Published in final edited form as:

Ear Hear. 2015 ; 36(5): 505–516. doi:10.1097/AUD.0000000000000173.

Relationship between signal fidelity, hearing loss and working memory for digital noise suppression

Kathryn Arehart¹, Pamela Souza², James Kates¹, Thomas Lunner^{3,4}, and Michael Syskind Pedersen⁵

¹Speech Language and Hearing Sciences, University of Colorado Boulder, 409 UCB, Boulder, CO 80309

²Communication Sciences and Disorders and Knowles Hearing Center, Northwestern University, Frances Searle Building 2-265, 2240 Campus Drive, Evanston, IL 60208, USA

³Eriksholm Res. Ctr., Oticon A/S, Linköping, Sweden Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

⁴Linnaeus Centre HEAD, Department of Behavioural Sciences and Learning, Linköping University, Sweden

⁵Oticon A/S, Smørum, Denmark

Abstract

Objectives—The present study considered speech modified by additive babble combined with noise-suppression processing. The purpose was to determine the relative importance of the signal modifications, individual peripheral hearing loss, and individual cognitive capacity on speech intelligibility and speech quality.

Design—The participant group consisted of 31 individuals with moderate high-frequency hearing loss ranging in age from 51 to 89 years (mean = 69.6 years). Speech intelligibility and speech quality were measured using low-context sentences presented in babble at several signal-to-noise ratios. Speech stimuli were processed with a binary mask noise-suppression strategy with systematic manipulations of two parameters (error rate and attenuation values). The cumulative effects of signal modification produced by babble and signal processing were quantified using an envelope-distortion metric. Working memory capacity was assessed with a reading span test. Analysis of variance was used to determine the effects of signal processing parameters on perceptual scores. Hierarchical linear modeling was used to determine the role of degree of hearing loss and working memory capacity in individual listener response to the processed noisy speech. The model also considered improvements in envelope fidelity caused by the binary mask and the degradations to envelope caused by error and noise.

Results—The participants showed significant benefits in terms of intelligibility scores and quality ratings for noisy speech processed by the ideal binary mask noise-suppression strategy.

Address for correspondence: Kathryn H. Arehart, University of Colorado, Department of Speech, Language, and Hearing Sciences, 409 UCB, Boulder, CO 80309, USA. kathryn.arehart@colorado.edu.

³The 0% error; 100-dB attenuation condition had an average quality rating of 0.567, so –6 dB SNR was used as a conservative lower boundary.

This benefit was observed across a range of signal-to-noise ratios and persisted when up to a 30% error rate was introduced into the processing. Average intelligibility scores and average quality ratings were well-predicted by an objective metric of envelope fidelity. Degree of hearing loss and working memory capacity were significant factors in explaining individual listener's intelligibility scores for binary mask processing applied to speech in babble. Degree of hearing loss and working memory capacity did not predict listeners' quality ratings.

Conclusions—The results indicate that envelope fidelity is a primary factor in determining the combined effects of noise and binary mask processing for intelligibility and quality of speech presented in babble noise. Degree of hearing loss and working memory capacity are significant factors in explaining variability in listeners' speech intelligibility scores but not in quality ratings.

INTRODUCTION

Signal processing algorithms for hearing aids are typically aimed at improving the audibility of the target signal. However, the signal modifications caused by those algorithms may also introduce unwanted distortion. Investigators have posited that understanding the tradeoffs between benefits and costs of hearing-aid signal processing is an important component of optimizing hearing-aid fittings for individual patients (e.g., Edwards 2007; Moore 2008; Rudner & Lunner 2013). One approach to analyzing tradeoffs is to investigate the relationship between listener characteristics and response to hearing-aid signal processing. For example, listeners with high working memory capacity benefit more than those with low working memory capacity from fast-acting wide-dynamic range compression, especially in modulated noise (Gatehouse et al. 2003, 2006a,b; Lunner & Sundewall-Thorén 2007; Ohlenforst et al. 2014). Arehart et al. (2013) showed that working memory, age and hearing loss all played a role in how listeners respond to frequency compression of noisy speech.

Designed to improve perception of speech that occurs in the presence of competing sounds, noise suppression is another hearing-aid signal processing algorithm that may be affected by individual listener factors. For example, Ng et al. (2013) reported that older listeners with hearing loss differed in their response to an implementation of a binary mask noise suppression algorithm according to their working memory capacity. That is, for listeners with higher working memory capacity, recall of words in sentences presented in babble backgrounds was significantly better when processed with a binary mask noise suppression algorithm compared to no noise suppression. Listeners with lower levels of working memory did not show a significant benefit from the noise suppression processing. A recent study by Neher et al. (2014) considered the effects of hearing loss and working memory on response to noise suppression for three-levels of a binaural-coherence strategy (none, moderate and strong). Results showed that intelligibility scores were unchanged or even worsened for speech processed with the noise suppression algorithm and that these effects were not significantly different for subjects with greater or lesser amounts of hearing loss or working memory capacity. However, for speech quality, working memory was a significant factor in that listeners with less working memory capacity preferred more aggressive (strong) noise suppression compared to moderate noise suppression. Finally, Healy et al. (2013) recently reported a trend for listeners with greater degrees of hearing loss to show

larger intelligibility improvements for noisy speech processed with a binary-masking noise suppression algorithm.

Taken together, the results of these recent studies of noise suppression and individual factors suggest that both working memory capacity and degree of hearing loss may play a role in perceptual responses to noise suppression processing. However, the results remain discrepant. Possible reasons for the discrepancy include differences in the experimental design, the range of signal and noise conditions considered in each experiment, and the ways in which the different noise-suppression algorithms influence the speech stimuli.

One way in which noise suppression modifies the signal is in changes to the envelope. Most noise-suppression algorithms apply a time-varying gain to the signal, with high gains for portions of the signal having a high signal-to-noise ratio (SNR) and reduced gains for the noisier portions of the speech (Kates 2008). Noise modifies the signal envelope in several ways. The noise fills in the valleys of the speech, thus reducing the modulation depth and the magnitude of the envelope modulation spectrum (Houtgast and Steeneken, 1985). Noise also introduces spurious envelope modulations unrelated to the speech that can interfere with intelligibility (Noordhoek & Drullman, 1997; Stone et al., 2011). Noise-suppression algorithms reduce the amplitude of the noisy portions of the speech-plus-noise signal, thus increasing the modulation depth of the noisy speech and partially restoring the envelope modulations to more closely follow those of the original noise-free speech. Several published models have related changes in the signal envelope to reductions in speech intelligibility and quality (e.g. Goldsworthy & Greenberg 2004; Huber & Kollmeier, 2006; Taal et al. 2011; Kates & Arehart 2014a). The purpose of the present study was to quantify the cumulative effects to the envelope in the case of noise-suppression processing and then relate these envelope effects to individual variation in intelligibility scores and quality ratings.

The experimental approach used in this study was to consider speech intelligibility and speech quality for low-context sentences mixed with babble noise over a wide range of SNRs. Noise suppression was implemented using an ideal binary mask processing algorithm (Kjems et al. 2009) which started with exact knowledge of the speech and noise signals. The separate speech and noise signals were each processed through a filter bank and then divided into short segments to produce time-frequency cells. The intensity of the speech in each cell was compared to that of the noise. If the SNR was greater than a pre-set threshold (or criterion), the cell was assumed to be primarily speech and the gain for that cell was set to 1. If the SNR was below the threshold, the cell was assumed to be noise and was attenuated by a specified amount.

Noise suppression with ideal binary mask processing results in improvement in intelligibility in both listeners with normal hearing and in listeners with hearing loss (e.g., Li & Loizou 2008; Wang et al. 2009). The size of the intelligibility improvements with ideal binary mask processing were consistently large compared to unprocessed noisy speech and tended to be greatest for processing thresholds between -20 dB and 0 dB SNR (e.g., Li & Loizou 2008) and for infinite attenuation compared to smaller amounts of attenuation (e.g., Brons et al. 2012). Sound quality, on the other hand, may be adversely affected by ideal binary mask

processing (Anzalone et al. 2006; Brons et al. 2012), especially for more aggressive levels of processing. Brons et al. (2012) reported that normal hearing listeners showed reduced overall preference and reduced perception of speech naturalness for ideal binary mask processing with infinite attenuation compared to 10 dB attenuation.

Degradation of the ideal binary mask processing significantly reduces speech intelligibility benefits. For example, Li and Loizou (2008) showed degraded intelligibility of processed noisy speech when two types of errors were intentionally introduced into the processing: a cell containing speech was attenuated instead of having unit gain (false negative), or a cell containing noise was preserved at full amplitude rather than attenuated (false positive). False positive errors had a greater impact on intelligibility than false negative errors. The errors in the Li and Loizou (2008) were independent; that is, there was no relationship between the error introduced in one cell and the error that might occur in another cell. However, intelligibility is reduced when the errors in adjacent cells are correlated instead of independent (Kressner and Rozell, 2014). Noise-suppression methods used in commercial devices are forced to estimate the noise intensity from the noisy speech, and the resulting estimate of the local SNR will be imperfect, leading to errors in the noise-suppression decisions. A simple first-order approximation is to introduce independent errors into the processing, while realizing that real-world noise-suppression performance may be worse at the same error rate. More practical noise-suppression methods such as spectral subtraction, while generally showing little to no improvement in intelligibility, do show significant improvements in speech quality (e.g., Arehart et al. 2003; Brons et al. 2012) and in ease of listening (Sarampalis et al. 2009).

In the present study, increasing error and attenuation decreased signal envelope fidelity, which was quantified using an envelope-based metric (Kates & Arehart, 2014a) based on the principle components of the short-time spectral variations in speech (Zahorian & Rotenburg 1981). A continuum of envelope fidelity was created by varying the accuracy of the ideal binary mask by introducing error and by varying the amount by which noisy cells are attenuated. The ideal binary mask with errors thus allows for the examination of the tradeoffs between restoration of the signal envelope due to binary mask processing and degradations of the signal envelope due to noise and imperfectly restored signal. The working hypothesis is that listener responses to the cumulative changes in envelope fidelity will depend on listener characteristics such as degree of hearing loss and working memory capacity.

METHOD

Listeners

The participants consisted of 31 individuals with moderate high-frequency hearing loss (Figure 1). Of this group, 17 were unaided, 13 wore bilateral hearing aids, and 1 wore a unilateral hearing aid. The mean age of the group was 69.6 years (range 51–89 years). All listeners had mild to moderately-severe sensorineural hearing loss, with air-bone gaps of 10 dB or less at octave frequencies from 0.5–4 kHz and normal tympanometric peak pressure and static admittance in both ears (Wiley et al. 1996). The correlation between age and high-frequency pure tone average (HFPTA) (500, 1000, 2000, 4000 Hz) was not significant

(Pearson correlation = 0.179; $p=0.336$). Testing was monaural. Typically, the better ear was chosen as the test ear and all results presented here refer to that ear. All listeners spoke English as their first or primary language. The listeners were administered the Mini-Mental Status Exam (MMSE) (Folstein, Folstein, & McHugh 1975) and all passed the test with a score of at least 26¹. The data were collected both at University of Colorado and at Northwestern University, with comparable equipment and protocols used at both sites. Local Institutional Review Boards reviewed and approved the human subjects' protocols. Listeners were reimbursed for their time.

Materials

Speech—The stimuli for the speech tasks were low-context sentences taken from the IEEE corpus (Rothauser et al. 1969). The sentences for both the intelligibility and quality tasks were spoken by a female talker and were first digitized at a 44.1 kHz sampling rate and then downsampled to 22.05 kHz. In an effort to represent conversational speech, the level of the sentences at the input to the hearing-aid simulation was set at 65 dB SPL. As described below, the actual presentation level for each subject was based on individualized amplification to compensate for elevated thresholds.

Noise—Noisy speech was created by digitally combining the sentence stimuli with four-talker babble. The SNRs ranged from -18 dB to $+12$ dB for intelligibility and -6 to $+12$ dB for quality (see Table 1). These SNRs were selected because they captured performance range from low to high intelligibility and from poor to excellent sound quality. For each SNR, the sentences were set to a level of 65 dB SPL and the babble level adjusted accordingly.

Signal Processing

Noise Suppression—The mixed speech-plus-noise stimuli were processed with a binary mask noise-suppression strategy (Kjems et al. 2009). The target speech signal, the masker signal and the speech-plus-noise mixture were separately converted into the frequency domain by analysis filterbanks consisting of 64 Gammatone filters (Patterson et al. 1995) with center frequencies equally distributed on the ERB scale (Glasberg & Moore, 1990) in the frequency range between 50 and 8000 Hz. The processing was done in time frames, each of which had a duration of 20 ms with an overlap of 10 ms resulting in frame shifts every 10 ms.

In each time-frequency unit, the local SNR was determined by comparing the intensities of the separate target and noise signals. The local SNR was then compared to a local criterion (LC) of 0 dB, resulting in an ideal binary mask decision equal to 1 if the local SNR was above the LC and 0 otherwise. The data of Kjems et al. (2009) indicate that an LC of 0 dB is most effective for SNRs in the range of approximately $+5$ to -10 dB. As the SNR is reduced below 0 dB, the LC of 0 dB causes more of the cells to be attenuated, and in the limit of a strongly negative SNR the entire signal is attenuated. Similar to the procedure in Li and Loizou (2008), errors were introduced into the ideal binary mask by randomly flipping a

¹One subject did not have MMSE administered.

certain percentage (0%, 10%, 30%) of the time-frequency units either from 0 to 1 or from 1 to 0. The binary patterns were converted into gain values, where values of 1 were converted into 0 dB gain and values of 0 were converted into an attenuation of either 10 dB or 100 dB. The noisy speech signal was then multiplied by the binary gain values to give the processed signal in the frequency domain. The processed signal was synthesized back into the time domain by use of a time-reversed Gammatone filterbank, thereby ensuring the same group delay in all frequency bands. The total ripple in the analysis/synthesis system was less than 0.1 dB.

Amplification—To assure audibility, the noisy sentences for each listener were subjected to the customized linear amplification prescribed by the National Acoustic Laboratory-Revised (NAL-R) algorithm (Byrne & Dillon, 1986) with the gain implemented using a 128-point linear-phase FIR digital filter.

Procedures

Stimulus presentation—The digitized sentences were first passed through a digital-to-analog converter (TDT RX6 or RX8), through an attenuator (TDT PA5), to a headphone buffer (TDT HB7), and then finally were presented to each listener (seated in a double-walled soundbooth) monaurally through a Sennheiser HD 25-1 earphone. Responses were collected using a monitor and computer mouse.

Speech intelligibility response task—Each trial consisted of a participant listening to and repeating a sentence randomly drawn from one of the 42 test conditions shown in Table 1 and described in more detail below. Subjects first heard 42 practice sentences (1 from each test condition) and then listened to 420 test sentences (with 10 sentences in each of the 42 conditions). No feedback was provided. The order of sentences and conditions was randomized across listeners. Scores were calculated based on the proportion of correctly-identified key words (10 sentences per condition and 5 words per sentence for 50 words per condition, per participant). Scoring was completed by the experimenter seated outside the sound booth. The proportion correct scores were transformed to rationalized arcsine units (RAU) (Studebaker 1985) for statistical analysis.

Speech quality response task—Following the intelligibility tests, listeners rated the quality of the processed speech. On each trial, listeners heard one presentation of two sentences (“Take the winding path to reach the lake. A saw is a tool used for making boards.”) spoken by a female talker and processed with one condition, randomly chosen from the 29 conditions listed in Table 1. The same sentences were used each time to avoid confounding effects of intelligibility, and allow for the desired focus on quality. Sound quality is multidimensional in nature (Arehart et al. 2007; Gabrielsson et al. 1988). However, specific quality aspects of speech processed by hearing aid signal processing algorithms are well predicted by metrics using a single “overall quality” rating scale (Arehart et al. 2010; Arehart et al. 2011; Kates & Arehart, 2010; 2014b). Accordingly, listeners rated the overall sound quality using a rating scale which ranged from 0 (poor sound quality) to 10 (excellent sound quality) (ITU 2003). The rating scale was implemented with a slider bar that registered responses in 0.05 increments. Listeners made

their selections from the slider bar displayed on the computer screen using a customized interface that included a point-and-click method for recording and verifying rating scores. The timing of presentation was controlled by the subject. Listeners rated one practice block followed by four test blocks. Each practice and test block contained 29 two-sentence presentations. No feedback was provided.

Factors and Groupings Used in the Statistical Analysis

Envelope fidelity metric—Changes in the signal caused by the noise and noise-suppression processing were measured using an existing envelope fidelity metric (Kates & Arehart, 2014a). The calculation started with an auditory model (Kates 2013) that included the filters, threshold, and dynamic-range compression found in a normal ear. The envelope signals at the output of the model were converted to dB sensation level. The envelopes in each frequency band were smoothed using 16-msec raised-cosine windows having a 50 percent overlap which were then resampled at 125 Hz, and a smoothed auditory spectrum was computed at each time sample. The smoothed spectra from the processed and reference signals were cross-correlated to produce a measure of envelope fidelity termed cepstral correlation (Kates & Arehart 2010; 2014a,b) which was related to the basic time-frequency modulation patterns of speech (Zahorian & Rothenberg 1981).

Perfect reproduction of the signal envelope produced a cepstral correlation value of 1, while envelopes that were completely unrelated produced a cepstral correlation value of 0. Additive noise reduced the time-frequency envelope fidelity by filling in the valleys of the signal envelope measured over time within one frequency band and by also reducing the spectral contrast measured across frequency bands at a specific time. The ideal binary mask can improve envelope fidelity in both time and frequency by reducing the amplitude of the signal cells that have been filled in by the noise. Errors in the binary mask, by either attenuating a cell that primarily contains the desired signal or failing to attenuate a cell containing that primarily contains noise, will reduce the envelope fidelity and hence the computed cepstral correlation.

Hearing loss and working memory groups—Listeners were categorized into two hearing loss groups (Mild HFPTA and Mod-severe HFPTA) using the median HFPTA (37.5 dB HL). Listeners were also categorized into two working-memory groups based on scores on the Reading Span Test (RST) (Daneman & Carpenter 1980; Rönnberg et al. 1989). During the RST, participants were presented with a series of sentences displayed on a computer screen and then were tasked with recalling (in correct order) the first or last words of the sentences. The instruction of whether to repeat the first or last words was provided after each sentence sequence. The scores from the RST were calculated based on the proportion of words that were correctly repeated, whether or not the sequential order was correct. The participants' RST scores ranged from 0.13 to 0.63. The two working-memory groups (High RST and Low RST) were determined by using the median RST score (38.9) as the cut-off criterion. Note, HFPTA was not significantly different between the two RST groups ($df=29$, $t=0.853$, $p=0.401$). Similarly, the RST scores were not significantly different between the HFPTA groups ($df=29$, $t=0.119$, $p=0.906$).

Statistical Analysis

Analysis of Variance (ANOVA)—The statistical analysis included two phases. The first phase utilized a repeated-measures ANOVA to address the first aim of characterizing the relative benefit of noise suppression on speech intelligibility and sound quality across a range of processing conditions and SNRs. This initial analysis facilitated comparison of the present results to previous studies and also served as a baseline for the analysis presented below regarding individual listener factors. The ANOVA included three within-subject factors (SNR, attenuation, error). The specific research questions addressed by the phase 1 analysis were as follows:

1. Do the mean intelligibility scores and quality ratings change as a function of SNR, binary mask attenuation and/or binary mask error?
2. What is the relative effect size of binary mask processing on mean intelligibility and quality relative to no processing?

Multi-level Model—This phase of the analysis specifically addressed the second aim of this study which was to determine whether the relationship between envelope fidelity and listener performance on intelligibility and quality tasks differed based on hearing loss (higher and lower HFPTA) and working memory (higher and lower RST scores). The specific analytical strategy used for phase 2 was multi-level modeling.

Multi-level modeling is considered a generalization of regression methods, but it is specifically designed for the analysis of repeated measures. In the literature, it is also referred to as hierarchical linear modeling, mixed models, random coefficient model, and longitudinal modeling (Fitzmaurice et al. 2004; Littell et al. 2006; Rogosa & Saner 1995; Singer & Willett 2003; Snijders & Bosker 1999). For the current data, the model was specified hierarchically using two levels (Level-1 and Level-2), and can be expressed symbolically as the predicted intelligibility/quality score (Y_{ij}) of listener i at condition j as a function of envelope fidelity (F_{ij}) and as a function of hearing loss ($HFPTA_i$), working memory (RST_i), and age (Age_i) with the following equation:

$$Y_{ij} = P_{0i} + P_{1i} * (F_{ij}) + e_{ij}, \quad (1)$$

where

$$P_{0i} = B_{00} + B_{01} * (HFPTA_i) + B_{02} * (RST_i) + B_{03} * (Age_i) + r_{0i} \quad (2)$$

$$P_{1i} = B_{10} + B_{11} * (HFPTA_i) + B_{12} * (RST_i) + B_{13} * (Age_i) + r_{1i} \quad (3)$$

Equation 1 represents the Level-1 model where intercept, P_{0i} , corresponds to the expected value of intelligibility/quality when F_{ij} equals zero, and the slope, P_{1i} , represents the rate of change in intelligibility/quality as fidelity increases. The coefficient, e_{ij} corresponds to the within-subject residual variance and the coefficients (r_{0i}, r_{1i}) represent the between-subject residual variance in intercept and slope. In order to facilitate the interpretation of the intercept coefficient, the fidelity metric was centered by subtracting its mean from the raw value. Centering the fidelity variable made the intercept interpretable as the value of the

outcome variable at the mean value of fidelity (0.5). Centering is a common strategy in multi-level modeling when the level-1 predictor at zero is not substantively meaningful. The adjusted Eq. (1) was written as follows:

$$Y_{ij} = P_{0i} + P_{1i} * (F - 0.5)_{ij} + e_{ij}, \quad (4)$$

Equation 2 and Equation 3 are the Level-2 full conditional models for intercept and slope. The beta coefficients (B_{ij}) represent the strength of the relation between intercept and slope with hearing loss (B_{01} , B_{11}), working memory (B_{02} , B_{12}), and age (B_{03} , B_{13}). Although the full conditional model is represented in Equations 2–3, there were three conditional models that were evaluated. The first, Model A, included the HFPTA group independently (not shown), Model B included HFPTA and RST (not shown), and Model C (Equation 2–Equation 3) included all predictors HFPTA, RST, and Age. The Level-1 and Level-2 models can be represented by a single equation by substituting equations 2–3 into equation 1 as follows:

$$\begin{aligned} Y_{ij} = & B_{00} + B_{01} * (HFPTA_i) \\ & + B_{02} * (RST_i) \\ & + B_{03} * (Age_i) + B_{10} * (F - 0.5)_{ij} + B_{11} * (HFPTA_i) \\ & * (F - 0.5)_{ij} \\ & + B_{12} * (RST_i) \\ & * (F - 0.5)_{ij} \\ & + B_{13} * (Age_i) \\ & * (F - 0.5)_{ij} \\ & + r_{0i} + r_{1i} * (F - 0.5)_{ij} + e_{ij} \end{aligned} \quad (5)$$

A formula suggested by Kreft and de Leeuw (1998) and Singer (1998) was used to evaluate the goodness of fit of the multi-level models evaluated in this analysis. The formula provides a pseudo R-square value of the total explainable variance that can be explained by the model by comparing the error terms in the unconditional model (no predictor variables) with the error terms of the conditional model(s) (one or more predictor variables).

The multi-level framework for the current analysis provided a statistical test for evaluating both within and between group differences in listener performance on intelligibility and quality tasks, and it provided a method of measuring whether the relation between speech perception and envelope fidelity depended on HFPTA, RST, and Age (Raudenbush & Bryk 2002). As such, the analytical strategy was particularly suited for the following research questions:

1. What are the predicted effects of degree of hearing loss, working memory capacity, and age on listener's intelligibility/quality scores, and do the strength of these effects change across the level of envelope fidelity scale?
2. Do degree of hearing loss, working memory capacity, and age relate to the rate of change among listener's intelligibility/quality scores across envelope fidelity?

3. How much variability in intelligibility/quality scores and their rate of change across envelope fidelity can be explained by degree of hearing loss, working memory capacity, and age?

RESULTS

Intelligibility ANOVA

Figure 2 shows the average intelligibility scores (in RAU units) across the 31 participants for each of the 6 SNRs for unprocessed noisy speech and for noisy speech processed with ideal binary mask using combinations of three error rates (0%, 10%, 30%) and two attenuation values (10 dB and 100 dB). The repeated-measures ANOVA used three within-subject factors (SNR, attenuation, and error). The ANOVA results indicated significant effects of SNR, attenuation and error. In addition, all interactions were found to be significant (Table 2). Pairwise comparisons with Bonferroni adjustments showed that the two attenuation conditions were significantly different from each other after controlling for error and SNR ($p < 0.05$). Pairwise comparisons with Bonferroni adjustments showed the three error conditions were significantly different from each other ($p < 0.05$) after controlling attenuation and SNR. Finally, pairwise comparisons with Bonferroni adjustments showed that all six SNRs were significantly different from each other ($p < 0.001$) after controlling for attenuation and error.

Sigmoidal fits (average $R^2 = 0.95$) to the intelligibility functions showed that the SNR for 50% intelligibility was greatest for the unprocessed condition (2.1 dB SNR) and was lowest for the 0% error, 100-dB attenuation condition (-6.3 dB SNR). The dB SNR change required to maintain 50% correct intelligibility (50 RAU) for each processing condition was used to quantify the amount of benefit of the ideal binary mask. The effective improvements in dB SNR for 50% intelligibility relative to the unprocessed condition ranged from 8.4 for the 0% error, 100-dB attenuation condition to 2.3 for the 30% error, 10-dB attenuation condition (Table 3).

Intelligibility Multi-level Model

Model fit—Hierarchical Linear Modeling procedures facilitate the exploration of linear relationships. Following the transformation of intelligibility scores from proportion correct to RAU units, mean intelligibility in RAU was considered as a function of cepstral-based envelope fidelity, and the linear-fitted trajectory was found to have an R^2 value of 0.91 (Figure 3). The R^2 values for individual linear plots ranged from 0.87 to 0.91. Higher order polynomial fits were also considered, but only improved R^2 to 0.92 for the data in Figure 3. Given that a linear model required fewer dimensions, the transformed intelligibility scale was used here for the application and interpretation of linear model techniques.

Variability in performance—The average estimate of speech intelligibility in RAUs at 0.5 envelope fidelity was 43.11 (SD=11.38), and the average estimate for the rate of change in speech intelligibility across the envelope fidelity continuum was 27.2 (SD=2.28) RAUs per envelope fidelity unit. The predicted 95% range of intelligibility scores was calculated to be 20.35 to 65.87 for intercept and 22.68 to 31.81 for slope. This range represented

statistically significant between-listener variability in intercept ($p < 0.001$) and slope ($p < 0.001$). Three conditional models (see methods) were evaluated in an effort to explain this between-listener variability.

Table 4 outlines the main effects and variability in intercept and slope explained for each of the conditional hierarchical models for intelligibility. The results for Model A demonstrated a significant main effect for HFPTA group on intercept and slope. HFPTA independently explained 64% of the variability in intercept and 51% of the variability in slope. Model B demonstrated that after controlling for HFPTA, RST group was also significant predictor for intercept. Working memory capacity explained an additional 5% of the variability in intercept for a total of 69%. However, RST group was not a significant predictor of slope when controlling for HFPTA. Lastly, Age was added in Model C. After accounting for HFPTA and RST, there was not a significant main effect for Age for either intercept or slope. Because Age was not a significant factor in either model for intercept or slope, it was not included in subsequent analyses. The omission of Age did not change the main effects for HFPTA and RST groups. The adjusted Eqs. (2) and Eqs. (3) were rewritten as Eq. (6) and Eq. (7):

$$P_{0i} = B_{00} + B_{01} * (HFPTA)_i + B_{02} * (RST)_i + r_{0i} \quad (6)$$

$$P_{1i} = B_{10} + B_{11} * (HFPTA)_i + B_{12} * (RST)_i + r_{1i} \quad (7)$$

The adjusted level-1 and level-2 models can then be represented as a single composite model as defined by Eq. (8).

$$Y_i = B_{00} + B_{01} * (HFPTA)_i + B_{02} * (RST)_i + B_{10} * (F - 0.5) + B_{11} * (HFPTA)_i * (F - 0.5) + B_{12} * (RST)_i * (F - 0.5) + r_0 + r_1 * (F - 0.5) + e_1 \quad (8)$$

The adjusted model can be interpreted in the following way. Intelligibility Y_i is a function of the overall intercept (B_{00}), the main effect of HFPTA (B_{01}), the main effect for RST group (B_{02}), and two cross-level interactions (B_{11} ; B_{12}) involving HFPTA group and RST group with envelope fidelity, respectively.

Table 5 summarizes the results for the multi-level model represented in Eq (8). The listeners with milder hearing loss (HFPTA < 37.5 dB) had an estimated intelligibility score that was 18.33 RAU units higher at 0.5 envelope fidelity than listeners in the moderate-severe HFPTA group ($B_{01} = 18.33$; $t = -7.74$, $p < 0.001$). There was also a significant difference in intercept between RST groups. Listeners in the High-RST group had significantly higher intelligibility scores than listeners in the Low-RST group ($B_{02} = 2.68$, $t = 2.15$, $p = 0.04$). The estimated effect size for listeners in the High-RST group was a 2.68 RAU gain when compared to the Low-RST group.

There was one significant interaction in the adjusted model for HFPTA group with envelope fidelity. This interaction demonstrated that, as envelope fidelity increased, listeners with milder hearing loss tended to have higher intelligibility scores compared to those with more hearing loss.

Figure 4 illustrates the final model for intelligibility (equation 8). It provides four different fitted trajectories of intelligibility as a function of envelope fidelity for all combinations of HFPTA and RST groups.

Quality ANOVA

Figure 5 shows average quality ratings (with standard error) across 30 participants² for each of the four SNRs for unprocessed noisy speech and for noisy speech processed with ideal binary mask using combinations of three error rates (0%, 10%, 30%) and two attenuation values (10 dB and 100 dB). Shown in Table 6, the results of the ANOVA indicated a significant effect of SNR, attenuation, and error. Significant interactions were found between attenuation and error and the three-way interaction between SNR, attenuation and error. Post-hoc comparisons with Bonferroni adjustments showed that attenuation conditions were significantly different from each other ($p < 0.05$), error conditions were significantly different from each other ($p < 0.05$) and all four SNRs were significantly different from each other ($p < 0.05$).

The dB SNR required for an average quality rating of 0.5 was used to quantify the benefit for quality ratings due to the ideal binary mask. Linear regression yielded accurate fits to the quality data and were therefore used to determine the SNR corresponding to 0.5 quality rating for each processing condition. Listed in Table 7, the effective improvements in dB SNR for a 0.5 quality rating relative to the unprocessed condition ranged from greater than 14.9 dB for the 0% error, 100-dB attenuation condition to -0.21 dB for the 30% error, 100-dB attenuation condition.

Quality Multi-level Model

Model fit—When quality was considered as a function of cepstral-based envelope fidelity for data averaged over all the subjects, the group fitted trajectory had an r -square value of 0.961 (Figure 6). Given such a high r -square, a linear model was determined to be an accurate representation of the relationship between the quality scale and envelope fidelity.

Variability in performance—The average estimate of sound quality ratings at 0.5 envelope fidelity was 0.50 ($SD = 0.11$), and the average estimate for the rate of change in quality ratings across the envelope fidelity continuum was 0.15 ($SD = 0.20$) points per fidelity unit. The predicted 95% range of quality rating was calculated to be 0.028 to 0.072 for intercept and 0.11 to 0.19 for slope. This range represented statistically significant between-listener variability in intercept ($p < 0.001$) and slope ($p < 0.001$), and three conditional models described in the methods section were evaluated.

Table 8 summarizes the results for the three conditional quality models. When HFPTA was included independently (Model A), it did not explain any variability in intercept or slope. Adding RST group (Model B) and Age (Model C) did not add any additional explanatory power. In other words, there were no significant main effects for degree of hearing loss,

²One of the 31 subjects was not able to complete the quality testing due to scheduling issues.

working memory capacity, and age on quality ratings. Table 9 provides the parameter estimates for the multi-level model (equation 8) in terms of quality.

DISCUSSION

Effects of Ideal Processing

Similar to other studies, the results of this study showed that ideal binary mask processing resulted in significant intelligibility improvements for listeners with hearing loss, with the largest improvements evident at -6 and 0 dB SNR. The average 8.5 dB improvement for the listeners in the 0% error, 100 -dB attenuation condition was comparable to the approximately 9 dB improvement in speech recognition thresholds for listeners with hearing loss reported by Wang et al. (2009) and by Anzalone et al. (2006) for sentences presented in speech-shaped noise with ideal binary mask processing with infinite attenuation, but was smaller than the 15.6 dB improvement reported by Wang et al. for sentences in cafeteria noise. Note that Wang et al. used an LC of -6 dB and that their interfering signal comprised a conversation between two talkers. This interfering signal would be expected to have a larger proportion of low-intensity cells than the multi-talker babble used in the present study. Their interference, when combined with their lower LC, would lead to a larger proportion of cells given a gain of 1 in their experiment in comparison with ours.

The present results also showed that the ideal binary mask improved quality perception: the participants rated noisy speech subjected to ideal mask processing significantly higher compared to no processing. Other studies explicitly addressing sound quality with ideal binary masking were more limited and less clear cut. Using ideal binary mask processing at -4 dB and $+4$ dB SNR with infinite attenuation, Brons et al. (2012) studied quality ratings for normal-hearing listeners and reported that processing resulted in less noise annoyance and less listening effort, but also worse speech naturalness and lower preference scores. When Brons et al. (2012) decreased the amount of attenuation to 10 dB using a tempered mask, perception of sound quality improved in terms of higher preference scores and higher ratings of speech naturalness. Anzalone et al. (2006) provided anecdotal evidence (based on listeners' informal reports) that normal-hearing listeners reported degraded sound quality for ideal binary mask processing but that hearing-impaired listeners did not.

The intelligibility and quality results presented here generally showed an advantage for the 100 -dB attenuation over the 10 -dB attenuation processing. Brons et al. (2012) also found an advantage for infinite attenuation over a procedure giving a maximum of 10 dB attenuation for intelligibility, but they found that the smaller amount of attenuation gave higher quality. It is not obvious *a priori* which amount of attenuation is preferable. The goal of the ideal binary mask processing is to match the envelope modulation of the processed noisy speech to that of the original noise-free signal. The 100 -dB attenuation will do a better job of matching the envelopes in the silent intervals between syllables, but the 10 -dB attenuation may do a better job in matching the reduced speech levels at the onset and offset of speech sounds such as bursts and fricatives. Thus, the highest quality may depend more on the details of the processing implementation (see below) than on the specific amount of noisy-signal attenuation.

Effects of Error

Both intelligibility and quality were adversely affected as the processing was deliberately degraded with errors. For example, intelligibility scores and quality ratings at -6 dB SNR showed less improvement due to binary mask processing for the 30% error condition compared to the 0% error condition. Li and Loizou (2008) also reported that intentionally-introduced error degraded the amount of intelligibility improvement provided by ideal binary mask processing. Their Figure 4 showed that with sentences presented with a two-talker babble at -5 dB SNR, normal-hearing listeners actually performed worse with 30% error compared to no processing. In the present study, the listeners with hearing loss had better intelligibility with 30% error compared to no processing. Hence, the listeners with hearing loss were actually able to tolerate more signal degradation in maintaining intelligibility improvement compared to the normal-hearing participants in Li and Loizou.

The independent errors used in this study are a first-order approximation to the more complicated error structure that occurs in real-world situations. A system embedded in a hearing aid will not have access to the separate speech and noise signals used in the ideal binary mask processing. Instead, the speech and noise level have to be estimated from the noisy speech signal, and errors will be made when classifying a time-frequency cell as being predominantly speech or predominantly noise. As shown by Kressner and Rozell (2014), errors in connected speech are partially correlated between neighboring cells. However, the independent errors used in this paper still give an indication of the interaction between speech intelligibility and cognition caused by spurious envelope modifications, and indicate the processing limitations that could be expected in devices such as hearing aids.

An example of ideal binary mask processing that could potentially be used in hearing aids was recently proposed by Healy et al. (2013). They used a neural network to classify time-frequency cells as being either speech or noise, with the neural network trained by comparing the classifier output to the ideal binary mask output for a set of noisy sentences. Their processing produced an accuracy rate, defined as hits (1s correctly assigned) minus false alarms (0s incorrectly classified as 1s), of about 80%. This accuracy corresponds to a random error rate of about 20% in the present study. They showed an improvement from 25 to 85 % correct for hearing-impaired listeners at an SNR of -2 dB. The data reported here, when interpolated to the same SNR and 20% error rate, showed similar behavior, with an improvement from about 20% to about 62%.

Comparisons across studies are limited by the fact that binary mask processing is implemented in different ways. The approach used in this paper and several others (Anzalone et al. 2006; Healy et al. 2013; Kjems et al. 2009; Wang et al. 2009) used a gammatone filterbank for the frequency analysis, while other implementations (Brons et al. 2012; Li & Loizou 2008) use a discrete Fourier transform (DFT) for the frequency analysis. In general, the DFT system gives poorer frequency resolution at low frequencies and higher resolution at high frequencies than a gammatone filterbank. The frequency analysis procedure may impact the ideal binary mask gain and error effects. At low frequencies, the DFT gain changes will span more than one auditory filter, and will thus create synchronized changes across auditory bands that could affect auditory stream segregation of the speech and noise (Grimault et al. 2002). At high frequencies, the DFT gain changes are narrower

than the auditory filters, so a change in a single cell will have a reduced impact on the overall loudness of the processed signal. Thus, even with similar amounts of attenuation and similar error rates, there could be subtle differences in the processing effects for the two different implementations.

Envelope-based Metric

A unique feature of this study is consideration of the cumulative effects on envelope fidelity caused by the binary mask processing, errors in the processing and additive noise. Additive noise modifies the short-time spectra by filling the valleys of the signals, reducing the temporal dynamic range and the spectral contrast of the noisy signal. The noise thus causes a reduction in the cross-correlation of the envelope variations of the noisy signal with those of the original signal. Noise suppression reduces the amplitude of the valley regions that have negative SNRs and thus restores the short-time spectral contrast to be closer to that of the original speech. However, the time-varying gain changes introduced by the noise suppression also apply amplitude modulation to the speech, resulting in modulation sidebands and an increase in distortion. The amount of distortion will in general be greatest for the largest gain changes. For example, at the onset of a burst leading into a vowel the speech goes from low intensity to high, so the SNR will cross the LC boundary and the noise-suppression gain will jump from attenuation to no suppression. Additional distortion for ideal binary mask errors will be generally associated with gain changes occurring for speech cells (false negatives) at positive SNRs and with noisy cells (false positives) at negative SNRs.

The envelope fidelity metric, despite showing excellent overall accuracy, also showed some systematic errors in predicting the subject intelligibility scores. In Figure 3, points below the diagonal line represent conditions for which the envelope metric predicted that intelligibility would be better than observed for the subjects, while for points above the line the predicted intelligibility is worse than observed for the subjects. The three points that fall furthest below the diagonal line are for the ideal binary mask processing with 100-dB attenuation and no decision errors for SNRs of -6 , -12 , and -18 dB. Used as a measure of signal fidelity, the envelope metric incorporated normal auditory threshold and was not adjusted for the shift in threshold associated with hearing loss. Because the LC was set to 0 dB, a large proportion of the cells for the negative SNRs were attenuated to below the threshold of audibility for the listeners with hearing loss and this loss of audibility was not reflected in the metric calculations. It is thus possible that the cells that remained after attenuation were sufficient to sketch out the shape of the short-time spectra used for the cepstral correlation calculation even though much of the speech was below the listeners' thresholds. Note that the negative SNRs with 10-dB attenuation did not produce the same systematic prediction errors, indicating that when most of the speech was above threshold the predictions were consistent with the listeners' scores.

The changes in the noise and distortion were both reflected in the speech envelope. The subject intelligibility scores and quality ratings, when averaged over the subjects, were accurately modeled by the envelope-based fidelity metric. The metric is equivalent to cross-correlating the variations in the smoothed short-time log-frequency spectra of the original

and processed noisy signals. A high value indicated that the spectral variations across time in the two signals match. Previous work in modeling speech intelligibility (Kates & Arehart 2014a) and speech quality (Kates & Arehart 2014b) has shown that envelope fidelity, particularly in reproducing the changes in the smoothed short-time spectrum from one segment to the next, is important for predicting both speech intelligibility and quality. The results of this experiment are consistent with these previous data.

The envelope metric was also sensitive to the different kinds of errors that occur in ideal binary mask processing. Li and Loizou (2008) divided errors into two classes. Their results showed that false positive errors caused a much greater reduction in speech intelligibility than false negative errors. A false positive error (0 changed to 1) would be expected to influence the envelope more than a false negative error (1 changed to 0). The inadvertent attenuation of a cell will tend to be smoothed over in extracting the envelope while an inadvertent peak will cause a noticeable increase in the envelope intensity. The accuracy of the envelope metric in modeling the subject intelligibility and quality scores indicates that the envelope metric effectively predicts the consequences of having both kinds of errors in the processed signal.

Effects of Hearing Loss and Working Memory

The accuracy of the metric in predicting average performance from the older listeners with hearing loss in this study indicates that changes in envelope fidelity are a primary factor in listeners' responses to the processed noisy speech. The present results also indicate that listeners' response to different amounts of signal modification caused by both noise and processing were significantly affected by degree of hearing loss and (to a lesser extent) working memory capacity. One factor that relates to degree of hearing loss is the sensation level of the amplified signal. Despite receiving customized amplification using a clinical fitting algorithm (Byrne & Dillon 1986), individual listeners in this study may only have had partial audibility restored. As Humes (2007) points out, cognitive factors may emerge as significant only when full audibility has been restored.

Other factors related to degree of hearing loss include deficits in supra-threshold spectro-temporal processing (Moore 2007). Listeners with greater degrees of hearing loss may have broader auditory filters (e.g., Laroche et al. 1992; Lutman et al. 1991). The auditory filter bandwidth may contribute to a smoothing of the internal spectral representation and thus reduce the effect of narrowband spectral changes introduced by the noise-suppression processing. In addition, listeners with greater degrees of hearing loss may be more affected by changes in envelope and temporal fine structure caused by the presence of noise and processing distortion (Anzalone et al. 2006; Arehart et al. 2013; Hopkins & Moore 2011; Lunner et al. 2012). Finally, older listeners may have reduced temporal fine structure processing (e.g., Grose & Mamo, 2010) and as such, may rely more on envelope cues and be more susceptible to changes to envelope fidelity.

The results of this study add to a varied literature on the relationship between listener characteristics and signal processing. Whereas degree of hearing loss and working memory capacity were significant factors for intelligibility scores in this study, Neher et al. (2014) did not find such a relationship. In addition, while working memory capacity and degree of

hearing loss did not explain variance on the quality rating task in this study, Neher et al. report that listeners with poorer working memory had a significant preference for stronger noise suppression compared to weaker noise suppression. Ng et al. (2013) reported a significant relationship between working memory and response to noise suppression in that listeners with better working memory capacity had more accurate recall of words with processing compared to no processing.

The mixed results with noise suppression may be due in part to the tasks used to quantify noise suppression benefit. While intelligibility and quality are important components to consider in noise suppression processing (Healy et al. 2013; Neher et al. 2014), listening effort and the ability to recall words in sentences are also important to consider in quantifying noise suppression benefit (Sarampalis et al. 2009; Ng et al. 2013). In addition, the effects of patient factors on response to signal processing may depend on the purpose of the hearing-aid signal processing. Wide-dynamic range compression and frequency compression both produce distorted signals where the nonlinear processing is intended to overcome one aspect of peripheral hearing loss. In wide dynamic range compression, the processing applies greater amplification to low-intensity portions of the speech signal than to the high-intensity portions, and thus, compensates for the elevated auditory threshold. In frequency compression, the processing shifts high frequencies lower, and thus, compensates for the reduced audibility of high-frequency speech sounds. In both cases, the listener with hearing loss is expected to extract useful speech information from an intentionally distorted signal, and working memory capacity appears to be an important factor in the ability of a listener to effectively process the distorted speech (e.g., Arehart et al. 2013). In contrast, working memory capacity may play a more limited role in ideal binary masking processing because its goal is to modify a noise signal so that its envelope more closely matches that of the original speech.

SUMMARY

The results indicate that envelope fidelity is a primary factor in determining the combined effects of noise and binary mask processing for intelligibility and quality of speech presented in babble noise. Degree of hearing loss and working memory capacity are significant factors in explaining variability in listeners' speech intelligibility scores but not in quality ratings.

ACKNOWLEDGMENTS

The authors thank Rosalinda Baca for substantive input with statistical analysis including the hierarchical linear modeling, Peggy Nelson for sharing speech materials, Laura Mathews, Katharine Miller, Efoe Femi Nyatepe-Coo, and Cory Portnuff for data collection, Ramesh Kumar Muralimanohar for software development and calibration, and Melinda Anderson and Katharine Miller for helpful discussions related to this work. A portion of these data was presented at the 2013 International Congress of Acoustics Montreal, Canada. (Proceedings of Meetings on Acoustics 19, 050084). This work was supported by National Institutes of Health grant R01 DC012289 (P.S. and K.A.) and by a grant to the University of Colorado by GN ReSound (K.A., JMK).

REFERENCES

Anzalone MC, Calandruccio L, Doherty KA, Carney LH. Determination of the potential benefit of time-frequency gain manipulation. *Ear Hear.* 2006; 27:480–492. [PubMed: 16957499]

- Arehart KH, Hansen JHL, Gallant S, Kalstein L. Evaluation of an auditory-masked-threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners. *Speech Commun.* 2003; 40:575–592.
- Arehart KH, Kates JM, Anderson MC. Effects of noise, nonlinear processing and linear filtering on perceived speech quality. *Ear Hear.* 2010; 31:420–436. [PubMed: 20440116]
- Arehart KH, Kates JM, Anderson MC, Harvey LO. Effects of noise and distortion on speech quality judgments in listeners with normal hearing and with hearing loss. *J Acoust Soc Am.* 2007; 122:1150–1164. [PubMed: 17672661]
- Arehart KH, Kates JM, Anderson MC, Moats P. Determining perceived sound quality in a simulated hearing aid using the international speech test signal. *Ear Hear.* 2011; 32:533–535. [PubMed: 21325947]
- Arehart KH, Souza P, Baca R, Kates JM. Working memory, age, and hearing loss: Susceptibility to hearing aid distortion. *Ear Hear.* 2013; 34:251–260. [PubMed: 23291963]
- Brons I, Houben R, Dreschler WA. Perceptual effects of noise reduction by time-frequency masking of noisy speech. *J Acoust Soc Am.* 2012; 132(4):2690–2699. [PubMed: 23039461]
- Byrne D, Dillon H. The National Acoustical Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear Hear.* 1986; 7:257–265. [PubMed: 3743918]
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *J Verbal Learn Verbal Behav.* 1980; 19:450–466.
- Edwards B. The future of hearing aid technology. *Trends Amplif.* 2007; 11:31–45. [PubMed: 17301336]
- Fitzmaurice, G.; Laird, N.; Ware, J. *Applied Longitudinal Analysis.* Hoboken, NJ: Wiley; 2004.
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975; 12:189–198. [PubMed: 1202204]
- Gabrielsson A, Schenkman BN, Hagerman B. The effects of different frequency responses on sound quality judgments and speech intelligibility. *J Speech Lang Hear Res.* 1988; 31:166–177.
- Gatehouse S, Naylor G, Elberling C. Benefits from hearing aids in relation to the interaction between the user and the environment. *Int J Audiol.* 2003; 42:S77–S85. [PubMed: 12918613]
- Gatehouse S, Naylor G, Elberling C. Linear and nonlinear hearing aid fittings-1: Patterns of benefit. *Int J Audiol.* 2006a; 45:130–152. [PubMed: 16579490]
- Gatehouse S, Naylor G, Elberling C. Linear and nonlinear hearing aid fittings-2: Patterns of candidature. *Int J Audiol.* 2006b; 45:153–171. [PubMed: 16579491]
- Glasberg BR, Moore BCJ. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 1990; 64:81–92. [PubMed: 1490904]
- Grimault N, Bacon SP, Micheyl C. Auditory stream segregation on the basis of amplitude-modulation rate. *J Acoust Soc Am.* 2002; 111:1340–1348. [PubMed: 11931311]
- Goldsworthy RL, Greenberg JE. Analysis of speech-based speech transmission index methods with implications for nonlinear operations. *J Acoust Soc Am.* 2004; 116:3679–3689. [PubMed: 15658718]
- Große JH, Mamo SK. Processing of temporal fine structure as a function of age. *Ear Hear.* 2010; 31:755–760. [PubMed: 20592614]
- Healy EW, Yoho SE, Wang Y, et al. An algorithm to improve speech recognition in noise for hearing-impaired listeners. *J Acoust Soc Am.* 2013; 134:3029–3038. [PubMed: 24116438]
- Hopkins K, Moore BCJ. The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J Acoust Soc Am.* 2011; 130:334–349. [PubMed: 21786903]
- Houtgast T, Steeneken HJM. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J Acoust Soc Am.* 1985; 77:1069–1077.
- Huber R, & Kollmeier B. POMO-Q: A new method for objective audio quality assessment using a model of auditory perception. *IEEE Trans Audio Speech Lang Proc.* 2006; 14:1902–1911.
- Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol.* 2007; 18:590–603. [PubMed: 18236646]

- International Telecommunication Union. ITU-R: BS. 1284–1, General Methods for the Subjective Assessment of Sound Quality. ITU: Geneva; 2003.
- Kates, JM. Digital Hearing Aids. San Diego, CA: Plural Publishing; 2008.
- Kates, JM. An auditory model for intelligibility and quality predictions. Proc. Mtgs. Acoust. 2013. (POMA) 19, 050184 (2013) ; <http://dx.doi.org/10.1121/1.4799223> Acoust. Soc. Am. 165th Meeting, Montreal, June 2–7, 2013
- Kates JM, Arehart KH. Hearing Aid Speech Quality Index. J Audio Engin Soc. 2010; 58:363–381.
- Kates JM, Arehart KH. The Hearing-Aid Speech Perception Index (HASPI). Speech Commun. 2014a; 65:75–93.
- Kates JM, Arehart KH. The Hearing-Aid Speech Quality Index (HASQI) Version 2. J Audio Engin Soc. 2014b; 62:99–117.
- Kjems U, Boldt JB, Pedersen MS, et al. Role of mask pattern in intelligibility of ideal binary-masked noisy speech. J Acoust Soc Am. 2009; 126:1415–1426. [PubMed: 19739755]
- Kreft, I.; de Leeuw, J. Introducing Multilevel Modeling. London, England: Sage; 1998.
- Kressner, AA.; Rozell, CJ. Int. Hear. Aid Conf. Vol. 16. Lake Tahoe, CA: 2014 Aug. The influence of structure in binary mask estimation error on speech intelligibility. 2014
- Laroche C, Hétu R, Quoc HT, et al. Frequency selectivity in workers with noise-induced hearing loss. Hear Res. 1992; 64(1):61–72. [PubMed: 1490902]
- Li N, Loizou PC. Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. J Acoust Soc Am. 2008; 123:1673–1682. [PubMed: 18345855]
- Littell, RC.; Milliken, GA.; Stroup, WW., et al. SAS System for Mixed Models. Second Edition. Cary, N.C: SAS Institute; 2006.
- Lunner T, Sundewall-Thorén E. Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. J Am Acad Audiol. 2007; 18:604–617. [PubMed: 18236647]
- Lunner T, Hietkamp RK, Andersen MR, Hopkins K, Moore BCJ. Effect of speech material on the benefit of temporal fine structure information in speech for normal-hearing and hearing-impaired participants. Ear Hear. 2012; 33(3):377–388. [PubMed: 22246137]
- Lutman ME, Gatehouse S, Worthington AG. Frequency resolution as a function of hearing threshold level and age. J Acoust Soc Am. 1991; 89:320–328. [PubMed: 2002171]
- Moore, BCJ. Cochlear Hearing Loss: Physiological, Psychological and Technical Issues Wiley-Interscience. 2nd Edition. England: John Wiley & Sons, Ltd; 2007.
- Moore BCJ. The choice of compression speed in hearing aids: theoretical and practical considerations and the role of individual differences. Trends Amplif. 2008; 12:103–112. [PubMed: 18567591]
- Neher T, Grimm G, Hohmann V, et al. Do hearing loss and cognitive function modulate benefit from different binaural noise-reduction settings? Ear Hear. 2014; 35:e52–e62. [PubMed: 24351610]
- Ng EH, Rudner M, Lunner T, et al. Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. Int J Audiol. 2013; 52:433–441. [PubMed: 23550584]
- Nordhoek IM, Drullman R. Effect of reducing temporal intensity modulations on sentence intelligibility. J Acoust Soc Am. 1997; 101:498–502. [PubMed: 9000740]
- Ohlenforst, B. Exploring the relationship between working memory, compressor speed and background noise characteristics. Masters thesis, Department of Acoustic Engineering, Centre for Applied Hearing Research, Technical University of Denmark; 2014.
- Patterson RD, Allerhand MH, Giguère C. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. J. Acoust. Soc. Am. 1995; 98:1890–1894. [PubMed: 7593913]
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd ed.. Newbury Park, CA: Sage; 2002.
- Rogosa D, Saner H. Longitudinal data analysis examples with random coefficient models. Journal of Educational and Behavioral Statistics. 1995; 20(2):149–170.
- Rönnerberg J, Arlinger S, Lyxell B, et al. Visual evoked potentials: Relation to adult speechreading and cognitive function. J Speech Hear Res. 1989; 32:725–735. [PubMed: 2601304]

- Rothauser E, et al. IEEE Subcommittee on Subjective Measurements IEEE Recommended Practices for Speech Quality Measurements. IEEE Transactions on Audio and Electroacoustics. 1969; 17:227–46.
- Rudner M, Lunner T. Cognitive spare capacity as a window on hearing aid benefit. *Seminars in Hearing*. 2013; 4:298–307.
- Sarampalis A, Kalluri S, Edwards B, et al. Objective measures of listening effort: Effects of background noise and noise reduction. *J Speech Lang Hear Res*. 2009; 52:1230–1240. [PubMed: 19380604]
- Singer J. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J Educ Behav Stat*. 1998; 24:323–355.
- Singer, JD.; Willett, JB. *Applied Longitudinal Data Analysis: Modeling change and event occurrence*. New York: Oxford University Press; 2003.
- Snijders, T.; Bosker, R. *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. 2000 ed.. Thousand Oaks CA: Sage Publications; 1999.
- Stone MA, Füllgrabe C, Mackinnon RC, Moore BCJ. The importance for speech intelligibility of random fluctuations in ‘steady’ background noise. *J Acoust Soc Am*. 2011; 30:2874–2881. [PubMed: 22087916]
- Taal CH, Hendriks RC, Heusdens R, Jensen J. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Trans Audio Speech and Lang Proc*. 2011; 19:2125–2136.
- Wang D, Kjems U, Pedersen MS, et al. Speech intelligibility in background noise with ideal binary time-frequency masking. *J Acoust Soc Am*. 2009; 125:2336–2347. [PubMed: 19354408]
- Wiley TL, Cruickshanks KJ, Nondahl DM, et al. Tympanometric measures in older adults. *J Am Acad Audiol*. 1996; 7:260–268. [PubMed: 8827920]
- Zahorian SA, Rothenberg M. Principal-components analysis for low-redundancy encoding of speech spectra. *J Acoust Soc Am*. 1981; 69:832–845.

FIGURE 1.

Audiograms are shown for the participants with hearing loss, with the average audiogram shown with a heavy line.

FIGURE 2.

Intelligibility scores (in RAU units, with SEs) are shown for seven levels of processing as a function of dB SNR. RAU, rationalized arcsine units; SNR, signal-to-noise ratio. The legend indicates the amount of noise attenuation in dB and the error rate in percent of cells.

FIGURE 3.

Average intelligibility scores (in RAU units) as a function of cepstral-based envelope distortion.

FIGURE 4.

Final model for intelligibility, showing four different fitted trajectories of intelligibility as a function of envelope fidelity for each HFPTA and RST group.

FIGURE 5.

Average quality ratings (with SEs) are shown for seven levels of processing as a function of dB SNR. The legend indicates the amount of noise attenuation in dB and the error rate in percent of cells.

FIGURE 6.
Average quality ratings as a function of cepstral-based envelope distortion.

Table 1

Test Conditions

Task	Attenuation (dB)	Error (%)	SNR (dB)
Intelligibility	10	0	-18, -12, -6, 0, 6, 12
	10	10	-18, -12, -6, 0, 6, 12
	10	30	-18, -12, -6, 0, 6, 12
	100	0	-18, -12, -6, 0, 6, 12
	100	10	-18, -12, -6, 0, 6, 12
	100	30	-18, -12, -6, 0, 6, 12
	No processing	No processing	-18, -12, -6, 0, 6, 12
Quality	10	0	-6, 0, 6, 12
	10	10	-6, 0, 6, 12
	10	30	-6, 0, 6, 12
	100	0	-6, 0, 6, 12
	100	10	-6, 0, 6, 12
	100	30	-6, 0, 6, 12
	No processing	No processing	-6, 0, 6, 12, No noise

Table 2

Results of repeated measures ANOVA for intelligibility scores, including the within- subject factors of SNR, attenuation (AT), and error (ER).

Factor	<i>df</i>	<i>F</i>	<i>p-value</i>	Partial η^2
<i>Within Subjects</i>				
Signal-to-Noise Ratio (SNR)	1.9, 58.4	1409.4	<0.001 *	0.979
Attenuation (AT)	1, 30	79.8	<0.001 *	0.727
Error (ER)	2, 60	381.4	<0.001 *	0.927
SNR x AT	3.4, 101.1	28.5	<0.001 *	0.487
SNR x ER	6.1, 184.4	47.7	<0.001 *	0.614
AT x ER	2, 60	14.2	<0.001 *	0.322
SNR x AT x ER	10, 300	6.1	<0.001 *	0.169

Significant effects ($p < 0.05$) are indicated with an asterisk (*).

Table 3

Estimate of dB SNR corresponding to average 50% intelligibility and corresponding dB benefit (relative to the unprocessed condition) shown for six ideal binary mask processing conditions.

Condition		dB SNR for 50% intelligibility	Effective dB improvement
Error (%)	Atten (dB)		
0	100	-6.3	8.4
10	100	-5.7	7.8
0	10	-4.2	6.3
10	10	-3.2	5.3
30	100	-0.7	2.8
30	10	-0.2	2.3
Unp	Unp	2.1	--

Table 4

HLM Model Summary Table Intelligibility

	Intercept		Slope	
	Variable	Explained Variance	Variable	Explained Variance
Model A	HFPTA **	64%	HFPTA **	51%
Model B	HFPTA ** RST *	69%	HFPTA ** RST	46%
Model C	HFPTA ** RST Age	68%	HFPTA ** RST Age	43%

Note. This table reports the main effects and variance explained for three hierarchical models. Intercept and slope were entered as outcome variables, and the independent variables included one or more of the following: HFPTA group, RST group, and Age. Intercept was defined as the predicted RAU Intelligibility score when fidelity was equal to 0.5 and slope was defined as the change in RAU intelligibility for one unit change in fidelity.

*
 $p < 0.05$

**
 $p < 0.01$

Table 5

Intelligibility coefficient estimations for the final two-level model.

<i>Fixed Effects</i>	<i>Coefficient</i>	<i>se</i>	<i>T Ratio</i>	<i>p-value</i>
Model for Intelligibility Intercept				
Intercept (B_{00})	70.65	3.30	21.39	<0.001*
HFPTA Group (B_{01})	-18.34	2.37	-7.74	<0.001*
RST Group (B_{02})	2.68	1.25	2.15	0.04*
Model for Intelligibility slopes				
Intercept (B_{10})	322.62	9.07	35.55	<0.001*
HFPTA Group (B_{11})	-33.16	7.54	-4.40	<0.001*
RST Group (B_{12})	1.37	3.93	0.35	0.729
<i>Random Effects</i>	<i>Variance Component</i>	<i>df</i>	<i>Chi-sq</i>	<i>p-value</i>
Level-2 Residual Variance Intercept (r_0)	40.23	28	152.97	<0.001*
Level-2 Residual Variance Slope (r_1)	279.83	28	51.05	0.005*
Level-1 Residual Variance (e_1)	375.02			

Significant effects ($p < 0.05$) are indicated with an asterisk (*).

Table 6

Results of repeated measures ANOVA for quality ratings, including the within-subject factors of SNR, attenuation (AT), and error (ER).

Factor	<i>df</i>	<i>F</i>	<i>p-value</i>	Partial η^2
<i>Within Subjects</i>				
Signal-to-Noise Ratio (SNR)	1.4, 39.3	353.8	<0.001*	0.924
Attenuation (AT)	1, 29	19.5	<0.001*	0.401
Error (ER)	1.2, 34.6	251.2	<0.001*	0.896
SNR x AT	3, 87	1.7	0.19	0.054
SNR x ER	3.6, 103.1	2.2	0.08	0.070
AT x ER	1.6, 44.4	66	<0.001*	0.695
SNR x AT x ER	3.8, 108.9	5.2	0.001*	0.151

Significant effects ($p < 0.05$) are indicated with an asterisk (*).

Table 7

Estimate of dB SNR corresponding to average quality rating of 0.5 and corresponding dB benefit (relative to the unprocessed condition) shown for six ideal binary mask processing conditions.

Condition		dB SNR for 0.5 quality rating	Effective dB improvement
Error (%)	Atten (dB)		
0	100	<-6.0	>14.9
0	10	0.8	8.1
10	100	0.8	8.1
10	10	3.6	5.3
30	100	9.1	-0.2
30	10	7.6	1.3
Unp	Unp	8.9	--

Table 8

HLM Model Summary Table Quality

	Intercept		Slope	
	Variable	Explained Variance	Variable	Explained Variance
Model A	HFPTA	0%	HFPTA	0%
Model B	HFPTA RST	0%	HFPTA RST	0%
Model C	HFPTA RST Age	0%	HFPTA RST Age	0%

Note. This table reports the main effects and variance explained for three hierarchical models. Intercept and slope were entered as outcome variables, and the independent variables included one or more of the following: HFPTA group, RST group, and Age. Intercept was defined as the predicted RAU Intelligibility score when fidelity was equal to 0.5 and slope was defined as the change in RAU intelligibility for one unit change in fidelity.

*
 $p < 0.05$

**
 $p < 0.01$

Table 9

Quality coefficient estimations for the final two-level model.

<i>Fixed Effects</i>	<i>Coefficient</i>	<i>se</i>	<i>T Ratio</i>	<i>p-value</i>
Model for Quality Intercept				
Intercept (B_{00})	0.46	0.05	8.51	<0.001 *
HFPTA Group (B_{01})	0.02	0.04	0.68	0.503
RST Group (B_{02})	-0.01	0.02	-0.80	0.432
Model for Quality slopes				
Intercept (B_{10})	1.27	0.13	9.49	<0.001 *
HFPTA Group (B_{11})	0.05	0.09	0.59	0.581
RST Group (B_{12})	-0.01	0.05	-0.12	0.907
<i>Random Effects</i>	<i>Variance Component</i>	<i>Df</i>	<i>Chi-sq</i>	<i>p-value</i>
Level-2 Residual Variance Intercept (r_0)	0.011	27	880.48	<0.001 *
Level-2 Residual Variance Slope (r_1)	0.056	27	131.86	<0.001 *
Level-1 Residual Variance (e_1)	0.009			

Significant effects ($p < 0.05$) are indicated with an asterisk (*).