

Corresponding author:

Borys Kowalewski, Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads 352, DK-2800 Lyngby, Denmark.

Email: bokowal@elektro.dtu.dk



& Delhorne, 2010; Edwards, 2002; Jenstad & Souza, 2005; Souza & Turner, 1998, 1999). On the other hand, with a very short release time, the gain follows the fast fluctuations of the signal, effectively reducing the temporal contrast and altering natural modulations in speech (Alexander & Rallapalli, 2017; Souza & Turner, 1996, 1998; Stone & Moore, 2003, 2004, 2007, 2008). The temporal characteristics of the speech signal provide important cues for speech intelligibility, especially for HI listeners (Rosen, 1992; Souza, Wright, Blackburn, Tatman, & Gallun, 2015). Temporal envelope distortion introduced by fast-acting amplification might therefore lead to a decrement in recognition performance (Jenstad & Souza, 2005, 2007; Walaszek, 2008). It is possible that optimal performance would be achieved if the time constants were adapted dynamically according to the current signal-to-noise ratio (SNR; Gatehouse, Naylor, & Elberling, 2003; Kates, 2010; Souza, Hoover, & Gallun, 2012; Yund & Buckles, 1995). For example, May, Kowalewski, Fereczkowski, and MacDonald (2017) evaluated different methods for blind estimation of the broadband SNR that could be applied in hearing aids. However, the relation between the optimal release time and SNR in connection to speech intelligibility is not yet known.

In the present study, it is hypothesized that potential negative effects of short release times will be more pronounced at higher SNRs, corresponding to higher speech input levels, where audibility and masking are less of a concern, and the compression is driven mostly by the speech signal. On the other hand, the additional gain applied by the fast-acting system is expected to provide an increasing benefit as the SNR decreases. To test these ideas, stimuli were designed to maximize the effects of compression release time. Consonant–vowel (CV) tokens were presented, and listeners were asked to report the consonant—a speech component that typically has a low intensity. The noise consisted of high-intensity bursts, separated by silent gaps and had very sharp onsets and offsets. The temporal onset of the CV token relative to the noise was controlled and chosen based on a previous study (Zaar, Kowalewski, & Dau, 2017). A wide range of SNRs and compression release times were tested to capture the potential interaction between the two factors. In addition, an analysis of the output level and audibility was performed for each experimental condition.

Methods

Listeners

Twelve young, normal-hearing (NH) listeners aged between 19 and 26 years (average age: 21.7 years) completed the task in the unaided condition. They all had

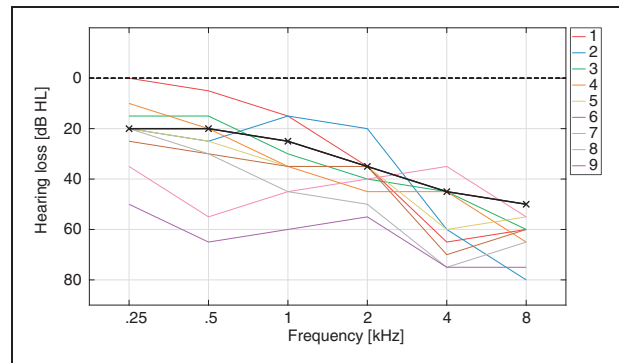


Figure 1. Pure-tone audiograms of the hearing-impaired listeners. Individual audiograms of Listeners 1 to 9 (from lowest to highest pure-tone average) are color coded according to the legend (color available online). The N2 standard audiogram (shown in black, Bisgaard et al., 2010) was used to derive the compression ratios, following the NAL-NL2 procedure (Keidser et al., 2011).

pure-tone thresholds lower than 20 dB HL in the 250 to 8000 Hz range. The aided conditions were completed by nine older HI listeners aged between 66 and 77 years (average age: 71.4 years). Their hearing losses ranged from mild to moderately severe and were most prominent at high frequencies (see audiograms in Figure 1). All participants provided informed written consent, and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

Stimuli

The target speech consisted of 15 consonant–vowel (CV) tokens: /bi, di, fi, gi, hi, ji, ki, li, mi, ni, pi, si, fi, ti, vi/ spoken by one male and one female talker (30 utterances in total), used previously by Zaar and Dau (2015). Four sound pressure levels (SPLs) were used: 45, 55, 65, and 75 dB. In the aided conditions, these were the levels at the *input* of the amplification system. Each utterance was presented to the listeners five times per test condition.

The noise was composed of five 100-ms bursts, separated by 100-ms silent gaps (corresponding to a 5-Hz repetition rate). Although rather artificial in nature, this noise envelope had the advantage that the sharp onsets and offsets and relatively long silent gaps emphasized the effects of compression time constants while ensuring that the consonant portion of speech would be affected by nonsimultaneous masking only (Desloge et al., 2017; Wilson et al., 2010). White noise (bandlimited between 0 and 22050 Hz) was chosen as a carrier to maximize masking of high-frequency consonants and to reduce potential spectral splatter due to noise onsets and offsets. The SPL was 65 dB, defined as the level of the noise bursts at the input to the hearing-aid simulator. The onset of the CV token was positioned 25 ms into

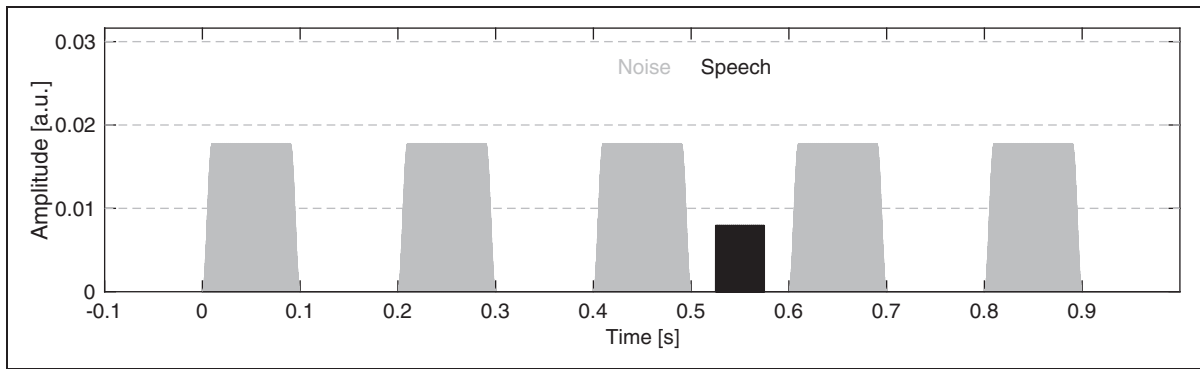


Figure 2. A schematic representation of the stimulus temporal envelope. The noise (shown in gray) consisted of 100-ms long bursts of white noise, separated by 100-ms silent gaps. The speech token (shown in black) was positioned 25 ms after the offset of the third noise burst.

the silent gap after the third noise burst, as shown schematically in Figure 2. The instantaneous SNR was therefore infinite. The broadband SNR values are still reported for consistency with previous literature. They are defined as the difference between the SPL of the token and the preceding noise burst.

Thirty noise waveforms (one per utterance) were pre-generated and stored as WAV files. Each utterance was always presented with the same noise recording. This was done to limit the across-repetition variability due to the random fluctuations in the Gaussian noise carrier, while preventing noise-learning effects that could occur if only one noise waveform was used for all utterances (Rhebergen, Maalderink, & Dreschler, 2017; Zaar & Dau, 2015).

Amplification

For the HI listeners, the stimuli were preprocessed using a hearing-aid simulator with eight independent compression channels, implemented in MATLAB. The signals were sampled at a rate of 44100 Hz and segmented into overlapping frames with a length of 256 samples and 75% overlap. A Hann analysis window was used. Each frame was zero padded to a length of 512 samples and a 512-point discrete Fourier transform (DFT) was computed. The DFT bins were grouped to create a bank of rectangular filters with 20% overlap and a spacing such that, given the DFT resolution, the desired number of channels was as uniform as possible on the equivalent-rectangular-band scale. The cutoff frequencies in Hz are provided in Table 1.

The compression thresholds (kneepoints) were frequency-dependent and calibrated so that compression in each channel was activated when the level of a broadband (white noise) input exceeded 50 dB SPL. The thresholds in each channel are provided in Table 1. Compression speed was dictated by the time constants of a one-pole-filter power-smoothing circuit

(the so-called RC time constants; Kates, 1993). The attack time was always 5 ms. The release time was 10 ms in the *fast* compression condition and 500 ms in the *slow* condition. The third condition was *linear*, which used the same maximum gain values as used in the compression conditions, but a compression ratio (CR) of 1:1, that is, no compression. It thus simulated an “idealized” hearing aid that never applies compression and provides the maximum possible amplification. Such high gain is unrealistic for high-intensity inputs, as it would be excessively loud. Thus, this condition served as a baseline for the behavior of compression systems but only for lower intensity speech inputs—at 45 and 65 dB SPL in quiet (see “Experimental Conditions and Training” section later).

The CRs in each channel were based on the NAL-NL2 target for the N2 audiogram (Bisgaard, Vlaming, & Dahlquist, 2010) using the *Slow* setting, which yields higher CRs (cf. Keidser, Dillon, Flax, Ching, & Brewer, 2011). The N2 audiogram was chosen because it was most representative of the participants’ hearing losses.¹ Thus, CRs were the same for all listeners, ranging between 1.1:1 and 2.3:1 (see Table 1). To maximize audibility for each participant, individualized linear gain was applied after compression. It was based on the insertion gain prescribed by NAL-NL2 for a 50 dB SPL input for individual audiograms. The total gain (after compression and linear compensation) for each time frame was then interpolated from the channel frequencies to the DFT bins. A 512-point inverse DFT was then computed for each frame and a Tukey window with a length of 512 samples, including 128-sample tapered ends, was applied. The output signal was synthesized using the overlap-add procedure and presented monaurally over Sennheiser HD650 headphones.

In all conditions, the level-detection circuit of the compression and the resulting gain were driven by the mixture of speech and noise. Thus, the gain applied to the clean speech in the quiet condition was not controlled

Table 1. Cutoff Frequencies, Compression Ratios, and Thresholds in Each of the Processing Channels.

Channel no.	1	2	3	4	5	6	7	8
Lower cutoff (Hz)	0	86	431	689	1292	2239	3618	4823
Upper cutoff (Hz)	172	517	861	1550	2584	4307	7752	22050
Compression ratio	1.1:1	1.1:1	1.3:1	1.5:1	1.8:1	2.1:1	2.3:1	2.0:1
Threshold (dB SPL)	22.8	32.9	32.9	35.9	37.7	39.7	42.7	48.9

Note. SPL = sound pressure level.

Table 2. Summary of Experimental Conditions: Presence of Noise, Input Level of Speech, and Type of Amplification Used.

	Speech input level (dB SPL)	NH		HI	
		Unaided	Linear	Slow	Fast
Quiet (shadow-controlled)	45	X	X	X	X
	65	X	X	X	X
	75	–	–	X	X
Noise	45	X	–	X	X
	55	X	–	X	X
	65	X	–	X	X
	75	X	–	X	X

Note. NH = normal-hearing; HI = hearing-impaired; SPL = sound pressure level.

by the clean speech signal but rather *shadow-controlled* by the mixture, that is, the gain applied to the speech signal was identical in noisy and quiet conditions. This setup allowed the investigation of the effects of the gain fluctuations (resulting from the presence of the interrupted noise) on the CV token without actually presenting the interferer to the listeners' ears.

Experimental Conditions and Training

The NH listeners were tested unaided, while HI listeners were always presented with amplified stimuli. Slow and fast compression were tested in all conditions, while linear amplification was tested only in a limited number of conditions. An overview of all experimental conditions is shown in Table 2.

Prior to the test runs, the listeners received training in each of the experimental conditions comprising of one presentation of each stimulus (each token spoken by each talker). This was done to familiarize the listeners with the speech material and the different processing types. Subsequently, the test was performed starting from the conditions assumed to be the easiest and progressively increasing the difficulty. Therefore, the quiet conditions were tested before the noisy conditions, and the highest input speech levels were tested first.

Results

For analyzing the quiet and noisy data sets, separate linear mixed-effects models were used with two fixed factors (*speech level* and *amplification type*) and one random factor (*listener*). Backwards elimination of nonsignificant effects was performed (Kuznetsova, Brockhoff, & Christensen, 2018), and the final model was used to establish significance between the results obtained with each *amplification type* at each *speech level*. To establish significance of differences between relevant levels of the interaction between *amplification type* and *speech level* (e.g., the difference in scores between amplification types at a given speech level or the differences across speech level for a given amplification type), the least-squares means approach was used with *p* values adjusted for multiple comparisons using the multivariate *t* adjustments (Lenth, 2017).

The distribution of the model residuals for the data obtained in the quiet conditions deviated from normal (it was “light-tailed”). Therefore, all recognition-rate data were transformed to rationalized arcsine units (RAU) prior to the statistical analysis. The transformation was not necessary for the data obtained in the noise-present conditions (the distribution of residuals was much closer to normal). Applying the transformation, however, did not affect the main conclusion of the analysis, and it was used in all cases for the sake of consistency. The recognition rates are nonetheless reported here in their nontransformed form for ease of interpretation.

Consonant Recognition in Quiet

The average consonant recognition rates for the stimuli presented in quiet are shown in Figure 3. It can be observed that the NH listeners (unaided) achieved recognition rates close to 100% at both speech input levels. The HI listeners (aided) performed much worse in all conditions and achieved maximum recognition rates of about 87% at 75 dB SPL. Significant differences were found between all amplification types for the lowest speech input level (45 dB SPL, $p < .05$ for linear vs. fast, $p < .001$ for linear vs. slow, $p < .01$ for fast vs. slow). The best recognition rate of 55% was achieved

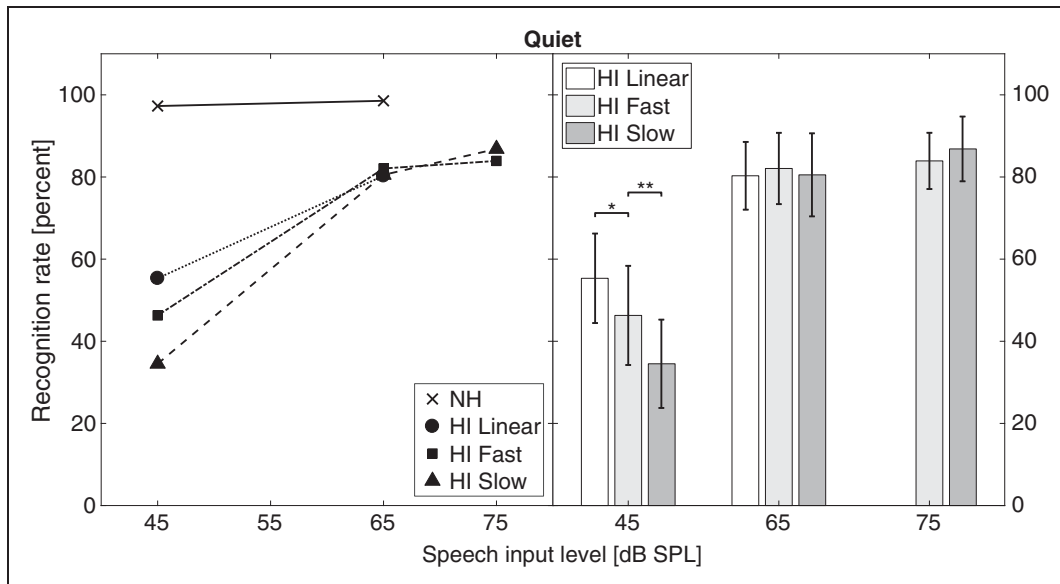


Figure 3. Averaged consonant recognition rates for speech tokens in quiet, “shadow-controlled” by the mixture of speech and noise. Left: normal-hearing (NH) unaided and hearing-impaired (HI) aided with three types of amplification. Right: Only the HI data replotted. The error bars indicate \pm one standard deviation. The significance levels are *.05, **.01, ***.001. NH = normal-hearing; HI = hearing-impaired; SPL = sound pressure level.

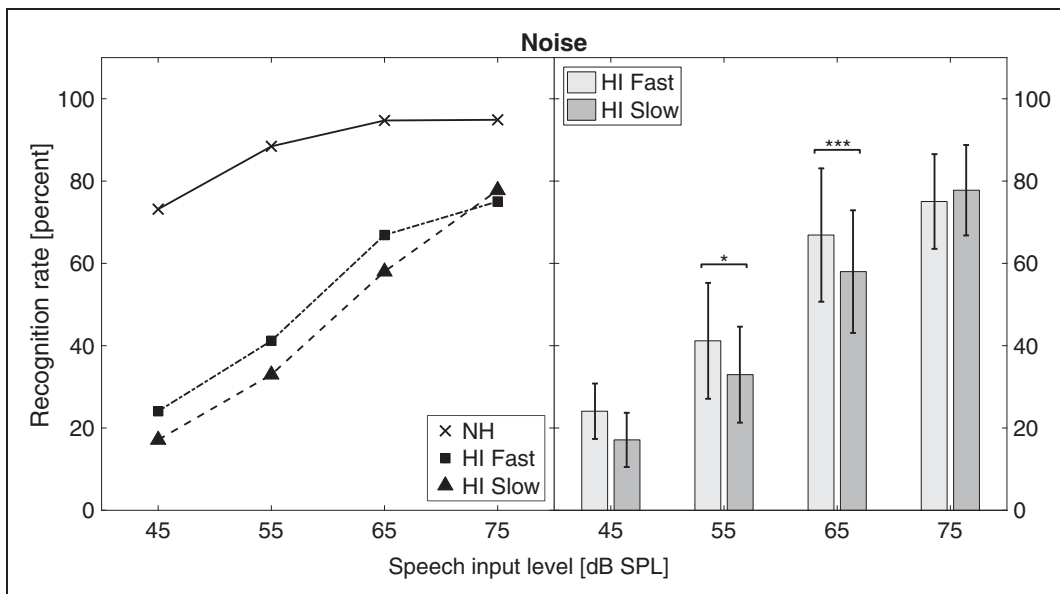


Figure 4. The same as Figure 3 but presented in 5-Hz interrupted Gaussian noise (noise level: 65 dB SPL). NH = normal-hearing; HI = hearing-impaired; SPL = sound pressure level.

with linear amplification (that provided the highest gain), followed by fast (46%) and slow compression (34%).

For the 65 dB SPL and 75 dB SPL speech input, no statistically significant differences between amplification types were observed. There is also no significant increase in performance between 65 and 75 dB SPL with either amplification type.

Consonant Recognition in Noise

The recognition rates in noise are shown in Figure 4. NH listeners achieved recognition rates of 95% for speech levels of 65 and 75 dB SPL (corresponding to SNRs of 0 and +10 dB). The rate decreased to 73% at 45 dB SPL (−20 dB SNR). Aided HI listeners achieved the highest

recognition rate of 77% at 75 dB SPL. For speech input levels of 55 and 65 dB SPL, the recognition rates observed with fast compression were 8% to 9% higher than with slow compression, with both differences being statistically significant ($p < .05$ and $p < .001$, respectively). At 45 and 75 dB SPL, there were no statistically significant differences in the recognition performance obtained with fast and slow compression.

Consonant Output Level

Different amplification strategies provide different gains, which might influence the measured consonant recognition performance. Therefore, an analysis of the output speech levels was conducted. The root mean square (RMS) level of a CV is highly dependent on the vowel portion. To assess the amount of gain applied to the consonant, the output levels of the consonant portions were therefore measured for all possible combinations of speech tokens, listeners' individual amplification settings, and input levels. The average values are shown in Figure 5. At each input level, the differences in output between the amplification conditions directly translate to the differences in the average gain they provide. The linear condition provided the highest gain and therefore the highest output level. At the lowest speech input level, fast- and slow-acting compression provide gain that is, respectively, 1.7 and 7.0 dB lower than linear. As the input level increases, the consonant at the output of both compressors receives progressively

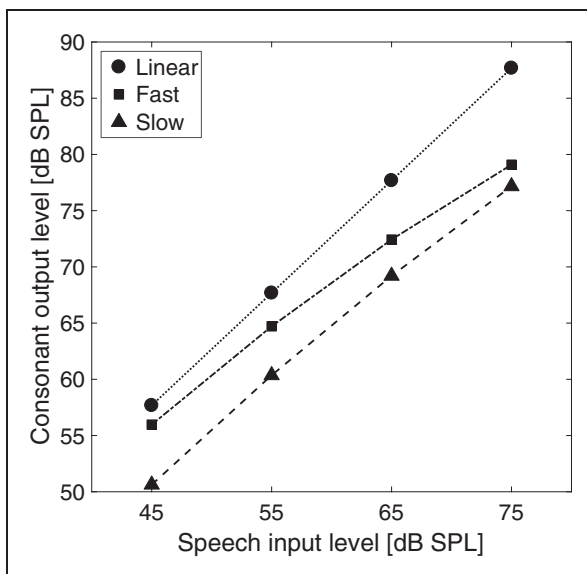


Figure 5. Averaged output level of the consonant portion of the stimulus for different amplification schemes, as a function of the overall speech input level.
SPL = sound pressure level.

less overall gain. This decrease in gain is more pronounced for fast-acting compression. At the highest input levels, the outputs of the two compression systems converge, resulting in a difference of less than 2 dB at 75 dB SPL speech input level.

Consonant Audibility and Performance

To further quantify the effect that the amount of audible information has on each listener's performance, a metric of audibility was computed for each combination of consonant, talker, listener, input level, and processing condition. The consonant-only portion of the processed stimulus (at the output of the hearing-aid simulator) was considered. The digital signal level was measured in octave bands with center frequencies from 250 to 8000 Hz and converted to dB SPL. The individual listeners' audiometric thresholds in dB SPL were subtracted from these values. A lower limit for the audiogram-corrected levels was assumed to be 0 dB (re. audiometric threshold). Subsequently, the audiogram-corrected levels were weighted using the band-importance functions for nonsense syllables from Pavlovic (1987) and summed across frequency yielding a measure of consonant audibility (in arbitrary units).

Each listener's audibility metric values obtained in each experimental condition were averaged across stimuli (consonant and talker). Figure 6 shows a scatter plot of the resulting averaged audibility metric versus the corresponding RAU-transformed recognition scores. An overall trend of performance improvement with increasing audibility can be observed. The Spearman's rank correlation coefficient (ρ) is 0.708 between the recognition score and the audibility index for the data in quiet and 0.797 for the data in noise. However, it is evident that the rate of performance growth changes when the audibility index reaches values of about 3 to 4. This is likely due to listeners reaching ceiling performance or, in some cases, due to a "roll-over effect." Therefore, a two-section piecewise linear model was fitted to the data. The best-fitting models are shown in black in Figure 6. In quiet, the RMS error of the model fit is 22.3 RAU. In noise, the RMS error is equal to 17.7 RAU. Most of the variability may be attributed to different listeners showing different sensitivity to a similar increase in audibility. Nevertheless, individual fits were not performed, due to the risk of overfitting to a limited number of data points per listener.

Discussion

Consonant recognition in noise was measured for HI listeners aided with a hearing-aid simulator and unaided NH listeners (used as a reference). Linear, fast- and slow-acting compression settings were considered. Interrupted noise was used as a masker, with sharp onsets and offsets

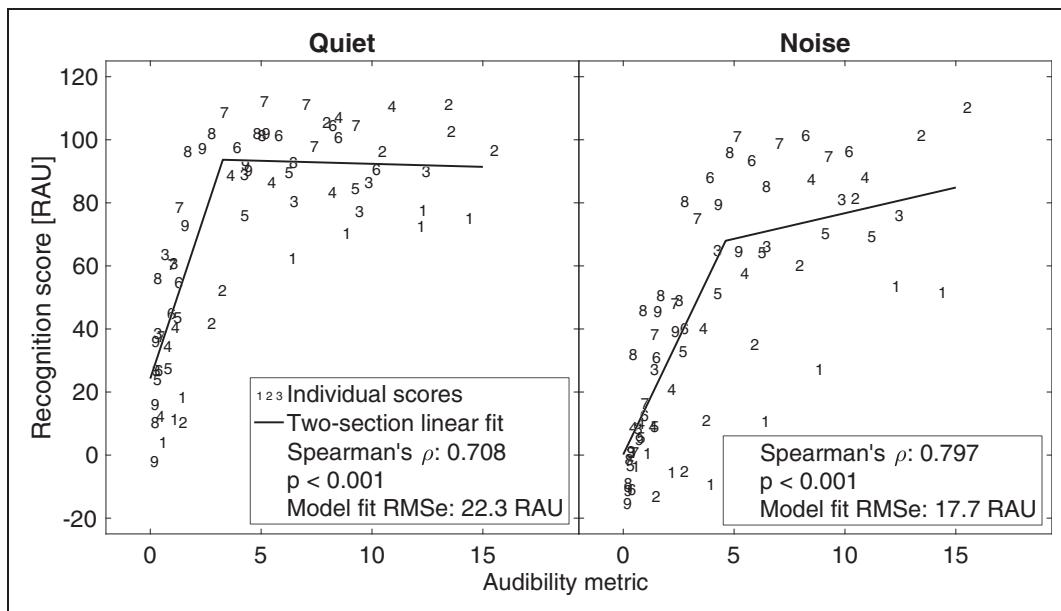


Figure 6. RAU-transformed recognition scores for each HI listener and experimental condition, averaged across stimuli (consonant and talker), versus the corresponding values of the audibility metric for the consonant portion of the CV stimulus. Left panel: results in quiet; right panel: results in noise. Numbers indicate individual listeners, numbered from lowest to highest pure-tone average as in Figure 1. The best fit two-section linear model is shown in black. The Spearman's rank correlation coefficient ρ for the data, the corresponding correlation p values, and the RMS error of the model fits are reported in each panel. RAU = rationalized arcsine unit; RMS = root mean square.

and relatively long, silent gaps, during which the speech token was presented. This was designed to maximize the differences between fast- and slow-acting compression. In addition, a quiet condition was considered in which the noise was not presented to the listeners, but the gain applied to the speech stimulus was shadow-controlled by the corresponding noisy mixture. Thus, it is possible to disentangle the effects of forward masking and reduced gain due to compression.

In the quiet, shadow-controlled condition, the consonant recognition rates obtained by HI listeners at low speech input levels strongly depended on the amplification type. The best performance was obtained with linear amplification and fast compression, which provided higher gain and thus better audibility than slow compression. However, the effect was present only at the lowest speech input level. At the speech levels of 65 and 75 dB SPL, there was no statistically significant difference in the average performance between the two compression conditions. There was also no increase in performance between 65 and 75 dB SPL, which suggests a ceiling effect. Audibility seems to be important for recognition in quiet, as the proposed metric of audibility is strongly positively correlated with listener's performance.

In contrast to the unaided NH listeners, HI listeners on average did not reach 100% performance, even when provided with a relatively high level of speech in quiet. However, there was a substantial spread in the

recognition scores among the listeners, as shown in the right panel of Figure 3. As seen in Figure 6, some listeners do not benefit from the relatively high audibility that was provided to them. Their poor performance may be attributed to suprathreshold distortion that potentially occurs in the impaired auditory system. Amplification is unlikely to reduce the consequences of this distortion and might even be detrimental at high levels. The interaction of cognitive status and fast-versus slow-acting compression has been investigated in many studies (e.g., Gatehouse et al., 2003; Lunner & Sundewall-Thorén, 2007; Ohlenforst, Souza, & MacDonald, 2016). While no cognitive assessment of the participants was conducted in the present study, it should be noted that the stimuli in the present study were nonsense CV tokens rather than intelligible speech. Thus, it was not possible for the HI listeners to make use of context or other top-down processing to improve performance. If the HI listeners in the present study are representative, one may surmise that near-normal speech intelligibility by HI listeners for contextual speech in quiet must require substantial top-down processing and listening effort.

Due to a large age difference between the NH and HI groups (nearly 50 years, on average), it is not possible to separate the effects of age and hearing loss on speech recognition. For example, Gordon-Salant, Yeni-Komshian, Fitzgibbons, and Barrett (2006) found that

age had an effect on the listeners' sensitivity to temporal cues used for consonant recognition.

In noise, fast-acting compression led to higher recognition rates for speech levels of 55 and 65 dB SPL, corresponding to a broadband SNR of -10 and 0 dB, respectively. Similar to the results in quiet, there was no statistically significant difference in performance at 75 dB SPL ($+10$ dB SNR) between the two compression conditions. Therefore, there was no statistically significant evidence for detrimental effects of fast-acting compression (e.g., due to temporal envelope distortion) on consonant recognition performance at higher speech input levels. However, possible ceiling effects in some of the HI listeners' data may be a confounding factor here. The difference in the overall gain applied by fast- and slow-acting compression becomes relatively small at 75 dB SPL, as seen in Figure 5. The additional 2 dB of gain provided by the fast-acting system might be insufficient to provide a measurable benefit in terms of speech recognition. Moreover, the average benefit at 45 dB SPL in noise was 7% but with a p value of .076. Therefore, it is possible that the difficulty of the task at the lowest speech level contributed to the large uncertainty (relative to the difference between the average scores between the two compression conditions) and that significance would have been reached if a greater number of listeners had been considered.

As the compressor was shadow-controlled by the mixture of speech and noise when it was applied to speech in quiet, it behaved identically in both conditions. Therefore, a difference in performance between the two conditions can be attributed only to the presence of noise (i.e., forward masking). In quiet, the relative benefit of fast versus slow compression was 12% (p value $< .01$) at 45 dB SPL but not statistically significant at 65 dB SPL (-20 and 0 dB SNR, respectively). In noise, on the other hand, the benefit was statistically significant at 55 and 65 dB SPL but not at 45 dB SPL. This may be due to the higher gain provided to the speech token by fast-acting compression, which may improve the recovery from the noise-induced forward masking (at least at SNRs close to 0 dB, i.e., speech levels close to 65 dB SPL). Even though the speech target and the interrupted-noise masker do not overlap in time, they likely interact in the auditory system due to its limited temporal resolution. One could hypothesize that successful recognition is dependent on an "internal" or "neural" SNR as "seen" by the higher stages of the auditory system rather than the "external" or "acoustical" SNR measured at the eardrum. The metric of audibility seems to be strongly correlated with individual performance in noise, with Spearman's ρ near 0.8 . In this context, greater values of the audibility metric could be linked to an increased ability to overcome forward masking, improving the internal SNR across a range of auditory channels. The limitation of this approach is that, due to the flat power spectrum density of the

masking noise, the power within an auditory channel increases with its center frequency. This could potentially lead to varying sensitivity to forward masking across frequency, which the proposed model does not take into account. However, in a yet unpublished study, the authors measured the recovery from forward masking using a wideband noise masker with a flat power spectrum density and tonal probes at 1 and 4 kHz. The results suggest that, at least in NH listeners, the recovery is less frequency-dependent than would be expected from the simple increase in power within the equivalent rectangular band between 1 and 4 kHz (the measured 2–3 dB increase in masked thresholds vs. the expected 5.4 dB). This could be explained, for example, by an internal compensation mechanism, such that at higher frequencies a lower SNR (a higher efficiency factor k , see Plack & Oxenham, 2002) is needed for the detection or, in the present case, for consonant identification.

Hearing-aid compression can also distort the envelope of the signal. Effects of CR and release time on envelope fidelity and speech recognition have been studied by Jenstad and Souza (2007), who quantified the amount of temporal envelope distortion using the envelope difference index (EDI; Fortune, Woodruff, & Preves, 1994). They found that the EDI (and hence the amount of distortion) increased with shorter release times and larger CRs. However, a detrimental effect on speech recognition was apparent only if the EDI values were above 0.25 , which was attainable only with CRs of 4:1 and higher. Because the CRs used in the present study did not exceed 2.3:1, they were not high enough to substantially alter the signal envelopes. Specifically, the median values of the EDI (computed from the consonant portions of all the stimuli used here) were 0.058 and 0.022 for the fast- and slow-acting compression, respectively. Thus, even though a short release time led to a higher EDI, the values were well below the 0.25 threshold proposed by Jenstad and Souza (2007), and therefore, the temporal envelope distortion likely had little influence on speech perception. The spectral envelope of the stimulus can also be affected by multichannel fast-acting compression. This effect was presumably small, but it has not been quantified in any way which poses a potential confound. However, because fast-acting compression led to higher consonant recognition at low and medium speech levels (45 dB SPL in quiet and 55 to 65 dB SPL in noise), any potential detrimental effects of compression were presumably offset by superior audibility at these levels. At high speech input levels, the initial consonant of the CV stimulus elicits an attack response. Therefore, the release time becomes less critical to how the two compression systems "reshape" the temporal and spectral envelopes. Moreover, at the highest input levels, there were no systematic differences in consonant confusions between fast- and slow-acting compression, which again

supports the conclusion that amplification-induced distortion did not play a role.

It is apparent that, at least in the conditions used in the present study, audibility of speech and the ability to overcome noise-induced forward masking are crucial for recognition. Using a short release time increases the energy of speech segments occurring in the dips of the interrupted noise by restoring gain more quickly. It is, however, unclear whether this would generalize to a broader range of interferers and speech stimuli. Recently, Desloge et al. (2017) proposed a system equalizing the short- and long-term energy of the signal, somewhat similar to instantaneous compression. This provided an increase in audibility of speech segments positioned in the gaps of interrupted and sinusoidally amplitude-modulated noise, which was reflected in a consonant recognition and sentence recognition improvement in HI listeners. Rhebergen et al. (2017) observed a similar benefit of fast-acting compression for sentence recognition when presented in 8-Hz interrupted noise at input SNRs from around -20 to -10 dB. However, for higher SNRs, compression was found to be detrimental, most likely due to reduction of speech peaks. In stationary noise, compression had no effect on recognition for SNRs up to -8 dB and was detrimental at higher SNRs. In all cases, the improvement and deterioration of speech recognition scores coincided, respectively, with an increase and a decrease in the long-term output SNR due to compression. The benefit of compression might also be less apparent in backgrounds that fluctuate but without marked temporal dips or with less coherent across-frequency comodulation (e.g., the “checkerboard noise” of Howard-Jones & Rosen, 1993). In their study of compression parameters, Alexander and Masterson (2015) used speech-shaped noise that was either unmodulated or had a temporal envelope derived from a two-talker mixture (in a manner similar to the International Collegium on Rehabilitative Audiology (ICRA) noise; Dreschler, Verschuure, Ludvigsen, & Westermann, 2001). In each condition, the sentences were mixed with noise at three different SNRs and processed in a compressor, using a 5-ms attack time and a release time of 40 ms (short) or 640 ms (long). With the shorter release time, the output level of speech increased. However, the fast-fluctuating gain amplified not only speech but also some segments of the background. As a result, the output long-term SNR decreased. This might have contributed to the lack of systematic benefit for intelligibility. It is possible that similar effects would have been observed in the present study if an additional masker had been added in the gaps.

Future research should focus on applying the shadow-control method together with the interrupted noise to study the complex interaction between compression, fluctuating noise and reverberation on speech recognition in HI listeners. Reverberation leads to “smearing” of noise

and speech across time. This can potentially increase self-masking and noise-induced masking (Nábělek, Letowski, & Tucker, 1989) and introduce additional envelope distortion that may interact with the temporal alterations occurring due to compression (Shi & Doherty, 2008). Results from Reinhart, Souza, Srinivasan, and Gallun (2016) suggest that both fast-acting compression and reverberation have detrimental effects on speech recognition and that these effects are additive, at least in the absence of background noise. To date, however, very few studies on hearing-aid amplification have considered noise in addition to reverberation. The presence of noise does not only increase the amount of energetic masking but also affects the temporal envelope transfer in a different way than deterministic envelope manipulations (Noordhoek & Drullman, 1997). A combination of noise, reverberation, and compression might lead to complex interactions, affecting speech intelligibility in a way that is difficult to predict. In the context of this study’s experimental setup, with reverberation present, two effects might offset the advantage from increased audibility in the noise gaps. First, the noise energy would “spill over” into the silent gaps. This could have similar effects as an added masker—it would be amplified by the compressor, reducing the long-term output SNR. Second, the temporal smearing would smooth out the internal signal representation that the compressor uses to determine the gain (for details see, e.g., Giannoulis, Massberg, & Reiss, 2012). This could lead to a more sluggish response and obscure the effects of varying time constants, as shown by Reinhart, Zahorik and Souza (2017). Precise control of the acoustical conditions together with the shadow-control approach employed here could help disentangle the previously mentioned phenomena.

Conclusion

A small but systematic benefit of fast-acting compression for consonant recognition was found in both the quiet and the noisy conditions for speech levels below 65 dB SPL (0 dB SNR in noise). Despite potentially detrimental speech envelope distortions, no significant detrimental effects of fast-acting compression were observed for speech recognition when the speech level exceeded the level of the noise. These findings suggest that fast-acting compression provides an audibility benefit and, possibly, an improved recovery from forward masking in fluctuating interferers when compared with slow-acting compression while not substantially compromising the perception of short CV tokens at higher SNRs. It is not yet clear whether these effects persist in more realistic conditions, that is, with longer speech stimuli (multisyllable words, sentences) in fluctuating interferers with softer onsets/offsets. For example, if a voiced interferer was used (such as a competing talker), such short time

constants might lead to detrimental intermodulation distortion offsetting the potential audibility advantage.

The findings from the present study and prospective future studies may inform SNR-dependent amplification strategies and individualized hearing-aid fitting strategies. The potential use of blind SNR estimation for hearing-aid applications has been investigated in May et al. (2017) and applied to dynamically manipulate compression parameters in real time (May, Kowalewski and Dau, 2018).

Declaration of Conflicting Interests

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



Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by funding from Sonova AG (Stäfa, Switzerland) and the Technical University of Denmark.

Note

1. The choice of N2 was determined as the best fit using a “voting” procedure based on the individual audiograms available in the database prior to conducting the experiments. New audiograms, reported here, were measured during the first experimental session, and a deterioration of the hearing loss was observed in some listeners. Therefore, the N3 audiogram (representative of a more severe hearing loss) likely would have been more appropriate, yielding different CRs. However, this could not have been determined until the measurements with all the listeners were completed. The linear gain was calculated during the measurement session after obtaining the new audiogram, and therefore it was fitted more appropriately.

ORCID iD

Borys Kowalewski  <http://orcid.org/0000-0001-6385-6154>
 Ewen N. MacDonald  <http://orcid.org/0000-0002-6973-2926>
 Tobias May  <http://orcid.org/0000-0002-5463-5509>

References

- Alexander, J. M., & Masterson, K. (2015). Effects of WDRC release time and number of channels on output SNR and speech recognition. *Ear and Hearing, 36*(2), e35–e49. 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 the Acoustical Society of America, 142*(2), 908–923. doi:10.1121/1.4997938
- Bisgaard, N., Vlaming, M. S. M. G., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification, 14*(2), 113–120. doi:10.1177/1084713810379609
- 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 the Acoustical Society of America, 141*(6), 4452–4465. doi:10.1121/1.4985186
- Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., & Delhorne, L. A. (2010). Speech reception by listeners with real and simulated hearing impairment: Effects of continuous and interrupted noise. *The Journal of the Acoustical Society of America, 128*(1), 342–359. doi:10.1121/1.3436522
- 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
- Edwards, B. (2002). Signal processing, hearing aid design, and the psychoacoustic Turing test. In: *IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Orlando, Florida, 13–17 May 2002, pp. 3996–3999. IEEE. doi:10.1109/ICASSP.2002.5745533
- Edwards, B. (2004). Hearing aids and hearing impairment. In S. Greenberg, W. A. Ainsworth & A. N. Popper, et al. (Eds), *Speech processing in the auditory system*. Vol. 18, pp. 339–421. Springer. Retrieved from <https://link.springer.com/book/10.1007/b973997>
- Fortune, T. W., Woodruff, B. D., & Preves, D. A. (1994). A new technique for quantifying temporal envelope contrasts. *Ear and Hearing, 15*(1), 93–99.
- 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*(Sup 1): 77–85. doi:10.3109/14992020309074627
- Giannoulis, D., Massberg, M., & Reiss, J. D. (2012). Digital dynamic range compressor design. *Journal of the Audio Engineering Society, 60*(6), 399–408.
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *The Journal of the Acoustical Society of America, 119*(4), 2455–2466. doi:10.1121/1.2171527
- Howard-Jones, P. A., & Rosen, S. (1993). Unmodulated glimpsing in “checkerboard” noise. *The Journal of the Acoustical Society of America, 93*(5), 2915–2922. doi:10.1121/1.405811
- 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*, 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. doi:10.1044/1092-4388(2007)078
- Kates, J. M. (1993). Optimal estimation of hearing-aid compression parameters. *The Journal of the Acoustical Society of America, 94*(1), 1–12. doi:10.1121/1.407078
- 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. R., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research, 1*(1S), 1–3. doi:10.4081/audiores.2011.e24
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). *Package "lmerTest." R Package, version 3.0-1*. Retrieved from <https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf>.
- Lenth, R. (2017). *Package 'lsmeans'. R package, version 2.27-61*. Retrieved from <https://cran.r-project.org/web/packages/lsmeans/lsmeans.pdf>.
- Lunner, T., & Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology, 18*(7), 604–617. doi:10.3766/jaaa.18.7.7
- May, T., Kowalewski, B., & Dau, T. (2018). Signal-to-noise-ratio-aware dynamic range compression in hearing aids. *Trends in Hearing, 22*(1), 1–12. doi:10.1177/2331216518790903.
- May, T., Kowalewski, B., Fereczkowski, M., & MacDonald, E. N. (2017). Assessment of broadband SNR estimation for hearing aid applications. In *IEEE international conference on acoustics, speech and signal processing (ICASSP)*, New Orleans, Louisiana, March 2017. doi:10.1109/ICASSP.2017.7952152
- Nábělek, A. K., Letowski, T. R., & Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *The Journal of the Acoustical Society of America, 86*(4), 1259–1265. doi:10.1121/1.398740
- Noordhoek, I. M., & Drullman, R. (1997). Effect of reducing temporal intensity modulations on sentence intelligibility. *The Journal of the Acoustical Society of America, 101*(1), 498–502. doi:10.1121/1.417993
- Ohlenforst, B., Souza, P. E., & MacDonald, E. N. (2016). Exploring the relationship between working memory, compressor speed, and background noise characteristics. *Ear and Hearing, 37*(2), 137–143. doi:10.1097/AUD.0000000000000240
- Pavlovic, C. (1987). Derivation of primary parameters and procedures for use in speech intelligibility predictions. *The Journal of the Acoustical Society of America, 82*(2), 413–422. doi:10.1121/1.395442
- Plack, C. J., & Oxenham, A. J. (2002). Linear and nonlinear processes in temporal masking. *Acta Acustica united with Acustica, 88*, 348–358.
- Reinhart, P., Zahorik, P., & Souza P. E. (2017). Effects of reverberation, background talker number, and compression release time on signal-to-noise-ratio. *The Journal of the Acoustical Society of America, 142*(1), EL130–135. doi:10.1121/1.4994683.
- Reinhart, P. N., Souza, P. E., Srinivasan, N. K., & Gallun, F. J. (2016). Effects of reverberation and compression on consonant identification in individuals with hearing impairment. *Ear and Hearing, 37*(2), 144–152. doi:10.1097/AUD.0000000000000229
- 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
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London, 336*(1278), 367–373. doi:10.1098/rstb.1992.0070
- Shi, L.-F., & Doherty, K. A. (2008). Subjective and objective effects of fast and slow compression on the perception of reverberant speech in listeners with hearing loss. *Journal of Speech, Language & Hearing Research, 51*(5), 1328–1340. doi:10.1044/1092-4388(2008/07-0196)
- 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., & 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 & Hearing Research, 41*(2), 315. 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.
- Souza, P. E., Wright, R. A., Blackburn, M. C., Tatman, R., & Gallun, F. J. (2015). Individual sensitivity to spectral and temporal cues in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research, 58*, 520–534. doi:10.1044/2015_JSLHR-H-14-0138
- Stone, M. A., & Moore, B. C. J. (2003). Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task. *The Journal of the Acoustical Society of America, 114*(2), 1023–1034. doi:10.1121/1.1784447
- Stone, M. A., & Moore, B. C. J. (2004). Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. *The Journal of the Acoustical Society of America, 116*(4), 2311–2323. doi:10.1121/1.1784447
- Stone, M. A., & Moore, B. C. J. (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *The Journal of the Acoustical Society of America, 121*(3), 1654–1664. doi:10.1121/1.2434754
- Stone, M. A., & Moore, B. C. J. (2008). Effects of spectrotemporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task. *The Journal of the Acoustical Society of America, 123*(2), 1063–1076. doi:10.1121/1.2821969
- 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*. Lyngby, Denmark: Technical University of Denmark.
- Wilson, R. H., McArdle, R., Betancourt, M. B., Herring, K., Lipton, T., & Chisolm, T. H. (2010). Word-recognition performance in interrupted noise by young listeners with normal hearing and older listeners with hearing loss. *Journal of the American Academy of Audiology, 21*(2), 90–109. doi:10.3766/jaaa.21.2.4

- Yund, E. W., & Buckles, K. M. (1995). Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids. *The Journal of the Acoustical Society of America*, 97(2), 1224–1240. doi:10.1121/1.412232
- 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
- Zaar, J., Kowalewski, B., & Dau, T. (2017). Effects of nonstationary noise on consonant identification. In *Proceedings of the International Symposium on Auditory and Audiological Research*, Nyborg, Denmark, August 2017.