DTU Library

The effect of hearing aid dynamic range compression on speech intelligibility in a realistic virtual sound environment

Mansour, Naim; Marschall, Marton; Westermann, Adam; May, Tobias; Dau, Torsten

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
10.1121/10.0008980

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Mansour, N., Marschall, M., Westermann, A., May, T., & Dau, T. (2022). The effect of hearing aid dynamic range compression on speech intelligibility in a realistic virtual sound environment. *Journal of the Acoustical Society of America*, *151*(1), 232-241. https://doi.org/10.1121/10.0008980

JASA ARTICLE

The effect of hearing aid dynamic range compression on speech intelligibility in a realistic virtual sound environment

Naim Mansour,[1,a)] Marton Marschall,[1,b)] Adam Westermann,[2] Tobias May,[1] and Torsten Dau[1,c)]
[1]*Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark*
[2]*Widex A/S, Lynge, Denmark*

ABSTRACT:
Measures of "aided" speech intelligibility (SI) for listeners wearing hearing aids (HAs) are commonly obtained using rather artificial acoustic stimuli and spatial configurations compared to those encountered in everyday complex listening scenarios. In the present study, the effect of hearing aid dynamic range compression (DRC) on SI was investigated in simulated real-world acoustic conditions. A spatialized version of the Danish Hearing In Noise Test was employed inside a loudspeaker-based virtual sound environment to present spatialized target speech in background noise consisting of either spatial recordings of two real-world sound scenarios or quadraphonic, artificial speech-shaped noise (SSN). Unaided performance was compared with results obtained with a basic HA simulator employing fast-acting DRC. Speech reception thresholds (SRTs) with and without DRC were found to be significantly higher in the conditions with real-world background noise than in the condition with artificial SSN. Improvements in SRTs caused by the HA were only significant in conditions with real-world background noise and were related to differences in the output signal-to-noise ratio of the HA signal processing between the real-world versus artificial conditions. The results may be valuable for the design, development, and evaluation of HA signal processing strategies in realistic, but controlled, acoustic settings.
 https://doi.org/10.1121/10.0008980

(Received 28 February 2021; revised 8 November 2021; accepted 20 November 2021; published online 13 January 2022)
[Editor: Joshua G Bernstein] Pages: 232–241

I. INTRODUCTION

Hearing aids (HAs) attempt to restore hearing-impaired (HI) people's ability to reliably explore their auditory world. The usage of modern digital HAs has been shown to improve a wearer's hearing ability in complex real-world environments (Noble and Gatehouse, 2006). However, the introduction of more sophisticated HA signal processing algorithms over the past decades has not led to a substantial increase in HA user satisfaction (Kochkin, 2002). While a satisfaction rating depends on many factors, such as ease of use and wearing comfort, improving speech intelligibility (SI) in noise remains one of the core purposes of a HA. However, HAs have failed to provide a consistent SI benefit across users (Kochkin, 2002). This may be partially due to the focus of current HA fitting procedures on restoring audibility, rather than addressing supra-threshold distortions which HI listeners commonly experience when listening to speech in noisy situations [e.g., Sanchez-Lopez *et al.* (2019)]. In addition, signal processing algorithms in HAs have mostly been optimized for SI using speech-recognition-in-noise metrics, such as speech reception thresholds (SRTs), obtained with artificial acoustic stimuli, which may not correlate well with HA satisfaction in the listeners' real-world experience (Bentler *et al.*, 1993; Cord *et al.*, 2007; Wu, 2010). Therefore, it may be worthwhile to explore SI in more realistic, ecologically valid ways, both in unaided conditions as well as in conditions aided by the HA.

Various studies have investigated the impact of HA processing on SI, widely varying in scope and methodology. A common approach has been to combine speech-shaped noise (SSN) or some type of babble noise as a masker with anechoic speech sentences as the target, both presented over headphones, with the recordings pre-processed to simulate the effect of HA amplification (Hunt *et al.*, 2019; Jirsa and Norris, 1982; Saunders and Kates, 1997; Souza *et al.*, 2015). Reverberant properties of both the background noise and the target speech, as well as effects of spatial source separation have often not been considered. In addition, head movements have largely been ignored in both the static playback of the stimuli and the HA processing. A few studies focused on the realism of the acoustic conditions and presented the noise and target speech stimuli over a spatially distributed set of loudspeakers, allowing for the use of physical HAs, either as a fitted commercial HA (Köbler and Rosenhall, 2002; Moore *et al.*, 1985; Wouters *et al.*, 1999), the participant's own HA (Best *et al.*, 2015), or a fully-controlled, real-time "master" HA (Hendrikse *et al.*, 2020; Oreinos and Buchholz, 2016; Seewald *et al.*, 1981). In most of these studies, the small number of loudspeakers and the involved

a)Electronic mail: naiman@dtu.dk, ORCID: 0000-0001-5673-6840.
b)ORCID: 0000-0003-2534-7062.
c)ORCID: 0000-0001-8110-4343.

playback methods did not allow for a realistic reproduction of a real-world spatial sound field, whereas Best *et al.* (2015) and Oreinos and Buchholz (2016) employed so-called virtual sound environments (VSEs), using a large number of loudspeakers to accurately reproduce real-world environments. Best *et al.* (2015) used a parametric room acoustic simulation technique to reproduce the spatialized sound field, while Oreinos and Buchholz (2016) employed recordings made in a room populated with simulated talkers, using higher-order Ambisonics (Daniel, 2000). Both studies still relied on simulated interferers in the noise masker, constructed using a number of anechoic speech samples placed in a room or room model, rather than *in situ* recordings of real scenes. In addition, neither study included an instrumental analysis to investigate the effect of the HA on its input signal, instead focusing entirely on psychoacoustic outcome measures.

In the present study, the effects of HA processing—specifically dynamic range compression (DRC)—on speech intelligibility were evaluated relative to unaided listening in realistic acoustic conditions representing and based on recordings of real-world environments. Best *et al.* (2015) found that HAs provided a greater benefit to SI in their simulated VSEs compared to more artificial masker types for the majority of listeners. Here, the effect of HA processing was evaluated inside recorded VSE using Ambisonic auralizations of spatial recordings (Mansour *et al.*, 2019). Two real-world scenes were recorded with a spherical microphone array using higher-order Ambisonics and reproduced inside a 64-loudspeaker loudspeaker array, providing the background-noise masker for the VSE. An artificial, quadraphonic SSN stimulus matched to the long-term average spectrum of one of the real-world scenes was used as a reference condition. Target sentences from the Danish Hearing in Noise Test (HINT) (Nielsen and Dau, 2011) were convolved with a room impulse response (RIR) recorded in the respective real-world scenes, obtaining spatialized target speech material for the SI task. A master HA was used to ensure consistency in HA processing across listeners and to allow for an instrumental analysis of HA benefit that could potentially explain aided SI improvements. Ten listeners with symmetric mild-to-moderate hearing loss carried out the adaptive SI task wearing two HA shells, i.e., simplified behind-the-ear HAs, controlled by the master HA, which implemented DRC based on the NAL-NL2 fitting rationale (Keidser *et al.*, 2011). The listeners also completed the SI task using an "unaided" reference strategy, i.e., without wearing a HA. It was assumed that, similarly to the results in Best *et al.* (2015), the real-world conditions would lead to decreased SI performance compared to the artificial reference condition, both with and without the help of a HA, and that the HA would render a different benefit in the real-world conditions compared to the artificial condition. Finally, an instrumental HA analysis was conducted to attempt to relate the HA's input-output signal-to-noise ratio (SNR) performance to the observed SI results.

II. METHODS

A. Virtual sound environment and spatial noise maskers

The Audiovisual Immersion Lab (AVIL) at the Technical University of Denmark, comprised of a fully spherical loudspeaker array mounted inside an anechoic chamber, provided the VSE which was used to play back the masker and target speech stimuli. The array consists of 64 loudspeakers (KEF LS50, KEF Audio, United Kingdom) at a distance of 2.4 m to a chair positioned in the center, with 24 loudspeakers separated by 15° in the horizontal ring (at 0° elevation).

Three spatial background-noise conditions were considered in the experiment. The first condition was based on a one-minute-long spatial recording of a real-world office meeting, obtained with a 32-channel microphone array (em32 Eigenmike, mh acoustics LLC, USA). In the office meeting scenario, 12 normal-hearing participants conversed in pairs around a square conference table (2.4 m long) inside a conference room. The microphone array was placed at head level in one of the seats around the table. Fourth-order Ambisonic encoding and decoding steps were applied to the recording to match the geometry of the loudspeaker array for playback. Ambisonic sound field processing was chosen because this technique can reproduce a recorded spatial sound field with physical accuracy by mapping microphone array recordings to spherical harmonic basis functions. The resulting signal was calibrated to the average broadband sound pressure level (SPL) of 73.5 dB observed in the original recording using the left ear of a head-and-torso simulator (HATS, B&K type 4128, Brüel & Kjær A/S, Denmark) placed on the chair in the center of the loudspeaker array, where the Ambisonic sound field was most accurately reconstructed (i.e., the sweet spot).

The second background-noise condition was constructed using a one-minute-long spatial recording of a public lunch. In this scenario, the 12 participants were seated around a rectangular lunch table (1 m in diameter) inside a large corporate restaurant. All recording and processing steps were the same as for the first masker, with the final signal calibrated to 75.5 dB SPL as the average level measured during recording.

The third background-noise condition included SSN, matched to the long-term average spectrum (LTAS) of the public-lunch recording. The SSN maskers were obtained by first recording the 64-channel public-lunch masker at the left and right ear of the HATS placed in the center of the array. Then, the LTAS of the left- and the right-ear recordings were computed separately using frames of 64 ms, Hann-windowing with 50% overlap, and smoothed over 1/3rd octave bands using a normalized Gaussian kernel. A white noise signal was then filtered using a linear-phase finite impulse response filter (FIR) matched to the magnitude spectrum of the LTAS. The resulting SSN signals were bandpass filtered between 20 Hz and 20 kHz using a 4th order Butterworth filter. Uncorrelated versions of the SSN masker derived from the left LTAS were played back over two loudspeakers in the horizontal ring at 45° and 135° azimuth, while the right, uncorrelated SSN maskers were played over the loudspeakers at 225° and 315° azimuth, creating a quadraphonic loudspeaker setup. The final, 4-channel SSN masker was calibrated in the center of the array to the same broadband level of 75.5 dB SPL as in case of the public-lunch masker.

B. Listeners

Ten hearing-impaired (HI) listeners, nine male and one female, participated in the experiment. They were between 62 and 84 years old with a median age of 71.5. All listeners had a symmetric sensorineural hearing loss not exceeding an N3/S1 hearing loss category (Bisgaard *et al.*, 2010) and showed word-discrimination scores higher than 90%. The listeners were seated in a chair in the center of the AVIL loudspeaker array and their SI was evaluated for the unaided hearing strategy and the aided hearing strategy detailed below, in each of the three background-noise conditions. The resulting six blocks were randomized across listeners according to a balanced 6-by-6 Latin square design (Bradley, 1958). The SI scoring was carried out by a Danish audiologist, while the master HA was monitored to ensure its proper processing of the input signal without delay or feedback. All participants provided informed consent and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

C. Speech intelligibility task

Using the spatialized noise maskers, an SI task was designed based on the HINT (Nielsen and Dau, 2011). The anechoic target speech sentences, on average 1.5 s long, were convolved with RIRs measured in the office meeting and the public-lunch scenario between a loudspeaker and the microphone array placed on opposite sides of the table. The direct-to-reverberant ratios (DRR) in the office meeting and the public-lunch scenario were 6.6 and 16.6 dB, respectively. The sentences were then calibrated at the left ear of the HATS placed in the center of the loudspeaker array. SRTs were determined using an adaptive, 1-up-1-down procedure, presenting the noise maskers at their constant broadband levels of 73.5 and 75.5 dB SPL for the office meeting and public-lunch recordings, and varying the level of the spatialized target sentences. The artificial SSN condition presented the noise masker at the same broadband level of 75.5 dB SPL as the public-lunch masker from which it was derived and utilized the target sentences that were convolved with the public-lunch RIR. This way, the only difference between the artificial SSN condition and the realistic noise conditions was the nature of the background noise.

D. Real-time hearing aid signal processing

To ensure that the HA signal processing operated consistently across listeners, a fully controlled, real-time HA system was implemented. The two HA shells (Signia, WSAudiology, Germany), each containing a microphone and a receiver, were connected *via* a custom preamplifier box and a sound card (RME Fireface 800, Audio AG, Germany) to a laptop. The HA signals were processed by the openMasterHearingAid framework (openMHA) (Herzke *et al.*, 2017), encapsulated in a MATLAB control layer.

1. HA algorithms

The basic building blocks of the HA signal processing chain consisted of input- and output-level equalization with clipping protection, as well as a filter bank decomposition with windowing and gain application. The input signal was sampled at a rate of 44.1 kHz in time windows of 64 samples. An 65-tap finite impulse response (FIR) input-equalization filter was applied to flatten the calibrated HA microphones' frequency response, while an output FIR filter was necessary to ensure that the frequency-dependent effect of the ear canal was compensated for in the calibrated HA receivers. The filter bank used an overlap-add strategy to process frames decomposed by a fast Fourier transform at a length of 512 time-domain samples, windowed by a 256-sample Hanning window with 50% overlap. Each frame was decomposed in the frequency domain into nine rectangular, 3/4-octave-wide, non-overlapping bands with the lowest center frequency at 177 Hz (177, 297, 500, 841, 1414, 2378, 4000, 6727, and 11 314 Hz).

The HA amplification employed DRC based on the listener's pure-tone audiogram thresholds and following the gain prescription of the National Acoustics Laboratory-Non Linear 2 (NAL-NL2) fitting rationale (Keidser *et al.*, 2011) (see the Appendix A for more details). The attack and release time constants for the DRC were set to 5 and 100 ms, respectively. These values correspond to a fast-acting compression scheme, where the output gain is adjusted relatively quickly after a change in input level, both at the onset and end of the level change.[1] Such a configuration aims at restoring audibility on short time scales corresponding to syllables or phonemes.

Other features, such as feedback cancellation, noise reduction and beamforming, common in commercially available hearing aids, were not included in the HA processing. This was done to focus on the essential components in the processing and exclude potentially confounding adaptive algorithms. The omission of feedback cancellation implied that the HAs needed to be equipped with a fully closed dome and were limited in their gain prescription to at most a N3 or S1 hearing loss category (Bisgaard *et al.*, 2010), or about a 40 dB hearing loss at 1 kHz.

2. HA implementation, fitting and evaluation

The HA processing chain was implemented in the openMHA framework. This plugin-based open source platform includes several basic HA features as well as HA calibration and validation tools and can interface in real-time

https://doi.org/10.1121/10.0008980

with input-output sound card channels using a desired sample rate and frame size (see Appendix C for more details).

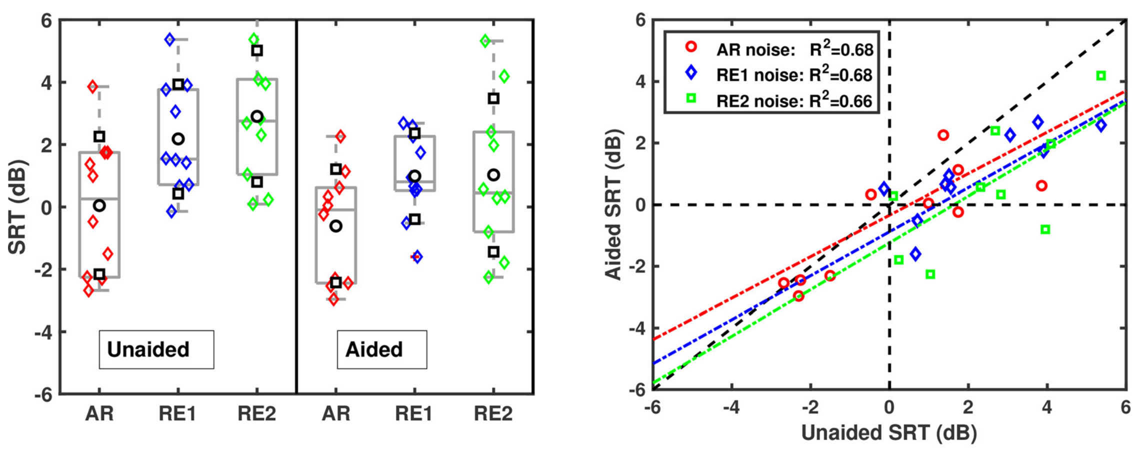
After the calibration of the HA microphones and receivers (see Appendix C for more details), a listener-specific validation routine was carried out, verifying that the HA receiver output levels matched the target gains of the NAL-NL2 rationale at the 65 dB SPL DRC knee point at each filterbank center frequency. The HAs were placed on the HATS, positioned on the chair in the center of the loudspeaker array and the HA receiver calibration values were fine-tuned until the measured gains at the HATS microphones deviated by less than 1 dB from the target gains. The processing delay between the sound card microphone input channels and receiver output channels was determined by feeding a test signal into each of the input channels and tracking the time it took for the signal to be processed by an openMHA instance and reach a receiver output channel. On average, the input-output delay amounted to 11 ms, which is more than the 5 ms delay commonly targeted in commercial HAs yet still considerable lower than the 20 ms limit tolerable to individuals with mild-to-moderate hearing loss (Stone and Moore, 1999).

E. Instrumental HA analysis

In addition to the SI assessment, an instrumental HA analysis was conducted to evaluate how the HA DRC processing affected the broadband SNR of target HINT sentences in the three noise conditions. With the HAs placed on the HATS in the center of the loudspeaker array, HA microphone recordings were made of 20-s excerpts of the two real-world background noises and the SSN as well as ten randomly selected HINT target sentences, spatialized in the same way as for the SI evaluation. Each sentence was superimposed onto 10 different noise clips within each background-noise excerpt, which was set to its respective broadband level. The sentences were scaled to achieve the desired SNR. Using the method proposed by Hagerman and Olofsson (2004), this process was then repeated while all noise clips were shifted in phase by 180°. For each speech-in-noise mixture, the in-phase and out-of-phase versions were then processed by an offline, file-based openMHA instance, configured to provide amplification with DRC according to the NAL-NL2 rationale fitted to the mean audiogram across all listeners. By adding or subtracting the output in-phase and out-of-phase mixtures, the separate speech and noise components can be recovered perfectly, assuming that the HA has a linear phase response. In this way, 100 speech-in-noise clips were analyzed per target SNR value, across a range from −12 to 12 dB (the approximate range of SNRs presented in the SI task), in 3-dB increments.

III. RESULTS

Figure 1(a) shows the measured SRTs across listeners for the unaided and the aided hearing strategies. Individual thresholds are plotted for the artificial SSN (AR, red circles), the real-world office meeting noise (RE1, blue diamonds), and public-lunch noise (RE2, green squares), as well as the median and 25th/75th percentiles of the SRT distributions (box plots) and the maximum/minimum values (whiskers). The mean values (black circles) and standard deviations (black squares) are also shown. A large variability of the SRTs across listeners can be observed in all conditions and for both processing strategies. For the unaided strategy, mean SRTs were 0.1 dB in the AR noise, 2.2 dB in the RE1 noise, and 2.8 dB in the RE2 noise. Similarly for the aided strategy, the mean SRTs were −0.6 dB in the AR noise, 1.0 dB in the RE1 noise, and 1.0 dB in the RE2 noise. The standard deviations for the unaided strategy were 2.2 dB in the AR condition, 1.8 dB in the RE1 condition, and 2.1 dB in the RE2 condition. For the aided strategy, the standard



(A) Speech reception thresholds (B) Unaided vs. aided SRTs

FIG. 1. (Color online) (A) Measured speech reception thresholds (SRTs) for the unaided and the aided hearing strategies, in the SSN (AR), the real-world office meeting noise (RE1), and public-lunch noise (RE2) conditions. The mean values (black circles) and standard deviations (black squares) are also shown. (B) Unaided vs aided SRTs and R^2 correlation factors in the three noise conditions; AR, RE1, and RE2. The least squares fits to each noise conditions are shown as dashed lines.

deviations were 1.8, 1.4, and 2.5 dB in the AR condition, RE1 condition, and RE2 condition, respectively.

A two-way repeated-measures analysis of variance revealed a significant effect of noise type [$F(1,9) = 23.2$, $p = 0.0001$] and hearing strategy [$F(1,9) = 47.7$, $p \leq 0.0001$]. There was no significant interaction between noise type and hearing strategy [$F(1,9) = 0.1, p = 0.78$]. However, to investigate the specific hypotheses on the effect of the noise type on the SI performance and on the potential HA benefit, pair-wise comparisons with a Bonferroni correction for three tests ($p = 0.167$) were carried out. Both for the unaided and the aided strategy, SRTs were significantly higher for the RE1 ($p = 0.001$ and $p = 0.006$) and RE2 condition ($p = 0.0002$ and 0.015) than for the AR condition. Comparisons between the RE1 and RE2 conditions did not reveal a significant difference ($p = 0.016$ and 0.94). Thus, the realistic noises made it consistently more difficult to understand speech compared to the artificial noise, regardless of whether HAs were used or not. Furthermore, there was a significant decrease in SRTs between the unaided and the aided hearing strategy for the RE1 ($p = 0.004$) and RE2 ($p = 0.003$) conditions, but no significant difference for the AR condition ($p = 0.13$). Consequently, the artificial noise did not show a significant HA benefit, while both realistic noise types did.

Figure 1(b) illustrates the unaided versus aided SRTs for each of the three noise conditions AR, RE1, and RE2, as well as their linear least squares fits, with respective R^2 correlation factors of 0.68, 0.68, and 0.66. This indicates that the aided strategy was strongly correlated with the unaided strategy across all noise types. Most data points lie below the 45° line and all least squares fits have slopes below one, demonstrating the increasingly beneficial effect of the HA DRC processing on SI with increasing values for the SRT (i.e., decreasing SI) in the unaided condition.

Table I shows the R^2 correlation factors between the listeners' SRTs and their the four-frequency-average hearing loss (4FAHL). Correlations were weak across all noise types and weakest for the realistic noise types, and there was virtually no difference between the unaided and the aided hearing strategy. Thus, the 4FAHL was a poor predictor of SI performance in the conditions considered in the present study, especially in realistic VSEs, regardless of whether a HA DRC processing strategy was active or not.

Figure 2 shows the output SNR distributions for the AR, RE1 and RE2 conditions at nine input SNRs between −12 and 12 dB, obtained from the instrumental HA analysis.

TABLE I. Unaided and aided R^2 correlation factors between the average four-frequency-average hearing loss (4FAHL) across participants and their speech reception thresholds for the three noise types.

R^2	Unaided	Aided
AR	0.27	0.26
RE1	0.15	0.19
RE2	0.09	0.07

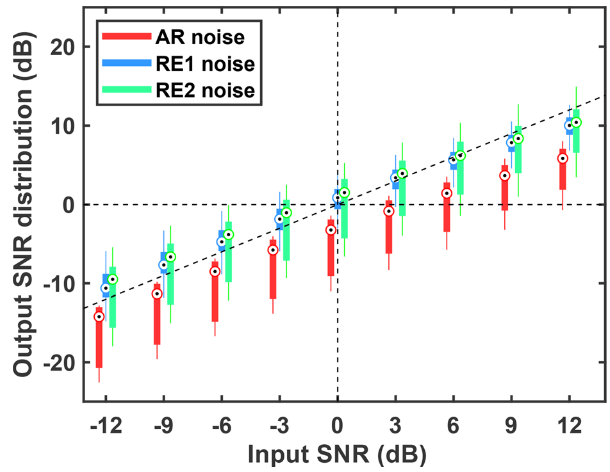


FIG. 2. (Color online) Output SNR distributions vs input SNR values resulting from the instrumental HA analysis for 10 target speech sentences superimposed onto 10 noise fragments at nine input SNRs between −12 dB and 12 dB, for the artificial noise type (AR) and the two realistic noise types (RE1 and RE2).

The AR noise produced considerably lower SNR values (red symbols) than the RE1 (blue) and RE2 conditions (green), for all input SNRs. For the RE1 and RE2 conditions, the HA DRC processing improved the output SNR of the provided speech-in-noise mixture by up to 2 dB at negative input SNRs. The effect decreased with increasing input SNR and at input SNRs beyond around 5 dB, the output SNR became smaller than the input SNR. For the AR condition, the HA DRC processing degraded the output SNR regardless of the input SNR, but again most strongly at positive input SNRs. The HA processing in the AR condition started to provide a positive median output SNR at an input SNR of 4.1 dB, while this happened at input SNRs of –1 and –1.8 dB, respectively, for the RE1 and RE2 conditions. The spread in the SNR output distribution was markedly larger in the AR and RE2 conditions relative to that in the RE1 condition.

Figure 3(a) displays the histograms of the noise levels (in dB) estimated for a ten-second excerpt of the AR (red), RE1 (blue), and RE2 (green) interferers as recorded by the left front HA microphone and normalized to unit average power. Each level estimate was obtained by calculating the average power over a time window of 5 ms, corresponding to the analysis window of the HA DRC processing. The histogram bin width of 1 dB corresponds to the resolution of the compression lookup table inside the HA. These histograms show that the RE1 and RE2 noise types exhibit considerably greater amplitude fluctuations over time than the stationary AR noise. Similarly, Fig. 3(b) shows the corresponding histograms of the speech levels across the ten HINT sentences used in the instrumental SNR analysis, calculated in the same way as in the case of the noise histograms. The histogram for the anechoic speech is shown in black, together with the histogram for the RE1 (in blue) and the AR/RE2 (in green) speech, as recorded by the left front HA microphone and normalized to unit average power. The histograms obtained for the anechoic and the AR/RE2

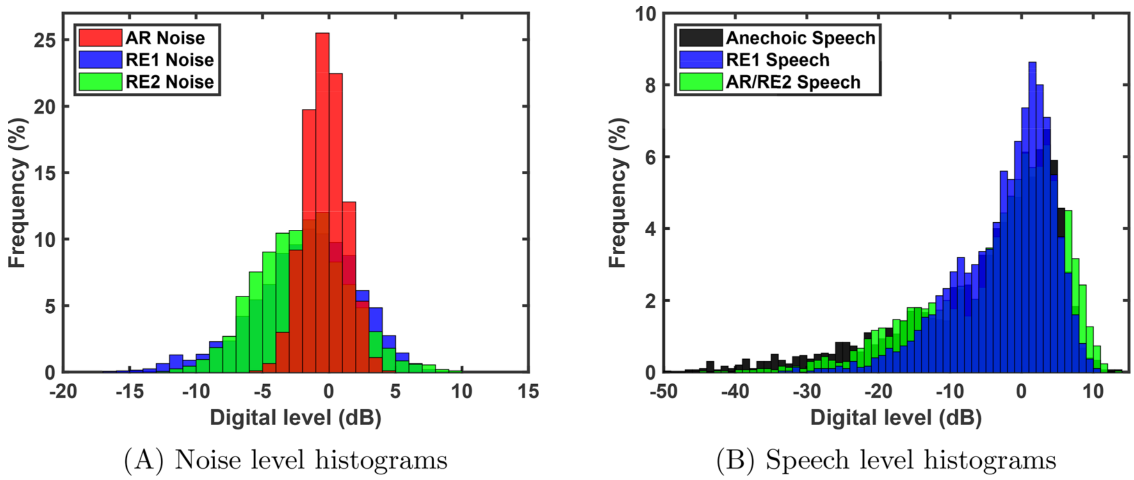


FIG. 3. (Color online) (A) Histograms of the noise levels (in dB) estimated for a ten-second segment of the AR (red), RE1 (blue), and RE2 (green) interferers, normalized to unit average power. Each level estimate was obtained by calculating the average power over a time window of 5 ms. (B) Corresponding histograms of the speech levels across the ten HINT sentences used in the instrumental SNR analysis, calculated in the same way as in the case of the noise histograms. The anechoic speech histogram is shown (black), alongside the RE1 (blue) and AR/RE2 (green) speech histograms, normalized to unit average power. The analysis window was the same as for the noise signals.

speech were similar in width whereas the RE1 histogram showed a narrower distribution. This was expected due to the much lower DRR in the office meeting scenario (RE1, 6.6 dB), introducing more reverberation into the anechoic HINT sentences than in the public-lunch scenario (RE2, 16.6 dB). The more reverberant RIR of the office meeting reduced the level fluctuations in the target speech signals, thereby also reducing the width of the RE1 SNR distribution (see Fig. 2) compared to the AR and RE2 SNR distributions.

Last, Table II displays values of the extended short-time objective intelligibility index (ESTOI) (Jensen and Taal, 2016) at the output of the HA processing across a range of input SNRs, for the same clean-speech and noisy-speech-mixture stimuli as used in the SNR instrumental analysis. The ESTOI metric was selected since it has been shown to predict intelligibility results well for both highly modulated and stationary noise types (Jensen and Taal, 2016). The results saturate for all conditions at the extremely negative and extremely positive SNRs of –30 and 30 dB, respectively. In the range from –6 to 6 dB SNR, the RE1 condition exhibits the highest ESTOIs, followed by the RE2 condition and then the AR condition. This indicates that the target speech was more intelligible, according to the model, in the RE1 and RE2 noise types than in the AR noise type.

TABLE II. Extended short-time objective intelligibility index (ESTOI) values at the output of the HA processing across a range of input SNRs, for the same clean-speech and noisy-speech-mixture stimuli as used in the SNR instrumental analysis.

	SNR (dB)						
ESTOI	−30	−6	−3	0	3	6	30
AR	0	0.18	0.26	0.33	0.42	0.50	0.98
RE1	0	0.29	0.38	0.47	0.56	0.65	0.99
RE2	0	0.23	0.31	0.39	0.48	0.57	0.97

IV. DISCUSSION

The main aim of this study was to investigate the impact of applying *in situ* background-noise recordings as maskers in a speech intelligibility task conducted with and without hearing aids. Consistently with previous work (Best *et al.*, 2015; Mansour *et al.*, 2019), employing realistic noise maskers resulted in increased SRTs compared to artificial noise maskers. This reduction in SI in realistic backgrounds containing a mixture of interfering talkers has been linked to energetic speech-on-speech masking of the interferers on the target speech in the spectral and modulation-spectral domains (Brungart *et al.*, 2006; Jørgensen and Dau, 2011) as well as informational masking caused by the intelligibility of the interferers (Westermann and Buchholz, 2015). In the present study, the LTAS of the AR noise was matched to that of the RE2 noise, such that the differences in the SRTs obtained in these conditions could not be caused by differences in the (long-term) spectral properties of these interferers. At the same time, SRTs did not differ between the RE1 and RE2 conditions. Despite room acoustic and spectral differences between the two realistic scenes, the stimuli were presented at rather similar broadband SPLs (73.5 and 75.5 dB) and contained the same number of nearby interfering talkers (10). This suggests that SI in realistic VSEs might be more strongly influenced by the overall background level and associated SNR as well as the number of interfering talkers, which have been found to be predictors of perceived scene complexity (Weisser *et al.*, 2019), than by the acoustic details of the background noise. However, while the RE1 and RE2 conditions resulted in similar SRTs, this may not necessarily be the case for other real-world scenarios with different room acoustic and spatial properties. The differences in results obtained across noise types between the aided and unaided strategies indicate that going from the artificial noise to the realistic noises degraded speech intelligibility, both with and without the help of a HA.

The applied HA processing led to improved SI in the realistic conditions, but not in the AR condition. Several factors could have contributed to this finding. First, hearing-impaired listeners that have their audibility (partially) restored through HA amplification may regain the ability to employ dip listening, which may be more beneficial in modulated interferers (such as the realistic noises in the present study) than in stationary noise interferers (Peters *et al.*, 1998).

Second, as revealed by the instrumental analysis, the HA DRC processing affected the different noise types differently. Due to the stationarity of the AR noise type, the applied compression resulted in a degradation in the output SNR of the HA across the entire input SNR range (see Fig. 2). In contrast, due to the wider dynamic range in the RE1 and RE2 noise types, the compressive processing was able to reduce the peak noise energy to a greater degree (at the same overall SPL), resulting in an improvement in the median output SNR at low input SNRs (up to about 3 dB). At highly positive SNRs, this effect disappeared (as can be seen in Fig. 2, above around 6 dB) because the speech signal, shared between all noise types, now determined the envelope of the mixture and thereby the impact of compression (Rhebergen *et al.*, 2009). Unaided SRTs in the RE1 and RE2 conditions, however, fell in the range where the HA DRC processing still resulted in an SNR improvement for most segments [roughly between 0 and 5 dB, see the ordinate of Fig. 1(b)], which likely contributed to the HA benefit observed in these conditions. Previous work (Boike and Souza, 2000; Moore *et al.*, 1999) also demonstrated the benefit of fast-acting DRC on the SNR of noise-dominated speech in modulated (artificial) maskers versus stationary maskers, as well as its beneficial effect on SI (Kowalewski *et al.*, 2018; Rhebergen *et al.*, 2009). A study by Naylor and Johannesson (2009) that included an instrumental analysis to investigate HA compression efficacy reported similarly sloped SNR input-output curves, further underscoring the effect of the noise fluctuations on the HA performance. The fact that the aided SRTs in the AR condition were not worse than the corresponding unaided SRTs may be explained by frequency-dependent audibility improvements provided by the HA, despite the reduction in broadband SNR. Overall, the presence or lack of benefit of compressive HA processing on SI, both in terms of instrumental and perceptual measures, depends strongly on the range within which SRTs are obtained, which, in turn, depends on the properties of the speech and noise stimuli presented to the HA and the listener. Specifically, HA benefit may have been greater for the artificial noise type at higher SNRs, and reduced for the realistic noise types at lower SNRs. It is therefore important to utilize realistic stimuli in aided SI paradigms, to ensure the HA will operate in a real-world SNR range.

Last, the ESTOI analysis revealed that predicted speech intelligibility in noise at the output of the HA was greater for the RE1 and RE2 noise types than for the AR noise type in the SNR range where unaided SRTs where obtained. This is in agreement with the input-output SNR analysis as well as the perceptually obtained SRTs.

The results of the present study suggest that a realistic VSE can provide an effective setting for evaluating the impact of HA signal processing algorithms on a listener's SI performance and for relating that performance to instrumental HA performance metrics. The present approach, however, did not consider visual information, which is known to greatly improve speech intelligibility and subsequently shift unaided SRTs ranges downward, especially at low SNRs (Sumby and Pollack, 1954). It is likely that the lack of visual cues also affected the SRTs of the listeners in the current experiment, which in turn may alter the aided benefit shown by the HA processing. Thus, an important next step would be to establish measurements of visually aided speech intelligibility in realistic VSEs. The results of the instrumental analysis showed that, with a shift to lower SNRs, the HA DRC processing continued to improve the output SNR in the realistic scenes, implying that a visually aided SI assessment, converging to lower SNRs, would continue to reveal an increased HA benefit for realistic versus artificial scenes. In addition, the interplay between evaluated SNR, hearing loss and HA performance makes comparing SI performance across listeners using only SRT measurements problematic (Naylor, 2016). Including a constant-SNR aided SI assessment in realistic VSEs would provide a measure of realistic HA performance at a constant operating point. Furthermore, only a simple HA was implemented in this study, excluding commonly used features such as noise reduction and beamforming, even though a HA simulator with any number of features could be used. This was done to be able to relate the impact of the different background-noise types on DRC to SI performance. As such, it is possible and even likely that a more advanced HA may render different performance benefits. This underscores the importance of evaluating HAs in realistic conditions, such as the ones suggested here, in order to faithfully capture their benefits. Last, it may be informative to consider psychoacoustic outcome measures beyond SI, such as localization or loudness perception in realistic VSEs. Combining aided psychoacoustic measures with instrumental HA analyses may provide a better understanding of how various signal processing strategies perform in real-life environments.

V. CONCLUSION

In this study, the effect of HA DRC processing on SI was investigated in two realistic acoustic scenes, constructed using spatial background recordings and anechoic speech samples convolved with RIRs measured *in situ*. It was found that both unaided and aided SRTs were significantly increased inside both realistic VSEs compared to a reference condition employing quadraphonic SSN as a masker. This is consistent with previous studies, which showed increased spectral and modulation-spectral energetic masking for realistic noises as well as a detrimental effect of intelligible speech in the maskers on SI (Best *et al.*, 2015; Mansour *et al.*, 2019). Despite the acoustic differences, SI was similar between the two realistic noise types, suggesting that precise acoustic details in the reproduced environment may be less important than its overall loudness and number of interfering sources.

HA DRC processing was found to provide a benefit to SI in the realistic scenes, but not in the artificial reference condition, which was consistent with output SNR differences observed between the realistic versus reference conditions. Results of an instrumental SNR analysis revealed that the stationary nature of the artificial noise led to consistently lower median SNRs at the output of the HA compared to the two realistic background noises.

The results of this study illustrate the relevance of evaluating the impact of HAs on SI in experimental conditions that are realistic, such that the SI task might produce SRTs that reflect real-life experience. By achieving this, HA processing strategies may then be tailored to SNRs that occur in real-world conditions and consequently are better matched to the user's every-day experiences.

ACKNOWLEDGMENTS

The research was supported by the Centre for Applied Hearing Research (CAHR) at the Technical University of Denmark and by Widex A/S. We would like to thank Rikke Sørensen for screening and scoring the participating listeners.

APPENDIX A: NAL-NL2 FITTING RATIONALE

The NAL-NL2 rationale attempts to restore speech intelligibility of a HA user while preserving comfortable normal-hearing loudness levels. This is achieved using a level- and frequency-dependent fitting formula, that specifies real-ear insertion gain (REIG) factors to be applied to the input signal at 50, 65, and 80 dB across 19 frequencies between 125 and 8 kHz, spaced 1/3rd of an octave apart. Linear level interpolation is applied in the 50–65 dB and 65–80 dB ranges, while linear gain is applied below 50 dB and above 80 dB. Further linear frequency interpolation is applied to match the center frequencies of the HA filter bank. The fitting formula is defined for a range of ten standard audiograms (Bisgaard *et al.*, 2010), from which the one is selected that minimizes the sum of absolute differences with the listener's available audiogram frequencies.

APPENDIX B: OPENMHA PROGRAMMING

The openMHA framework uses a custom configuration language allowing for line-by-line human-readable text commands to be inserted in a configuration file in order to build a HA signal processing chain. A basic code sample is shown in Fig. 4. Elements of the microphone and receiver level equalization stages are shown, as well as excerpts from the algorithm chain, the filter bank settings and the input-output channels.

```
mha.plugin_name = mhachain
...
mha.calib_in.peaklevel = [111.3537 112.166 111.0489 112.1515]
mha.calib_in.fir   = [[-8.2451e-05 -0.00013246 -0.00018752 ...]
...
mha.calib_out.peaklevel = [112.5364 110.1078]
mha.calib_out.fir = [[-0.00019513 -0.00018781 -0.00025777 ...]
...
mha.mhachain.overlapadd.mhachain.algos = ...
                        [route:left_in acSteer:mvdr steerbf:left ...
                         route:right_in steerbf:right route:out ...
                         fftfilterbank dc combinechannels]
...
mha.mhachain.overlapadd.mhachain.fftfilterbank.unit = Hz
mha.mhachain.overlapadd.mhachain.fftfilterbank.f = ....
             [177 297 500 841 1414 2378 4000 6727 11314]
...
io.con_in = [system:capture_1 system:capture_2 ...
                system:capture_3 system:capture_4]
io.con_out = [system:playback_3 system:playback_4]
```

FIG. 4. Excerpts from an openMHA configuration file.

All of the used processing algorithms were sourced from available openMHA plugins. To control the basic operations of toggling the HA processing and loading appropriate configuration files, openMHA uses Java interface libraries that can be invoked from encapsulating MATLAB functions. This allowed for a homogeneous implementation of the HA processing alongside the MATLAB-based spatial sound processing programming. In addition, the openMHA framework provides a graphical user interface for deriving appropriate configuration file values based on an input audiogram and a desired fitting rationale. The NAL-NL2 rationale was implemented specifically for this experiment.

APPENDIX C: HA CALIBRATION

The microphones and receivers in the master HA were calibrated to ensure proper operation with the listeners. To calibrate the four HA microphones for an equal level across frequencies, the HAs were placed in an portable anechoic enclosure inside of which third-octave band white noise bursts around the HA filter bank center frequencies were played at a level of 80 dB SPL. The obtained digital levels for each of the microphones were then transformed to the appropriate openMHA configuration peak values and FIR filters using openMHA-provided scripts. The validation of the microphone calibration consisted of rerunning the calibration procedure and verifying the correct 80 dB SPL values across the considered frequencies.

The HA receiver calibration ensured that the prescribed gain of the fitting rationale accounted for the natural frequency dependency of the ear canal, resulting in the REIG. By definition, the REIG is equal to the difference between the real-ear aided gain and the real-ear unaided gain (REUG), referring to the gain at the eardrum compared to a reference point at the canal entrance, with the HA inserted into the ear canal and turned on or not inserted, respectively. Here, a single set of general REUG values was measured for use across all listeners using the ear canals of the HATS, which was placed in the center of the loudspeaker array.

Specifically, 1/3rd octave band noise around the HA filter bank center frequencies was played from the frontal horizontal loudspeaker in the array, calibrated to 80 dB SPL by a calibrated reference microphone placed at the entrance to the HATS' left and right ear canals. Simultaneously and without the HAs inserted, the levels at the calibrated ear drum HATS microphones were recorded. Subtraction of the ear canal entrance levels from the ear drum levels resulted in the left and right REUGs. The HA receivers were calibrated by repeating this procedure with the HAs inserted into the ear canals, configured to unit gain. The measured levels at the ear drum microphones then had to match the desired level of 80 dB SPL after subtraction of the REUG. This procedure ensured that potential additional corrections to the receiver equalization filter, e.g., caused by the fact that the closed HA ear tip attenuated sound entering the ear canal in a frequency-dependent way, were taken into account in the calibration.

[1]To limit the maximum output level of the receiver signal, an additional ultra-fast-acting compressor was applied to the output signal, with attack and release times of 2 and 5 ms, respectively. This soft clipping protection mechanism compressed the signal at a slope of 0.5 when broadband levels of 0.8 or higher relative to the digital maximum were detected (Herzke *et al.*, 2017)

Bentler, R. A., Niebuhr, D. P., Getta, J. P., and Anderson, C. V. (**1993**). "Longitudinal study of hearing aid effectiveness. II: Subjective measures," J. Speech Lang. Hear. Res. **36**(4), 820–831.

Best, V., Keidser, G., Buchholz, J. M., and Freeston, K. (**2015**). "An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment," Int. J. Audiol. **54**(10), 682–690.

Bisgaard, N., Vlaming, M. S., and Dahlquist, M. (**2010**). "Standard audiograms for the iec 60118-15 measurement procedure," Trends Amplif. **14**(2), 113–120.

Boike, K., and Souza, P. (**2000**). "Effect of compression ratio on speech recognition in temporally complex background noise," in *International Hearing Aid Conference, Lake Tahoe, CA*.

Bradley, J. V. (**1958**). "Complete counterbalancing of immediate sequential effects in a Latin square design," J. Am. Stat. Assoc. **53**(282), 525–528.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**(6), 4007–4018.

Cord, M., Baskent, D., Kalluri, S., and Moore, B. (**2007**). "Disparity between clinical assessment and real-world performance of hearing aids," Hear. Rev. **14**(6), 22.

Daniel, J. (**2000**). "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimedia" ("Sound field representation and its application to the transmission and reproduction of complex auditory scenes in a multimedia context"), Ph.D. thesis, Université Paris 6, Paris, France.

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

Hendrikse, M. M., Grimm, G., and Hohmann, V. (**2020**). "Evaluation of the influence of head movement on hearing aid algorithm performance using acoustic simulations," Trends Hear. **24**, 233121652091668.

Herzke, T., Kayser, H., Loshaj, F., Grimm, G., and Hohmann, V. (**2017**). "Open signal processing software platform for hearing aid research (openMHA)," in *Proceedings of the Linux Audio Conference*, pp. 35–42.

Hunt, R., Bell, S., and Simpson, D. (**2019**). "Predicting the impact of hearing aid processing on speech intelligibility," J. Acoust. Soc. Am. **146**(4), 2920–2920.

Jensen, J., and Taal, C. H. (**2016**). "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," IEEE/ACM Trans. Audio Speech Lang. Process. **24**(11), 2009–2022.

Jirsa, R. E., and Norris, T. W. (**1982**). "Effects of intermodulation distortion on speech intelligibility," Ear Hear. **3**(5), 251–256.

Jørgensen, S., and Dau, T. (**2011**). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," J. Acoust. Soc. Am. **130**(3), 1475–1487.

Keidser, G., Dillon, H., Flax, M., Ching, T., and Brewer, S. (**2011**). "The NAL-NL2 prescription procedure," Audiol. Res. **1**(1), 88–90.

Köbler, S., and Rosenhall, U. (**2002**). "Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification" ("Localización horizontal y discriminación del lenguaje con adaptación unilateral y bilateral de auxiliares auditivos"), Int. J. Audiol. **41**(7), 395–400.

Kochkin, S. (**2002**). "10-year customer satisfaction trends in the US hearing instrument market," Hearing Rev. **9**(10), 14–25.

Kowalewski, B., Zaar, J., Fereczkowski, M., MacDonald, E. N., Strelcyk, O., May, T., and Dau, T. (**2018**). "Effects of slow-and fast-acting compression on hearing-impaired listeners' consonant–vowel identification in interrupted noise," Trends Hear. **22**, 233121651880087.

Mansour, N., Marschall, M., Westermann, A., May, T., and Dau, T. (**2019**). "Speech intelligibility in a realistic virtual sound environment," in *23rd International Congress on Acoustics*, Deutsche Gesellschaft für Akustik eV, pp. 7658–7665.

Moore, B., Alcántara, J., Stone, M., and Glasberg, B. (**1999**). "Use of a loudness model for hearing aid fitting: II. Hearing aids with multi-channel compression," Brit. J. Audiol. **33**(3), 157–170.

Moore, B. C., Laurence, R. F., and Wright, D. (**1985**). "Improvements in speech intelligibility in quiet and in noise produced by two-channel compression hearing aids," Brit. J. Audiol. **19**(3), 175–187.

Naylor, G. (**2016**). "Theoretical issues of validity in the measurement of aided speech reception threshold in noise for comparing nonlinear hearing aid systems," J. Am. Acad. Audiol. **27**(7), 504–514.

Naylor, G., and Johannesson, R. B. (**2009**). "Long-term signal-to-noise ratio at the input and output of amplitude-compression systems," J. Am. Acad. Audiol. **20**(3), 161–171.

Nielsen, J. B., and Dau, T. (**2011**). "The Danish hearing in noise test," Int. J. Audiol. **50**(3), 202–208.

Noble, W., and Gatehouse, S. (**2006**). "Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ)" ["Efectos de la adaptación uni o bilateral de auxiliares auditivos en las habilidades medidas la escala de cualidades auditiva, espacial y del lenguaje (SSQ)]," Int. J. Audiol. **45**(3), 172–181.

Oreinos, C., and Buchholz, J. M. (**2016**). "Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids," J. Am. Acad. Audiol. **27**(7), 541–556.

Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**(1), 577–587.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2009**). "The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise," J. Acoust. Soc. Am. **126**(6), 3236–3245.

Sanchez-Lopez, R. H., Dau, T., and Jepsen, M. L. (**2019**). "Hearing-aid settings in connection to supra-threshold auditory processing deficits," in *Proceedings of the International Symposium on Auditory and Audiological Research*, Vol. 7, pp. 221–228.

Saunders, G. H., and Kates, J. M. (**1997**). "Speech intelligibility enhancement using hearing-aid array processing," J. Acoust. Soc. Am. **102**(3), 1827–1837.

Seewald, R. C., Bornstein, S. P., and Randolph, K. J. (**1981**). "Speech intelligibility as a function of hearing aid microphone location," J. Acoust. Soc. Am. **70**(S1), S108–S108.

Souza, P. E., Arehart, K. H., Shen, J., Anderson, M., and Kates, J. M. (**2015**). "Working memory and intelligibility of hearing-aid processed speech," Front. Psychol. **6**, 526.

Stone, M. A., and Moore, B. C. (**1999**). "Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses," Ear Hear. **20**(3), 182–192.

Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**(2), 212–215.

Weisser, A., Buchholz, J. M., and Keidser, G. (**2019**). "Complex acoustic environments: Review, framework, and subjective model," Trends Hear. **23**, 233121651988134.

Westermann, A., and Buchholz, J. M. (**2015**). "The influence of informational masking in reverberant, multi-talker environments," J. Acoust. Soc. Am. **138**(2), 584–593.

Wouters, J., Litière, L., and Van Wieringen, A. (**1999**). "Speech intelligibility in noisy environments with one-and two-microphone hearing aids," Audiology **38**(2), 91–98.

Wu, Y.-H. (**2010**). "Effect of age on directional microphone hearing aid benefit and preference," J. Am. Acad. Audiol. **21**(2), 78–89.