Understanding variability in individual response to hearing aid signal processing in wearable hearing aids

Pamela Souza

Department of Communication Sciences and Disorders and Knowles Hearing Center

Northwestern University

Kathryn Arehart

Department of Speech Language Hearing Sciences

University of Colorado at Boulder

Tim Schoof

Department of Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences

University College London

Melinda Anderson

Department of Otolaryngology

University of Colorado Anschutz Medical Campus

Dorina Strori

Department of Communication Sciences and Disorders and Department of Linguistics

Northwestern University

Lauren Balmert

Biostatistics Collaboration Center, Department of Preventive Medicine

Feinberg School of Medicine, Northwestern University

Abstract

Objectives. Previous work has suggested that individual characteristics, including amount of hearing loss, age, and working memory ability, may affect response to hearing aid signal processing. The present study aims to extend work using metrics to quantify cumulative signal modifications under simulated conditions to real hearing aids worn in everyday listening environments. Specifically, the goal was to determine whether individual factors such as working memory, age, and degree of hearing loss play a role in explaining how listeners respond to signal modifications caused by signal processing in real hearing aids, worn in the listener's everyday environment, over a period of time.

Design. Participants were older adults (age range 54-90 years) with symmetrical mild-to-moderate sensorineural hearing loss. We contrasted two distinct hearing aid fittings: one designated as mild signal processing and one as strong signal processing. Forty-nine older adults were enrolled in the study and thirty-five participants had valid outcome data for both hearing aid fittings. The difference between the two settings related to the wide dynamic range compression (WDRC) and frequency compression features. Order of fittings was randomly assigned for each participant. Each fitting was worn in the listener's everyday environments for approximately five weeks prior to outcome measurements. The trial was double blind, with neither the participant nor the tester aware of the specific fitting at the time of the outcome testing. Baseline measures included a full audiometric evaluation as well as working memory and spectral and temporal resolution. The outcome was aided speech recognition in noise.

Results. The two hearing aid fittings resulted in different amounts of signal modification, with significantly less modification for the mild signal processing fitting. The effect of signal processing on speech intelligibility depended on an individual's age, working memory capacity, and degree of hearing loss. Adults who were older demonstrated progressively poorer speech recognition at high levels of signal modification. Working memory interacted with signal processing, with individuals with lower working memory demonstrating low speech intelligibility in noise with both processing conditions, and individuals with higher working memory demonstrating better speech intelligibility in noise with the mild signal processing fitting. Amount of hearing loss interacted with signal processing, but the effects were very small. Individual spectral and temporal resolution did not contribute significantly to the variance in the speech intelligibility score.

Conclusions. When the consequences of a specific set of hearing aid signal processing characteristics were quantified in terms of overall signal modification, there was a relationship between participant characteristics and recognition of speech at different levels of signal modification. Because the hearing aid fittings used were constrained to specific fitting parameters that represent the extremes of the signal modification that might occur in clinical fittings, future work should focus on similar relationships with a wider range of signal processing parameters.

Understanding variability in individual response to hearing aid signal processing in wearable hearing aids

Current hearing aids offer a variety of signal processing options. Common approaches include fast- or slow-acting multichannel wide dynamic range compression (WDRC), noise suppression, and feedback suppression. More recently, a number of products have also offered frequency lowering (either frequency compression or frequency transposition). Each feature is intended to improve aided speech perception and/or sound quality, and for the most part there is evidence for benefit of those features (Bentler, 2005; Bentler, Wu, Kettel, & Hurtig, 2008; Simpson, 2009; Souza, 2002, 2016). However, each type of processing may be advantageous for some but not all listeners. In some cases, the variability among participants means that some listeners simply do not benefit from a particular strategy. In other cases, some listeners may be negatively affected.

While it may be possible to increase the benefit of a specific signal processing approach by using different parameter settings, definitive evidence to guide selection of those parameters is not yet available. Even with clinicians' best attempts to make parameter adjustments that optimize signal processing for each listener, there is considerable variability in listener response. Several studies have explored the factors underlying this variability. A general approach of such work has been to manipulate one setting of a specific feature, and relate that manipulation to individual abilities. In an early demonstration that specific listener factors could affect outcome in response to signal processing, Gatehouse and colleagues (Gatehouse, Naylor, & Elberling, 2006a) showed that listeners with a varied listening environment and better cognitive ability had better aided speech perception with fast-acting than with slow-acting WDRC, whereas listeners with a more restricted listening environment and lower cognitive ability performed better with slow-acting WDRC. A relationship between cognitive ability and compression speed has since been affirmed in a number of other studies (e.g., Foo, Rudner, Rönnberg, & Lunner, 2007; Lunner & Sundewall-Thoren, 2007; Ohlenforst, MacDonald, & Souza, 2015; Souza & Sirow, 2014).

Frequency compression has generated more treatment uncertainty, with studies of adult listeners showing only a subset of treated individuals received benefit (Picou, Steven, & Ricketts, 2015; Souza, Arehart, Kates, Croghan, & Gehani, 2013). This finding has been proposed to be related to the level of signal manipulation versus the improvement in audibility (Brennan, Lewis, McCreery, Kopun, & Alexander, 2017; Souza et al., 2013). Presumably, if improved audibility is the dominant effect, speech recognition will be better with frequency compression. If signal manipulation is the dominant effect (without a significant improvement in audibility), speech recognition will be worse with frequency compression. That idea is consistent with data showing that frequency compression benefits occur mainly for listeners with poorer high-frequency thresholds (e.g., Shehorn, Marrone, & Muller, 2017; Souza et al., 2013).

A similar audibility-to-modification tradeoff has been tested for digital noise reduction, usually by manipulating either the strength of the noise reduction algorithm and/or the extent of the "error" (i.e., the degree to which noise components are inadvertently retained and speech components are inadvertently removed) (Arehart, Souza, Kates, Lunner, & Pedersen, 2015; Desjardins & Doherty, 2014; Neher, 2014; Neher, Grimm, & Hohmann, 2014; Ng et al., 2014; Ng, Rudner, Lunner, Pedersen, & Rönnberg, 2013). Some studies found working memory was a predictor of response to such signal modifications (e.g., Arehart, Souza, Baca, & Kates, 2013; Ng et al., 2013) while others did not (Neher et al., 2014).

In many studies of hearing aid signal processing, the algorithms are described in terms of parameter settings and not in terms of how the signal is actually being modified. We have approached this issue by using a metric (Kates & Arehart, 2014a, 2014b) that directly quantifies the changes in the time-frequency modulation of the signal. Using such a metric, we found that individual factors predict variability in how listeners respond to greater amounts of signal modification. For example, recent studies by our research group have reported that the intelligibility of noisy speech processed with simulations of frequency compression (Arehart et al., 2013), of noise suppression (Arehart et al., 2015), and of WDRC combined with frequency compression (Souza, Arehart, Shen, Anderson, & Kates, 2015) are systematically related to changes in signal modification. Specifically, listeners with better hearing, better working memory and/or who were younger had better intelligibility than listeners with worse hearing, poorer working memory and/or who were older; and the magnitude of the intelligibility difference increased with more signal modification.

The idea that manipulation of signal processing parameters affects speech intelligibility may be interpreted in the context of perceptual models (e.g., Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008). The premise is that the stored lexical representations by which meaning is assigned to acoustic patterns represent the unmodified speech signal. When acoustic patterns are substantially modified—as may be the case with some signal processing parameters, and/or in high levels of background noise--it may be more difficult for the listener to match those acoustic patterns to stored lexical information. The process whereby the altered acoustic pattern is deliberately reconciled to the lexically stored item requires that more cognitive resources be deployed. This process is proposed to draw on working memory capacity. Accordingly, participants with lower working memory capacity may be at a disadvantage when listening to a modified speech signal.

The present study aims to extend our work using metrics to quantify cumulative signal modifications under simulated conditions *to real hearing aids worn in everyday listening environments*. Use of wearable aids coupled to appropriate earmolds incorporates acoustic effects that are not captured by laboratory simulations. Such data can also move beyond time-limited laboratory work to consider the experience gained with a new signal processing approach over the duration of hearing aid use. Ng and colleagues (Ng et al., 2014) recently suggested that acclimatization to signals in everyday environments may modulate or alter the factors predicting individual response; and that the lexical "mismatch" postulated by perceptual models may contribute to a greater extent early in use of the hearing aid. On the other hand, some studies suggested that working memory continues to influence response to signal processing even after a period of acclimatization (e.g., Gatehouse, Naylor, & Elberling, 2006b). Therefore, we considered the extent which individual factors such as working memory, age, and degree of hearing loss play a role in explaining how listeners respond to signal modifications caused by signal processing in real hearing aids, worn in the listener's everyday environment, after a period of acclimatization.

To that end, we designed a trial in which we contrasted two distinct fittings: one with mild signal processing expected to result in relatively little signal modification and one with strong signal processing expected to result in larger amounts of signal modification, as quantified by our signal fidelity metrics. Each fitting was worn in the listener's everyday environments for approximately five weeks prior to outcome measurements, to allow time for acclimatization to occur. To maintain a high level of scientific integrity, the trial was double blind, with neither the participant nor the tester aware of the specific fitting at the time of the outcome testing. As in our laboratory work, a goal was to assess whether the response to signal modification due to signal processing was predicted by individual factors.

Methods

Data were collected at two sites, Northwestern University and the University of Colorado at Boulder, following the same protocol and equipment, as described below.

Participants

Audiometric inclusion criteria were bilateral sensorineural hearing loss with a four-frequency pure-tone average (PTA; 0.5, 1, 2, 4 kHz) in each ear of at least 30 dB HL, audiometric thresholds through 3 kHz no poorer than 70 dB HL, symmetrical hearing loss (between-ear PTA difference ≤ 15 dB), and normal tympanograms bilaterally (Wiley et al., 1996). None of the participants had worn hearing aids in the previous year. The participants were all native speakers of American English, had good self-reported health, normal or corrected-to-normal vision (≤20/50 on the Snellen Eye Chart), and passed the Montreal Cognitive Assessment (Nasreddine et al., 2005) with a score of 22 or better. A group of 49 older adults were enrolled for this study. Two participants withdrew from the study before they were fit with hearing aids (one because of loudness sensitivity concerns, and the other for personal reasons). Five participants withdrew from the study shortly after their first hearing aid fitting. Of these five individuals, three could not tolerate the strong signal processing fitting even after adjustments, one could not tolerate the mild signal processing fitting even after adjustments, and one was unable to correctly insert the hearing aid after repeated practice and reinstruction and did not like how the hearing aid felt once inserted. An additional two participants were later excluded because they did not wear the hearing aids for the minimum required hours of use per day (i.e., might not have acclimatized to the signal processing). Therefore, 40 older adults aged 54 – 90 years (mean age 72 years; 19 women) were ultimately included in the dataset. Their audiograms are shown in Figure 1 and distribution of hearing thresholds to age in Figure 2. Mean high-frequency pure-tone averages (2, 3, 4 kHz) were 51 dB HL (range 32-75 dB HL) in the right ear and 52 dB HL (range 30-87 dB HL) in the left ear. Higher age was not significantly associated with greater high-frequency hearing loss (right ear high-frequency pure-tone average: r=.23, p=.15; left ear high-frequency pure-tone average: r=.32, p=.05). Mean unaided monosyllabic word recognition scores (NU6 presented at 30 dB SL re: PTA) were 90.7% correct for the right ear and 85.7% correct for the left ear. Mean unaided (bilateral) QuickSIN score was 4.6 dB.

Study Timeline

The study consisted of eight visits of approximately 2 hours each. Baseline measures (described in detail below) and earmold impressions were obtained during the first two visits. At the third visit, the participant was fit with hearing aids. One week after the first fitting, the participant returned to the clinic for a follow-up appointment. Three weeks after the hearing aid fitting s/he was contacted by telephone to assess any problems. The participant returned for an evaluation at week five or six (depending on participant schedule constraints). Following the first set of outcome measurements, the fitting was transitioned to the second fit where the timeline repeated (fitting, one-week in-person follow-up, three-week telephone follow-up, final evaluation at five or six weeks post-fitting).

The fitting order was randomly chosen for each participant. The study was double-blinded. The audiologist who conducted the hearing aid fittings and the in-person and telephone follow-ups knew the fitting order, but the participant and the experimenter who conducted the baseline and outcome measure visits did not. Hearing aid fittings and adjustments took place in a quiet examination room. Baseline and outcome measures were obtained in a double-walled sound booth.

All study procedures were approved by the Institutional Review Boards of Northwestern University and the University of Colorado-Boulder. Participants completed an informed consent process and were paid for their participation. Participants received an hourly compensation rate for the study visits. To improve retention, participants received bonus payments at the first and second outcome visits.

Baseline Measures

Working memory. The reading span test, developed by Rönnberg and colleagues (Rönnberg, Arlinger, Lyxell, & Kinnefors, 1989), was used to measure working memory capacity. This task taxes information storage and rehearsal and requires information processing. Participants were asked to read sentences on a computer screen, which appeared one word or word pair at a time. Words or word pairs were presented at a rate of 0.8 s/word. At the end of each sentence, participants were asked to judge whether the sentence made semantic sense or not (e.g., "The train" "sang" "a song", or "The captain" "sailed" "his boat"). The inter-sentence interval, during which participants had to make the semantic judgment, was 1.75 s. These sentences appeared in blocks of 3-6 sentences. At the end of each block, participants were asked to recall either the first or the last word in each sentence and to repeat those words (in any order). Participants received training on one block of three sentences. The percentage of correctly recalled words was taken as the measure of working memory capacity.

Spectral and temporal resolution. We reasoned that if the purported benefit of both fast-acting WDRC and frequency compression is to increase the amount of audible speech information, it is not only necessary that the information be suprathreshold but that the listener be able to resolve that information. At least one previous study (Kates et al., 2013) has shown that spectral resolution explains a portion of the variance in response to frequency compression. In addition, individual temporal and/or spectral resolution may influence benefit of fast-acting WDRC (Davies-Venn & Souza, 2014; Dreschler, 1989).

Accordingly, temporal resolution was measured using a gap detection task (Brennan, Gallun, Souza, & Stecker, 2013). The carrier signal was a broadband noise spanning 0.1 – 10 kHz, with a duration of 250 ms, tapered on and off across 10 ms. Gaps were introduced using 0.5 ms cosine squared ramps. Gap detection thresholds were determined using a three-alternative forced choice task, following a two-down one-up rule, thus tracking 70.7% correct (Levitt, 1971). The initial gap duration was 100 ms and changed by a factor of 1.4 and, after the first four reversals, by a factor of 1.2 on subsequent trials. Visual correct-answer feedback was provided.

Stimuli were presented monaurally to the better ear via Sennheiser HD-25 headphones at 35 dB SL with respect to the four-frequency PTA. Participants started with a practice block, followed by two test blocks. A block terminated after 10 reversals. Gap detection thresholds were computed as the mean across the final six reversals in a block, with a final score based on the average of two test blocks.

Spectral resolution was measured using a spectral ripple detection task (Won, Drennan, & Rubinstein, 2007). The stimuli consisted of a weighted sum of 800 sinusoidal components ranging from 100 to 5000 Hz. Spectral ripples were introduced by adjusting the amplitudes of the components using a full-wave rectified sinusoidal envelope on a logarithmic scale. Ripple stimuli with 16 different densities (ripples per octave) were generated. The ripple densities differed by ratios of 1.414, ranging from 0.125 to 22.628 ripples/octave. The peak-to-valley ratio of the ripples was 30 dB. The stimuli were subsequently filtered using a long-term speech-shaped filter. The stimuli had a duration of 500 ms and were tapered on and off across 150 ms. For each ripple density, a reference and a test stimulus were generated that differed only in terms of the phase of the ripples (by $\pi/2$).

Ripple detection thresholds were determined using a three-alternative forced choice procedure tracking 70.7 % correct (two-up, one-down; Levitt, 1971). Each trial consisted of two reference stimuli and one test stimulus (inverted phase). The participant's task was to determine which of the three sounds was different (i.e. the inverted-phase test stimulus). Each block started with 0.176 ripples/octave and increased or decreased at subsequent trials in equal ratio steps of 1.414. The presentation level was roved across an 8 dB range (in 1 dB steps) to minimize level cues. No feedback was provided. Stimuli were presented monaurally to the better ear via Sennheiser HD-25 headphones at 35 dB SL with respect to the 4-frequency PTA. Participants received training on one practice block, which terminated after four reversals. The experiment consisted of two test blocks, with ten reversals per block. The results reported here are the mean ripple densities across the final six reversals.

Loudness discomfort levels. To assist in setting hearing aid maximum output, frequency-specific loudness discomfort levels (LDLs) were measured for both ears using warble tones at 0.5 and 3 kHz. Following an ascending procedure, consistent with loudness scaling as described by Cox et al. (Cox, Alexander, Taylor, & Gray, 1997), participants were asked to indicate when the stimulus became uncomfortably loud.

Hearing aids. All hearing aids were 20-channel behind-the-ear (BTE) devices. Hearing aids were fit using slim tube and custom earmolds. Every earmold was a vinyl canal mold with a 2 mm vent. A canal lock was added for some participants to address retention problems. The manufacturer provided an experimental version of the fitting software allowing for manipulation of the WDRC time constants. The hearing aids were programmed by creating a custom program with all noise reduction features and feedback management disabled and the directional microphones set to omni-directional. All push buttons and volume controls were also disabled. The goal of these fitting constraints was to ensure the participant listened to sound processed with the desired hearing aid parameters.

Participants wore the same hearing aids programmed to two different settings: a strong signal processing setting and a mild signal processing setting. The difference between the two settings related to the WDRC and frequency compression features. In the mild signal processing setting the hearing aids were programmed with slow (attack: 1160 ms, release: 6900 ms) WDRC time constants and frequency compression was disabled. In the strong signal processing setting WDRC time constants were set to fast (attack: 13 ms, release: 59 ms) and frequency compression was enabled. Both fittings employed a compression limiter to control maximum output. The order of the two fittings was counterbalanced across participants. Participants were instructed to wear the hearing aids at least five hours/day.

The first hearing aid fitting was completed by matching real ear aided response (REAR) targets using the NAL-NL2 (Dillon, Keidser, Ching, Flax, & Brewer, 2011) prescribed response. The hearing aid gain was fit to target, with a goal of being within 3 dB of the prescribed REAR from .25-2 kHz and within 5 dB between 2 and 6 kHz for the International Speech Test signal (ISTS; Holube, Fredelake, Vlaming, & Kollmeier, 2010) presented at 55, 65, and 75 dB SPL.

Regardless of the signal processing setting, all first fits were matched to target with slow WDRC and FC turned off (Figure 3). If a participant was to be fit with the strong signal processing setting first, after REAR measurements and gain adjustments were completed the hearing aids were set to the strong signal processing condition by adjusting compression speed to fast and activating frequency compression. Frequency compression was initially set with a compression ratio of 3:1 and a cutoff frequency of 1.9 kHz. If the participant found the sound quality objectionable, the audiologist reduced the extent of frequency compression until sound quality was deemed acceptable, to a minimum compression ratio of 2:1 and maximum cutoff frequency of 2.2 kHz. Across all participants, the mean frequency compression ratio after adjustment was 2.67 and the mean cutoff frequency after adjustment was 2.1 kHz. After fast-acting compression was active and frequency compression was active and adjusted, real-ear measurements were rerun for documentation, without making further gain adjustments.

At the second fitting, if the fitting was to be mild signal processing, the compression speed was changed to slow and frequency compression was deactivated. If the fitting was to be strong signal processing, the compression speed was changed to fast and frequency compression was activated and adjusted using the criteria described above. No gain changes were made at the time of the second fitting in order to assure similar amounts of gain between the first and second fittings. However, real-ear testing was repeated for documentation purposes.

Hearing aid follow-up

One week after the first hearing aid fitting, the participant returned for a follow-up visit. At this visit the participant completed the Practical Hearing Aid Skills Test-Revised (PHAST-R; Desjardins & Doherty, 2012) with the exception of the sections related to adjusting manual controls (which were disabled for the duration of the study). The PHAST allowed the audiologist to verify the participant was able to use the hearing aids in a consistent manner across sites and across participants. Hearing aids were checked and concerns addressed to the extent allowed by study constraints. In case of physically uncomfortable fits, the earmolds were modified or remade. If a participant's reports were consistent with too much gain, the gain was reduced with the constraint that the hearing aids were still within the gain tolerances established above. If a participant reported bothersome sound quality with frequency compression, compression parameters were adjusted as described above. If any programming changes were made, real-ear verification was repeated. In addition, datalogging was performed to confirm the participant was wearing the hearing aids for at least 5 hours/day. If lower use was noted, it prompted an inquiry into factors limiting use.

Outcome Measures

Speech recognition. Speech recognition was measured using low-context sentences (Rothauser et al., 1969) spoken by a female American English speaker. Each sentence contained five keywords. The sentences were presented in four-talker babble at fixed signal-to-noise ratios (SNR) of 0, 5, and 10 dB, representing a range of realistic listening situations (Hodgson, Steininger, & Razavi, 2007; Olsen, 1998). The babble began 3 seconds prior to sentence onset and continued for an additional 0.5 seconds after the sentence had been presented. The desired SNRs were obtained by adjusting the level of the masker while keeping the level of the speech fixed at 65 dB SPL, as measured in soundfield at the position of the listener's head. The stimuli were presented using a Mac Mini computer connected to a speaker (KEF, iQ1) via an external amplifier. The speaker was placed in front of the participant at a distance of 1 meter.

The participants' task was to repeat the sentences as best as they could. The experimenter recorded the number of correctly repeated key words for each sentence. The stimuli were presented at an inter-stimulus interval of 4.5 seconds. Sentences were presented in blocks of ten, containing 50 keywords in total. Two blocks were presented at each SNR. The order of sentence lists and SNRs was randomized across participants.

Results

Baseline Measures

Working memory. The mean reading span score was 34.1%, with a range from 11.1% to 55.6%. The distribution of scores (Figure 4) was very similar to previously reported results for older listeners (e.g., Souza & Arehart, 2015). Higher age was not significantly associated with poorer reading span scores (r=-.10, p=.53)

Temporal and spectral resolution. The mean gap detection score was 8.7 ms, with a range of 3.1-10.7 ms. Gap detection was not related to pure-tone average (r=.13, p=.42). The mean ripple score was 3.0 ripples/octave, with a range of 0.2-6.0 ripples/octave. Ripple detection was negatively related to pure-tone average (r= -0.45, p=.003), such that listeners with poorer hearing also had poorer spectral resolution. The scores and their relationships with hearing thresholds were consistent with published values for participants with similar age and hearing loss (Davies-Venn, Nelson, & Souza, 2015; Henry, Turner, & Behrens, 2005).

Hearing aid use. For each fitting (irrespective of signal processing), mean hearing aid use was 9 hours per day (range 5-17 hours). Mean PHAST score was 97% (range 86%-100%). Our mean PHAST score was higher than the mean scores of 78%-88% reported by Desjardins and Doherty (2009; 2012) for experienced users, perhaps reflecting the structured nature of our fitting appointments, including time dedicated to hearing aid instruction.

Outcome Measures

Thirty-five participants had valid outcome data for both the mild and strong signal modification hearing aid fittings. An additional five participants only had outcome measures data for one of the two hearing aid fittings. Two of those participants withdrew from the trial shortly after the second hearing aid fitting because they could not tolerate the second (strong signal processing) settings. Another participant was dropped after the second fitting when that participant lost multiple study hearing aids. Due to a fitting error, two participants were not fit correctly in the strong signal processing condition (compression speed was incorrectly set to slow), so their outcome scores for the strong signal processing setting were removed from the data set.

Signal modification. To quantify the amount of cumulative signal modification caused by hearing aid signal processing, we calculated metric values that were customized for each individual participant's hearing loss. Following the procedures of Kates, Arehart, Anderson, Muralimanohar, and Harvey (2018), acoustic recordings were made for speech stimuli processed through the study hearing aids. The changes in the time-frequency modulations of the signal were then calculated based on differences between the reference and test conditions. The reference signal was speech in quiet at the input to the hearing-aid microphone, to which NAL-R equalization was applied. The test conditions included the speech (in quiet and in the four-talker babble at 0, 5 and 10 dB SNR) processed through the hearing aid for each participant's user settings for both the strong and mild signal processing conditions.

Both the reference and test conditions were processed through an auditory model that considered the user's audiogram. The metric was calculated by first processing the reference and test conditions through an auditory model of the impaired auditory system (Kates, 2013) that was customized for each listener based on their audiogram and that took into account changes that hearing loss has on auditory filtering and nonlinearities. The model produced output envelope signals that were expressed in dB above the normal or impaired auditory threshold. The envelope in each frequency band was smoothed using a 62.5 Hz lowpass filter implemented using a sliding raised-cosine window, and the smoothed envelope was resampled at 125 Hz. A smoothed version of the log magnitude spectrum produced by the auditory model was then computed at each time sample. The cross-correlation of the smoothed spectra from the reference and processed signals was computed to produce the cepstral correlation (Kates et al., 2018), which measures the degree to which the time-frequency envelope modulation of the processed signal matches that of the reference. The cepstral correlation values are related to the time-frequency modulation patterns of speech that are used in speech recognition (Zahorian & Rothenberg, 1981).

The closer the metric value is to 1 the less signal modification was caused by the hearing aid signal processing. The metric values (Table 1) showed more signal modification for the strong signal processing fit compared to the mild signal processing fit. As expected, average metric values also decreased as the level of the noise increased. Figure 5 shows the metric values for the mild and strong signal processing conditions for individual listeners for each SNR condition. First, we considered any relationships between amount of hearing loss and metric values. After correcting for multiple correlations the only significant relationships were for mild signal processing at 0 dB (r=.60, p<.001) and 5 dB (r=.47, p=.004), where metric values improved slightly with more hearing loss. Recall that the metric expresses envelope relative to effects of individual hearing loss, including auditory thresholds. For listeners with more hearing loss, less of the noise is above threshold, resulting in slightly better envelope fidelity.

Second, to verify whether the two hearing aid fittings resulted in different amounts of signal modification, a linear mixed-effects model with cepstral correlation as the dependent variable was performed in R using the *lme()* function from the *nlme* package. The model included signal processing (mild vs. strong) and SNR (planned contrasts comparing 0 vs. 5 dB and 5 vs. 10 dB SNR) and their interactions as fixed factors, and participant as a random intercept. The results, summarized in Table 2, confirmed that the two hearing aid fittings resulted in different amounts of signal modification, with significantly less modification (i.e, higher signal fidelity) for the mild signal processing fitting. SNR also affected the signal fidelity, with lower SNRs resulting in lower signal fidelity. There was no significant interaction between SNR and hearing aid fitting, suggesting that the amount of signal modification introduced by the different hearing aid fittings did not depend on the SNR.

Speech recognition. The distribution of aided speech-recognition scores are shown in Figure 6 for both fittings and the three signal-to-noise ratios. The speech recognition data were analyzed using a logistic mixed-effects model in R (using the *glmer()* function from the *lme4* package). The logistic mixed-effects model offers a number of advantages over alternative approaches, such as the commonly-used rationalized arcsine transform (Studebaker, 1985). The logistic transform converts percent correct scores (based on our binary outcome variable) into a range from $-\infty$ to ∞, which means that floor and ceiling effects are not a limitation. (For a more detailed discussion, the interested reader is referred to Hilkhuysen [2015]). The dependent variable was a binary outcome measure indicating whether the keywords were correctly repeated or not. Based on the individual characteristics identified in our previous work as having predictive value, the mixed-effects model included SNR (planned contrasts comparing 0 vs. 5 dB and 5 vs. 10 dB), hearing aid fitting (mild or strong signal processing), PTA (0.5, 1, 2, 4 kHz, mean across the ears), age, and reading span test (RST) score as well as two-way interactions between hearing aid signal processing and the participant characteristics (age, PTA, and RST) as fixed effects. Continuous variables (age, PTA, and RST score) were all centered by subtracting the mean before they were entered into the model. This allowed for better interpretability of regression coefficients, particularly for interaction terms. To account for correlation of observations from the same participant or same sentence, the model included random effects for participant and keyword. In addition, test session (outcome A or B) was added to the model as a fixed effect because model comparisons based on the Akaike Information Criterion (AIC; Akaike, 1974) indicated improvement in model fit. Further model comparisons showed that adding testing site (Northwestern University, University of Colorado at Boulder) or the measures of spectro-temporal processing (gap detection and spectral ripple detection) to the model did not result in a better fit for the data. These variables were therefore not included in the model.

The final model, including fixed and random effects, explained 80.6% of the variance in the data, as indicated by the conditional R^2 (Nakagawa & Schielzeth, 2013). The marginal R^2 indicated that 65.7% of the variance in the data was explained by the fixed effects alone.

The results, summarized in Table 3, showed significant interactions between hearing aid processing and age, between hearing aid processing and RST score, and between hearing aid processing and PTA. Specifically, an increase in age was associated with a larger decrease in odds of correctly answering for strong signal processing than for mild signal processing. For example, for strong signal processing, a 10 year increase in age was associated with a 21% decrease in the odds of correctly repeating a word, holding all other variables constant. However, for mild signal processing, a 10 year increase in age was associated with a 3% decrease in odds of correctly repeating a word, holding all other variables constant. This significant interaction effect is illustrated in Figure 7, in terms of the predicted probability of correctly repeating a word under three levels of signal-to-noise ratio (SNR).

Additionally, the effect of RST score was found to be greater for mild signal processing. For example, a one percent increase in RST score was associated with a 1% increase in odds of correctly repeating for strong signal processing, and a 4% increase for mild signal processing, holding all other variables constant. Figure 8 illustrates this relationship (in terms of predicted probabilities) at the three SNR levels.

The effect of PTA interacted with signal processing, but the difference was very small (odds ratio for PTA under strong signal processing = 0.96; odds ratio for PTA under mild signal processing 0.958) and the interaction is likely due to score compression at the extremes of the probability range. To illustrate this, consider the relationships shown in Figure 9. The lines representing predicted probabilities for mild and strong signal processing are essentially parallel except for minimum predicted scores (i.e., PTA > 50 dB HL at 0 dB SNR) and maximum predicted scores (i.e., PTA < 40 dB HL at 10 dB SNR). In other words, the PTA x signal processing interaction is likely related to the range of difficulty of the selected SNRs.

In addition to the interaction effects, the main effects for SNR and session were also found to be statistically significant. Session 2 was associated with a 57% increase in odds of correctly repeating a word, compared to session 1 (odds ratio = 1.57). Similarly, higher levels of SNR were associated with greater odds of correctly repeating a word (SNR 0 vs 5 dB: odds ratio = 7.5, SNR 5 vs 10 dB: 3.1).

Discussion

The purpose of the trial described here was, essentially, a proof of concept: when the consequences of a specific set of hearing aid signal processing characteristics were quantified in terms of overall signal modification, was there a relationship between participant characteristics (age, hearing loss, and/or working memory) and recognition of speech at different levels of signal modification? Such a relationship had been shown in our laboratory work (Arehart, Kates, & Souza, 2014; Arehart et al., 2013; Arehart et al., 2015; Kates, Arehart, & Souza, 2013; Souza et al., 2015) but it was unknown whether the same relationships would be demonstrated with wearable hearing aids which operated in a more multifaceted way (i.e., with dynamic gains and compression characteristics) and after a period of acclimatization. We were keen to test our hypotheses in wearable hearing aids, to more closely represent real-life aided listening for the population of interest.

Our results indicated that some relationships held true in this study as they had under more constrained laboratory simulations. With regard to age, adults who were older demonstrated progressively poorer speech recognition at high levels of signal modification. Figure 7 illustrates that while the differences in predicted probabilities between SNR levels and modification are present across the entire age range tested, the effects of these factors are larger for the oldest listeners.

In previous work, response to strong or to mild signal processing was associated with differences in working memory. Specifically, in our laboratory studies, listeners with higher working memory performed similarly with strong and mild processing and listeners with lower working memory performed more poorly with strong than with mild modification processing (Arehart et al. 2013, 2015). Some wearable aid studies (e.g., Gatehouse et al., 2006) have also shown that listeners with lower working memory are the most sensitive to processing differences, albeit without direct quantification of signal modification. Such findings are consistent with models of working memory (Rönnberg et al., 2013; Rönnberg et al., 2008) which argue that a mismatch between the expected acoustic patterns and stored lexical representations taxes low working memory capacity and results in degraded scores.

The present data also show an interaction between working memory and signal modification, but the statistical model predicts that when other participant factors have been controlled for, the largest differences between strong and mild processing will occur for listeners with higher working memory capacity (Figure 8). This is a different result than previous studies, which mostly showed the largest differences between strong and mild processing for listeners with lower working memory capacity. Further research is needed to confirm and explain this pattern. The listeners tested here were very similar in age, amount of hearing loss and distribution of working memory scores to those tested in previous studies. Experimental differences relative to previous work include: a much larger number of compression channels; longer attack and release times for the mild processing condition; frequency-gain response closely constrained to a validated prescriptive procedure; and listeners without previous hearing aid experience. It is possible that some of those differences affected the working memory-distortion relationships (i.e., the slope of the predicted probability lines in Figure 8). This will be an important area for future examination in order to understand how patient factors should direct treatment when that treatment uses advanced technology hearing aids.

Some authors have argued that acclimatization will minimize the interaction with working memory as listener “learn” the new patterns (Ng et al., 2014; Rudner, Foo, Rönnberg, & Lunner, 2009). On the other hand, a number of studies have shown that the contribution of working memory (and presumed lexical “mismatch”) is maintained even after weeks of hearing aid use (Gatehouse et al., 2006b). It may be that a very long period of acclimatization is needed--perhaps even years of experience--before the impact of working memory is diminished (Rahlmann et al., 2017). Nonetheless, the persistence of the effect after six weeks of hearing aid use suggests that the working memory contribution is at least fairly robust.

Predictions based on the current data (Figures 7-9) did not indicate that strong processing will provide better speech recognition than mild processing for any listener, regardless of severity of hearing loss, age, or working memory capacity. An advantage of high-modification processing is thought to be due to improved audibility of phonetic contrasts. For example, fast WDRC may provide relatively greater gain to short-duration, low-intensity consonants than would occur with slow compression. Frequency compression is expected to improve audibility for otherwise inaudible high-frequency phonemes. However, both manipulations may introduce distortions that offset any audibility advantages (for a model of such tradeoffs, see Leijon & Stadler, 2008).

To the extent that fast WDRC and/or frequency compression offer an audibility advantage over amplification with only slow compression, it may not have occurred for the strong signal processing condition used here due to a combination of effects: (a) an excellent match to target through 6 kHz in the mild signal modification condition; (b) a group of listeners with relatively good high-frequency thresholds (i.e, few listeners with steeply sloping severe loss who would be unlikely to achieve audibility through high-frequency gain alone); and (c) test materials that allowed use of linguistic experience to infer presence of some high-frequency, less-audible sounds (such as the plural /s/ being simultaneously cued by verb plurality). Such a combination of effects, in which modification (distortion) outweighs audibility improvement, might explain why the high modification processing resulted in lower scores in general, and in particular why the mild-vs-strong signal modification difference was larger for listeners with higher working memory in this study.

The measures of spectral and temporal resolution did not add to the predictive value of the model, despite spectral resolution having predicted response to signal processing in a previous study (Kates et al., 2013). However, there were also important differences. In Kates et al., more extreme frequency compression parameters would have substantially altered spectral cues such as vowel formant spacing and overall spectral shape. It may be that spectral (or temporal) resolution ability is only important when the listener receives signal processing that challenges the limits of spectral ability. To put this another way, if listeners in this study had sufficient spectral and temporal resolution to discriminate the cues received through the fitted hearing aids, there might be no predictive value to measuring more fine-grained resolution.

In a separate paper (Anderson, Rallapalli, Schoof, Souza & Arehart, 2018), we report subjective outcome data collected for the same cohort. Subjective ratings of speech intelligibility and quality were consistent with the measured intelligibility scores. On average, participants reported higher speech intelligibility and quality for the mild signal processing than for the strong signal processing. Interestingly, the range of subjective ratings was larger for the strong signal processing, suggesting that there may be greater variability among listeners who receive strong processing (with some rating it much more favorably than others), compared to a narrower range of individual ratings when listeners receive mild processing.

A small number of participants enrolled in the study rejected the fitted hearing aids on the basis of sound quality. Among the 49 originally enrolled participants, four rejected the fitted hearing aids on the basis of sound quality, for a rejection rate of 8%. There was no obvious pattern to the rejections, which occurred during fittings of both mild and strong processing, and for participants with both higher and lower working memory scores and who had audiograms and loudness discomfort levels nearly identical to participants who completed the study. The dropout rate is consistent with the 6-13% of clinical (non-research) hearing aid wearers who reject their aids as having unacceptable sound quality (e.g., Bertoli et al., 2009; Kochkin, 2000). Moreover, in the present study, that rejection rate occurred when only certain adjustments were permitted. As one example, one participant who withdrew had requested that overall gain of his aid be decreased to a level that was more than 5 dB below target. That adjustment might have been allowed clinically, but did not comply with our study protocol. It is probable that some of the participants who withdrew might have continued wearing the study hearing aids had they been given wider latitude for hearing aid adjustments.

While the purpose of the present trial was not to mimic a clinical scenario *per se* it is of interest to consider the extent to which the present high- and low-signal modification fittings might occur in typical practice. The WDRC speed and frequency compression parameters applied here were chosen to mimic the range of signal modification values used in our previous laboratory work, rather than as clinically typical values. Default manufacturer's parameters for frequency lowering, for example, would likely result in a higher cutoff frequency and lower compression ratio than used in the present study. There are no prescribed values for WDRC speed, although slightly more products use a slower compression speed (Rallapalli, Mueller, & Souza, 2018). Our focus was not on the specific parameters but on the aggregate signal modification created by those parameters. Indeed, the range of signal modification created (ranging from approximately 0.3 to 0.7, depending on the signal processing and the input SNR) was similar to that seen in clinically-fit hearing aids. For example, a signal modification range of approximately 0.2 to 0.8 has been reported for user settings of clinically-fit hearing aids for adults (Kates et al., 2018; Rallapalli, Anderson, Kates,, Sirow, Arehart, & Souza, 2018) and children (Anderson, Mowery, & Uhler, 2018), using a similar metric approach.

In summary, results of the present study were consistent with previous work using hearing aid simulations in that there was a relationship between participant characteristics (other than the audiogram) and recognition of speech at different levels of signal modification. The relevant participant characteristics included age and working memory. However, the present data also diverge from laboratory results in that the largest processing differences occurred for listeners with higher working memory capacity. The data broadly support inclusion of patient factors other than pure-tone thresholds in the hearing aid fitting process. Barriers to inclusion of patient factors include clinician access to appropriate tests (such as working memory tests) and an efficient method to measure aggregate signal modification created by a complex set of signal processing parameters. Work continues in our laboratories to understand the level of signal modification that would occur with a wider range of signal processing parameters and hearing aid features and to seek practical solutions for clinical implementation.

Acknowledgments

All authors contributed equally to this work. P.S. and K.A. designed and managed the experiment. T.S., M.A. and D.S. performed the experiments. T.S. created and managed the study database and supervised data collection for the Evanston site. M.A. supervised data collection for the Boulder site. T.S. and L.B. performed the statistical analysis. P.S. wrote the main paper. All authors discussed the results and implications and commented on the manuscript at all stages. The project was supported by NIH grant R01 DC012289 to P.S. and K.A. We thank James Kates, Ramesh Muralimanohar, and Jing Shen for helpful discussions regarding study design and analysis; Cynthia Erdos, Laura Mathews, Elizabeth McNichols, Arianna

Mihalakakos, Melissa Sherman, Kristin Sommerfeldt, and Varsha Rallapalli for their assistance with data collection and management; and Christine Jones and Olaf Strelcyk for project support. The project was a registered NIH clinical trial (ClinicalTrials.gov Identifier: NCT02448706). Data management via REDCap is supported at Feinberg School of Medicine by the Northwestern University Clinical and Translational Science (NUCATS) Institute. Research reported in this publication was supported, in part, by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Anderson, M. C., Arehart, K., & Souza, P. E. (in press). Survey of current practice in the fitting and fine-tuning of common signal-processing features in hearing aids for adults. *Journal of the American Academy of Audiology*.

Anderson, M., Rallapalli, V., Schoof, T., Souza, P., & Arehart, K. (2018). The use of self-report measures to examine changes in perception in response to fittings using different signal processing parameters. *International Journal of Audiology, 57*, 809-815.

Anderson, M., Mowery, P., & Uhler, K. (2018). Relationship between personal hearing aid settings and infant speech discrimination. Poster presented at the American Academy of Audiology conference, Nashville, TN.

Arehart, K. H., Kates, J. M., & Souza, P. (2014). The role of metrics in studies of hearing and cognition. *ENT Audiology News, 23*, 92-93.

Arehart, K. H., Souza, P., Baca, R., & Kates, J. M. (2013). Working memory, age, and hearing loss: susceptibility to hearing aid distortion. *Ear and Hearing, 34*, 251-260.

Arehart, K. H., Souza, P., Kates, J. M., Lunner, T., & Pedersen, M. S. (2015). Relationship among signal fidelity, hearing loss, and working memory for digital noise suppression. *Ear and Hearing, 36*, 505-516.

Baker, R. J., & Rosen, S. (2002). Auditory filter nonlinearity in mild/moderate hearing impairment. *Journal of the Acoustical Society of America, 111*, 1330-1339.

Bentler, R. (2005). Effectiveness of directional microphones and noise reduction schemes in hearing aids: a systematic review of the evidence. *Journal of the American Academy of Audiology, 16*, 473-484.

Bentler, R., Wu, Y. H., Kettel, J., & Hurtig, R. (2008). Digital noise reduction: outcomes from laboratory and field studies. *International Journal of Audiology, 47*, 447-460.

Brennan, M. A., Gallun, F. J., Souza, P. E., & Stecker, G. C. (2013). Temporal resolution with a prescriptive fitting formula. *American Journal of Audiology, 22*, 216-225.

Brennan, M. A., Lewis, D., McCreery, R., Kopun, J., & Alexander, J. M. (2017). Listening effort and speech recognition with frequency compression amplification for children and adults with hearing loss. *Journal of the American Academy of Audiology, 28,* 823-837.

Cox, R. M., Alexander, G. C., Taylor, I. M., & Gray, G. A. (1997). The contour test of loudness perception. *Ear and Hearing, 18*, 388-400.

Davies-Venn, E., Nelson, P., & Souza, P. (2015). Comparing auditory filter bandwidths, spectral ripple detection, spectral ripple discrimination and speech recognition: normal and impaired hearing. *Journal of the Acoustical Society of America, 138*, 492-503.

Davies-Venn, E., & Souza, P. (2014). The role of spectral resolution, working memory, and audibility in explaining variance in susceptibility to temporal envelope distortion. *Journal of the American Academy of Audiology, 25*, 592-604.

Desjardins, J. D., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and Hearing, 35*, 600-610.

Desjardins, J. L., & Doherty, K. A. (2009). Do experienced hearing aid users know how to use their hearing aids correctly? *American Journal of Audiology, 18*, 69-76.

Desjardins, J. L., & Doherty, K. A. (2012). The Practical Hearing Aids Skills Test-Revised. *American Journal of Audiology*, *21*, 100-105.

Dillon, H., Keidser, G., Ching, T., Flax, M. R., & Brewer, S. (2011). The NAL-NL2 prescription procedure *Phonak Focus*: Phonak AG.

Dreschler, W. A. (1989). Phoneme perception via hearing aids with and without compression and the role of temporal resolution. *Audiology*, *28*, 49-60.

Foo, C., Rudner, M., Rönnberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology*, *18*, 618-631.

Gatehouse, S., Naylor, G., & Elberling, C. (2006a). Linear and nonlinear hearing aid fittings--2. Patterns of candidature. *International Journal of Audiology*, *45*, 153-171.

Gatehouse, S., Naylor, G., & Elberling, C. (2006b). Linear and nonlinear hearing aid fittings--2. Patterns of candidature. *International Journal of Audiology*, *45*, 153-171.

Henry, B. A., Turner, C. W., & Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. *Journal of the Acoustical Society of America*, *118*, 1111-1121.

Hilkhuysen, G. (2015). RM-ANOVA on RAUs vs mixed effects logistic regression: a ruling of the high court for statistics. Poster presented at the SPiN workshop, Copenhagen, Denmark. Retrieved 11/30/2018 from http://www.spin2015.dk/download-abstracts.

Hodgson, M., Steininger, G., & Razavi, Z. (2007). Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments. *Journal of the Acoustical Society of America*, *121*, 2023-2033.

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology, 49,* 891-903.

Kates, J. M. (2013). An auditory model for intelligibility and quality predictions. *Proceedings of Meetings on Acoustics, 19*, 050184.

Kates, J. M., & Arehart, K. (2010). The Hearing-Aid Speech Quality Index (HASQI) *Journal of the Audio Engineering Society, 58*, 363-381.

Kates, J. M., & Arehart, K. H. (2014a). The hearing-aid speech quality index (HASQI) version 2. *Journal of the Audio Engineering Society, 62*, 99-117.

Kates, J. M., & Arehart, K. H. (2014b). The Hearing Aid Speech Perception Index (HASPI). *Speech Communication, 65*, 75-93.

Kates, J. M., Arehart, K. H., Anderson, M. C., Muralimanohar, R. K., & Harvey, L. O. (2018). Using perceptual metrics to measure hearing-aid performance. *Ear and Hearing, 39,* 1165-1176.

Kates, J. M., Arehart, K. H., & Souza, P. (2013). Integrating cognitive and peripheral factors in predicting hearing-aid processing benefit. *Journal of the Acoustical Society of America, 134*, 4458-4469.

Leijon, A., & Stadler, S. (2008). Fast amplitude compression in hearing aids improves audibility but degrades speech information transmission. *2008 16th European Signal Processing Conference*, 1-5.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467.

Lunner, T., & Sundewall-Thoren, 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*, 604-617.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133-142.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society, 53*, 695-699.

Neher, T. (2014). Relating hearing loss and executive functions to hearing aid users' preference for, and speech recognition with, different combinations of binaural noise reduction and microphone directionality. *Frontiers in Neuroscience, 8*, 391.

Neher, T., Grimm, G., & Hohmann, V. (2014). Perceptual consequences of different signal changes due to binaural noise reduction: do hearing loss and working memory capacity play a role? *Ear and Hearing, 35*, e213-227.

Ng, E. H., Classon, e., Larsby, B., Arlinger, S., Lunner, T., Rudner, M., & Ronnberg, J. (2014). Dynamic relation between working memory capacity and speech recognition in noise during the first 6 months of hearing aid use. *Trends in Hearing, 18*, 1-10.

Ng, E. H., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *International Journal of Audiology*.

Ohlenforst, B., MacDonald, E., & Souza, P. (2015). Exploring the relationship between working memory, compressor speed and background noise characteristics. *Ear and Hearing, 37*, 137-143.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett & Fidell (1977) report. *American Journal of Audiology, 7*, 21-25.

Picou, E. M., Steven, C., & Ricketts, T. A. (2015). Evalulation of the effects of nonlinear frequency compression on speech recognition and sound quality for adults with mild to moderate hearing loss. *International Journal of Audiology, 54*, 162-169.

Rahlmann, S., Meis, M., Schulte, M., Kiebling, J., Walger, M., & Meister, H. (2017). Assessment of hearing aid algorithms using a master hearing aid: the influence of hearing aid experience on the relationship between speech recognition and cognitive capacity. *International Journal of Audiology*, 1-7.

Rallapalli, V., Mueller, A., & Souza, P. (2018). *Survey of hearing aid signal processing features across manufacturers*. Paper presented at the American Academy of Audiology, Nashville, TN.

Rallapalli, V., Anderson, M., Kates, J., Sirow, L., Arehart, K., Souza, P. (2018). Quantifying the range of signal modification in clinically-fit hearing aids. Poster presented at the International Hearing Aid Conference, Tahoe City, CA.

Rönnberg, J., Arlinger, S., Lyxell, B., & Kinnefors, C. (1989). Visual evoked potentials: relation to adult speechreading and cognitive function. *Journal of Speech, Language, and Hearing Research, 32*, 725-735.

Rönnberg, J., Lunner, T., Zekveld, A., Sorqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7*, 31.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47 Suppl 2*, S99-105.

Rothauser, E. H., Chapman, W. D., Guttman, N., Silbiger, H. R., Hecker, M. H. L., Urbanek, G. E., . . . Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions Audio and Electroacoustics, 17*, 225-246.

Rudner, M., Foo, C., Rönnberg, J., & Lunner, T. (2009). Cognition and aided speech recognition in noise: specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scandinavian Journal of Psychology, 50*, 405-418.

Shehorn, J., Marrone, N., & Muller, T. (2017). Speech perception in noise and listening effort of older adults with nonlinear frequency compression hearing aids. *Ear and Hearing, Epub ahead of print*.

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

Souza, P. (2002). Effects of compression on speech acoustics, intelligibility and speech quality. *Trends in Amplification, 6*, 131-165.

Souza, P. (2016). Speech perception and hearing aids. In G. R. Popelka, B. C. J. Moore, R. R. Fay & A. Popper (Eds.), *Hearing Aids* (pp. 151-180). Switzerland: Springer International Publishing.

Souza, P., & Arehart, K. H. (2015). Robust relationship between reading span and speech recognition in noise. *International Journal of Audiology, 54*, 705-713.

Souza, P., Arehart, K. H., Kates, J. M., Croghan, N. B., & Gehani, N. (2013). Exploring the limits of frequency lowering. *Journal of Speech, Language, and Hearing Research, 56*, 1349-1363.

Souza, P., Arehart, K. H., Shen, J., Anderson, M., & Kates, J. M. (2015). Working memory and intelligibility of hearing-aid processed speech. *Frontiers in Psychology, 6*.

Souza, P., & Sirow, L. (2014). Relating working memory to compression parameters in clinically fit hearing aids. *American Journal of Audiology, 23*, 394-401.

Souza, P., Wright, R., & Bor, S. (2012). Consequences of broad auditory filters for identification of multichannel-compressed vowels. *Journal of Speech, Language and Hearing Research, 55,* 474-486.

Wiley, T. L., Cruickshanks, K. J., Nondahl, D. M., Tweed, T. S., Klein, R., & Klein, B. E. (1996). Tympanometric measures in older adults. *Journal of the American Academy of Audiology, 7*, 260-268.

Won, J.-H., Drennan, W., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology, 8*, 384-392.

Zahorian, S. A., & Rothenberg, M. (1981). Principal-components analysis for low-redundancy encoding of speech spectra. *Journal of the Acoustical Society of America, 69*.

Figures

Figure 1. Left and right ear audiograms for the test group. The thick dark line shows the group mean.

Figure 2. Distribution of hearing loss (expressed as the average of .5, 1, 2, and 3 kHz in the right ear; left ear was similar) as a function of participant age. Each data point represents a single study participant.

Figure 3. NAL-NL2 prescribed (dashed lines) and measured (solid lines) real ear aided response (REAR) for each ear and processing condition, averaged across all study participants. From top to bottom of each panel the lines show data for 75, 65, and 55 dB SPL input levels, respectively. The roll-off of the measured high-frequency REAR in the “strong” processing reflects the expected effect of frequency compression.

Figure 4. Distribution of working memory (expressed as percent correct words correctly repeated during the Reading Span test) as a function of participant age. Each data point represents a single participant.

Figure 5. Metric values (cepstral correlation) for the mild and strong signal processing conditions for each of the presented signal to noise ratios, shown as a function of the listener’s 4-frequency (.5, 1, 2, 3 kHz) pure-tone average. Although aided speech recognition in quiet was not measured, the metric for quiet speech is shown for information purposes. Each data point shows the metric difference for the right ear of an individual participant (left ear was similar).

Figure 6. Aided speech recognition for each fitting as a function of signal-to-noise ratio. Boxes show the interquartile range. The middle line of each box shows the median value.

Whiskers show 1.5 times the interquartile range. Small circles and asterisks indicate values that extend outside the whiskers by more than 1.5 and 3 times the interquartile range, respectively.

Figure 7. Relationship between marginal predicted probabilities of correctly repeating a sentence and *age*, at different levels of signal modification (mild vs strong) and SNR (0, 5, 10 dB). All other covariates (reading span score, pure-tone average, and test session) were held constant, set to their mean or respective reference group.

Figure 8. Relationship between marginal predicted probabilities of correctly repeating a sentence and *reading span score*, at different levels of signal modification (mild vs strong) and SNR (0, 5, 10 dB). All other covariates (age, pure-tone average, and test session) were held constant, set to their mean or respective reference group.

Figure 9. Relationship between marginal predicted probabilities of correctly repeating a sentence and *pure-tone average*, at different levels of signal modification (mild vs strong) and SNR (0, 5, 10 dB). All other covariates (reading span score, age, and test session) were held constant, set to their mean or respective reference group.