

Investigating Differences in Preferred Noise Reduction Strength Among Hearing Aid Users

Trends in Hearing
2016, Vol. 20: 1–14
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516655794
tia.sagepub.com



Tobias Neher^{1,2} and Kirsten C. Wagener^{2,3}

Abstract

Even though hearing aid (HA) users can respond very differently to noise reduction (NR) processing, knowledge about possible drivers of this variability (and thus ways of addressing it in HA fittings) is sparse. The current study investigated differences in preferred NR strength among HA users. Participants were groups of experienced users with clear preferences (“NR lovers”; $N = 14$) or dislikes (“NR haters”; $N = 13$) for strong NR processing, as determined in two earlier studies. Maximally acceptable background noise levels, detection thresholds for speech distortions caused by NR processing, and self-reported “sound personality” traits were considered as candidate measures for explaining group membership. Participants also adjusted the strength of the (binaural coherence-based) NR algorithm to their preferred level. Consistent with previous findings, NR lovers favored stronger processing than NR haters, although there also was some overlap. While maximally acceptable noise levels and detection thresholds for speech distortions tended to be higher for NR lovers than for NR haters, group differences were only marginally significant. No clear group differences were observed in the self-report data. Taken together, these results indicate that preferred NR strength is an individual trait that is fairly stable across time and that is not easily captured by psychoacoustic, audiological, or self-report measures aimed at indexing susceptibility to background noise and processing artifacts. To achieve more personalized NR processing, an effective approach may be to let HA users determine the optimal setting themselves during the fitting process.

Keywords

hearing loss, hearing aids, noise reduction, individual differences, personalized treatment

Date received: 12 December 2015; revised: 12 March 2016; accepted: 14 March 2016

Introduction

Digital hearing aids (HAs) are typically equipped with a range of signal processing algorithms including directional processing, noise reduction (NR), and amplitude compression (e.g., Dillon, 2012). A number of studies have indicated that individual HA users can respond very differently to these types of algorithms (e.g., Gatehouse, Naylor, & Elberling, 2006; Houben, Dijkstra, & Dreschler, 2012a; Keidser, Dillon, Convery, & Mejia, 2013; Lunner, 2003). As a consequence, it is of interest to understand these differences better, so that possible avenues for more personalized algorithm settings can be identified. Although considerable progress has been made with respect to individualizing amplitude compression systems, the same is not true for other types of HA algorithms.

The current study focused on individual differences in NR outcome. Generally speaking, NR processing does not improve speech intelligibility in noise, but the

attenuation of noisy signal components can lead to improved listening comfort, albeit at the cost of added processing artifacts (e.g., Bentler, Wu, Kettel, & Hurtig, 2008; Loizou & Kim, 2011). In other words, NR processing involves a trade-off between desirable noise attenuation and undesirable speech distortions (e.g., Kates, 2008), and there are indications that HA users respond differently to these conflicting effects (Houben et al., 2012a; Marzinzik, 2000). In a number of recent studies, we have investigated the influence of individual factors on experienced HA users’ preference for, and speech recognition with, different NR settings (Neher, 2014; Neher,

¹Medizinische Physik, Oldenburg University, Oldenburg, Germany

²Cluster of Excellence Hearing4all, Oldenburg, Germany

³Hörzentrum Oldenburg GmbH, Oldenburg, Germany

Corresponding author:

Tobias Neher, Department of Medical Physics and Acoustics, Carl-von-Ossietzky University, D-26111 Oldenburg, Germany.
Email: tobias.neher@uni-oldenburg.de



Grimm, Hohmann, & Kollmeier, 2014; Neher, Wagener, & Fischer, 2016). Our data analyses revealed considerable inter-individual variability in preferred NR setting. Furthermore, they indicated that preferred NR strength varies with input signal-to-noise ratio (SNR). That is, our participants generally favored stronger NR processing at 4 dB SNR than at 0 and -4 dB SNR. Regarding individual influences, we saw indications that participants with higher pure-tone average hearing thresholds (PTAs) and poorer cognitive performance, as assessed using a reading span test (Neher et al., 2014) or a measure of “executive control” (Neher, 2014; Neher et al., 2016), prefer stronger NR than participants with lower PTAs and better performance on those measures (see also Participants section). This could indicate that the former types of participants are more affected by noise and less by speech distortions, whereas for the latter types of participants the opposite may be true.

While these results provide some indications in terms of how NR processing may be personalized, the observed relations with hearing loss and cognitive factors only accounted for some of the variability in NR preference. Because strong NR can impair speech intelligibility (e.g., Loizou & Kim, 2011; Neher, 2014), it is important to be able to identify candidates for strong NR reliably. Thus, the main objective of the current study was to investigate alternative means of predicting NR preference. We investigated if preference for strong (or weak) NR processing is associated with increased (or decreased) susceptibility to background noise and decreased (or increased) sensitivity to speech distortions. To that end, we retested some of the participants from our earlier studies on a number of measures designed to tap into aspects related to noise acceptance and distortion sensitivity. More specifically, we included two psychoacoustic or audiological measures as well as a novel “sound personality” questionnaire covering domains such as noise sensitivity or importance of sound quality as potential candidates for predicting NR preference. A secondary aim was to confirm the differences in preferred processing strength across listeners and input SNRs found previously. In this way, we wanted to examine the consistency of these judgments over time. To that end, we had our participants adjust the NR to their preferred level at two input SNRs (i.e., 0 and 4 dB). On the basis of the insights gained in this manner, we aimed to lay the basis for a clinically feasible way of personalizing NR processing in HAs.

Previous research into individual differences in preferred NR strength is scarce, especially as far as HA users are concerned. Houben, Dijkstra, and Dreschler (2012b) conducted a study with 10 normal-hearing participants and observed a large spread in preferred NR settings. In another study, Houben et al. (2012a) used a method of self-adjustment to investigate preferred NR strength with 10 normal-hearing and 7 hearing-impaired

listeners. Again, they found considerable spread, which was of comparable magnitude in both groups. Using 12 normal-hearing and 12 hearing-impaired participants, Brons, Dreschler, and Houben (2014) extended these results by additionally assessing their participants’ sensitivity to distortions of the signal mixture, the target speech, and the background noise caused by NR processing. On average, the hearing-impaired listeners tended to have higher detection thresholds for the different types of signal distortions than the normal-hearing listeners, and their inter-individual threshold differences were also larger.

The study of Brons et al. (2014) constitutes a first step toward elucidating differences in NR outcome among listeners with normal and impaired hearing based on psychoacoustic measurements. So far, however, no corresponding steps seem to have been taken to elucidate such differences among HA users. Not only does this apply to how HA users respond to signal distortions but also to how they respond to noise (which NR schemes are designed to attenuate). In the field of audiology, the Acceptable Noise Level (ANL) measure of Nabelek, Tucker, and Letowski (1991) has frequently been used to investigate the relation between response to noise and NR outcome (e.g., Fredelake, Holube, Schlueter, & Hansen, 2012; Mueller, Weber, & Hornsby, 2006; Peeters, Kuk, Lau, & Keenan, 2009; Wu & Stangl, 2013). Up until now, however, its ability to account for NR preference does not seem to have been examined. Furthermore, although some researchers have attempted to employ self-report measures for that purpose, these endeavors have hitherto been unsuccessful (Recker, McKinney, & Edwards, 2011).

The current study sought to address these shortcomings. Its aims were to investigate (a) the long-term consistency and SNR dependence of NR preference and (b) the ability of a number of psychoacoustic, audiological, and self-report measures aimed at indexing noise acceptance, distortion sensitivity, and other sound personality traits to explain (or predict) NR preference. Regarding the first aim, we hypothesized that for the participants tested here (i.e., experienced HA users), NR preference would generally be stable across time. Furthermore, we expected to find that with increasing input SNR stronger NR processing would be preferred. Regarding the second aim, we anticipated that participants with a preference for stronger NR processing would be more susceptible to background noise and less sensitive to speech distortions, whereas for participants with a preference for weaker NR processing the opposite would be true.

Materials and Methods

Ethical approval for all experimental procedures was obtained from the ethics committee of the University

of Oldenburg (reference number DRS.21/20/2013). Prior to any data collection, written informed consent was obtained from all participants. Participants were paid on an hourly basis for their participation.

Participants

The participants were recruited from a cohort of 60 habitual HA users who had all taken part in our two previous studies (Neher, 2014; Neher et al., 2016). These studies had taken place about 1 year prior to the measurements reported here. At that point in time, each participant had had at least 9 months of HA experience. For the current study, we initially reanalyzed the preference judgments from these studies, which we had obtained with the (binaural coherence-based) NR algorithm tested here (Neher, 2014) as well as a different (single-microphone, modulation-based) NR algorithm implemented in wearable HAs (Neher et al., 2016). Our motivation for considering the data from both studies (and hence two different algorithms) was to obtain indices of our participants' general liking of NR processing. Both sets of preference judgments were based on a large number of pairwise comparisons of inactive, moderate, and strong NR. More specifically, the judgments were proportional values (with a range of 0 to 1) reflecting how much a given NR setting was preferred to the other ones. For the current study, we calculated an aggregate preference score per participant and NR setting by averaging the two sets of preference judgments obtained at 0 and 4 dB SNR. On the basis of the resultant scores, we then identified those 2×15 HA users with the clearest dislikes ("NR haters") or preferences ("NR lovers") for strong NR processing. Because 3 of these 30 participants were unavailable at the time of testing, the current study was carried out with 27 participants (13 NR haters, 14 NR lovers). For 23 of them (11 NR haters, 12 NR lovers), preferred NR strength was unambiguous in the sense that the scores for inactive NR were much higher than the ones for strong NR or vice versa (mean scores 11 NR haters: 0.70, 0.54, and 0.26 for inactive, moderate, and strong NR, respectively; mean scores 12 NR lovers: 0.19, 0.55, and 0.76 for inactive, moderate, and strong NR, respectively). For the two remaining NR lovers, the scores for moderate and strong NR were equally high (mean scores: 0.22, 0.64, and 0.64 for inactive, moderate, and strong NR, respectively), while for the two remaining NR haters the scores for moderate NR were somewhat higher than the ones for inactive NR (mean scores: 0.50, 0.74, and 0.26 for inactive, moderate, and strong NR, respectively). Thus, except for a couple of "borderline cases" per group that tended to converge at moderate NR (i.e., especially the two NR haters), the two groups were well separated in terms of preferred NR strength.

The 27 participants of the current study were aged 61 to 81 years. They all had symmetrical sensorineural hearing impairment defined as (a) asymmetries in air-conduction thresholds of no more than 15 dB HL across ears for the standard audiometric frequencies from 0.5 to 4 kHz and (b) air-bone gaps no larger than 15 dB HL at any audiometric frequency between 0.5 and 4 kHz. Furthermore, all of them had previously passed a number of sensory and neuropsychological screening tests (cf., Neher, 2014). Three independent t tests (all $|t|_{25} < 1.4$, all $p > .17$) revealed that the two groups of participants did not differ in terms of age (mean ages: 73 vs. 70 years), PTAs across 500 Hz to 4 kHz and both ears (mean PTAs: 44 vs. 47 dB HL), or performance on the aforementioned reading span test (Carroll et al., 2015; mean scores: 39 vs. 40% correctly recalled target words). Another independent t test ($t_{25} = 2.1$, $p = .048$) revealed that the NR haters had higher scores on the aforementioned measure of executive control than the NR lovers (Zimmermann & Fimm, 2012; mean scores: 93 vs. 81% correctly responded to target stimuli). This difference in executive control performance is consistent with our previous findings concerning individual influences on NR outcome (see Introduction section). Based on these, however, one would also expect a group difference in PTAs. While there was a trend for the NR lovers to have higher PTAs than the NR haters (see above), this difference was not statistically significant. Presumably, this was related to a loss of statistical power due to the much smaller cohort tested this time ($N = 27$ in the current study vs. $N = 60$ in the previous studies).

Physical Test Setup

All testing was carried out under headphones in a sound-proof booth. Inside the booth, a touch screen displayed the graphical user interfaces (GUIs) used during the measurements (see below). All measurement software was implemented in Matlab (MathWorks, Natick, USA). It was run on a personal computer (PC) located outside the booth that was equipped with an RME (Haimhausen, Germany) DIGI96/8 soundcard. The soundcard was connected to a Tucker-Davis Technologies (Alachua, USA) HB7 headphone buffer and a pair of Sennheiser (Wennebostel, Germany) HDA200 headphones used for stimulus presentation. Calibration was carried out using a Brüel & Kjær (B&K; Nærum, Denmark) 4153 artificial ear, a B&K 4134 1/2" microphone, a B&K 2669 preamplifier, and a B&K 2610 measurement amplifier.

The measurement PC was connected to another PC also located outside the booth and equipped with an RME Digiface soundcard via a local area network and an optical digital audio interface. On this additional PC, a simulation of a bilateral HA fitting implemented on the

Master Hearing Aid research platform (Grimm, Herzke, Berg, & Hohmann, 2006) was run, which could be controlled from the measurement PC. The additional PC received the stimuli from the measurement PC via the optical digital audio interface, processed them in real-time, and then routed them back to the measurement PC via the optical digital audio interface.

Speech Stimuli

The stimuli used for the current study closely resembled those we had used previously. They were based on recordings from the Oldenburg sentence test (Wagener, Brand, & Kollmeier, 1999). To simulate a realistically complex listening situation, we convolved these recordings with publicly available pairs of head-related impulse responses measured in a reverberant cafeteria using a head-and-torso simulator equipped with two behind-the-ear HA dummies (Kayser et al., 2009). Each HA dummy consisted of the microphone array housed in its original casing, but without any of the integrated amplifiers, speakers, or signal processors commonly used in HAs. For the current study, we used the measurements made with the (omnidirectional) front microphones of each HA dummy and a source at an azimuth of 0° and a distance of 1 m from, and at the same height as, the head-and-torso simulator. For the interfering signal, we used a publicly available recording made in the same cafeteria with the same setup during a busy lunch hour (Kayser et al., 2009). This recording, which is several minutes in length, is characterized by continuous unintelligible speech babble, occasional parts of intelligible speech from nearby speakers, as well as sporadic transient sounds from cutlery, dishes, and chairs. During the measurements, we presented this recording at a nominal sound pressure level (SPL) of 65 dB and mixed it with the target sentences, the level of which we adjusted to produce a given SNR.

HA Processing

The HA processing also closely resembled what we had used previously (cf., Neher, 2014). It included binaural coherence-based NR (Grimm, Hohmann, & Kollmeier, 2009), individual linear amplification according to the “National Acoustic Laboratories-Revised Profound” prescription rule (Dillon, 2012), and a 32-tap finite impulse response filter that compensated for the uneven frequency response of the headphones. All processing was carried out at a sampling rate of 44.1 kHz.

The NR algorithm tested here relies on estimates of the binaural coherence (or interaural similarity) for distinguishing between desired and undesired acoustic information. As such, it requires the exchange of information across the left and right devices in a bilateral

fitting. An implicit assumption made in the design of this algorithm is that incoherent signal components constitute detrimental acoustic information for the user (because they typically are due to strong reflections or diffuse background noise) and thus can be attenuated. First, the binaural coherence of the ear input signals is estimated as a function of time and frequency. The estimates produced in this manner can take on values between 0 and 1. A value of 0 corresponds to fully incoherent (or diffuse) sound, while a value of 1 corresponds to fully coherent (or directional) sound. Because of diffraction effects around the head, the coherence is always high at low frequencies. At frequencies above about 1 kHz, the coherence is low for diffuse and reverberant signal components, but high for the direct sound from nearby directional sources (e.g., talkers). Due to the spectro-temporal fluctuations contained in speech, the ratio between incoherent and coherent signal components may vary across time and frequency. By applying appropriate time- and frequency-dependent gains to the noisy (binaural) input signal, this ratio can be improved. These gains are obtained by applying an exponent, α , to the coherence estimates and then mapping the resultant values to the intended gain range.

In the current study, we used a gain range of -30 to 0 dB and a 40-ms integration time constant for estimating the binaural coherence. To vary the strength of the applied NR processing, we varied the parameter α . Setting α to 0, 0.75, or 2 resulted in the inactive, moderate, or strong NR settings we had tested previously (Neher, 2014). Figure 1 illustrates the effect of varying α on the mapping function between the binaural coherence estimates and NR gains. As can be seen, larger α -values lead to greater attenuation of signal

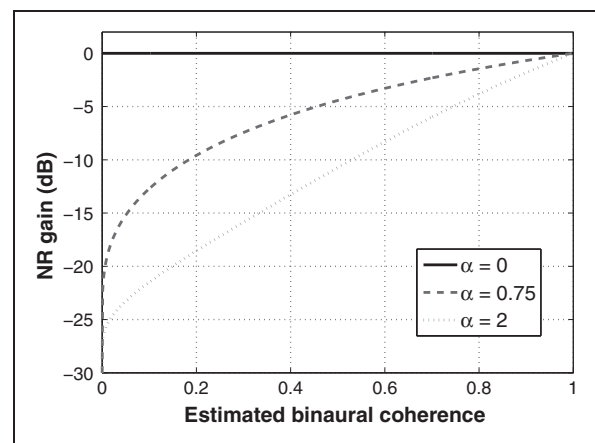


Figure 1. NR gain as a function of the estimated binaural coherence for three values of α (i.e., the parameter determining the NR strength) corresponding to the inactive ($\alpha = 0$), moderate ($\alpha = 0.75$), and strong ($\alpha = 2$) NR settings.

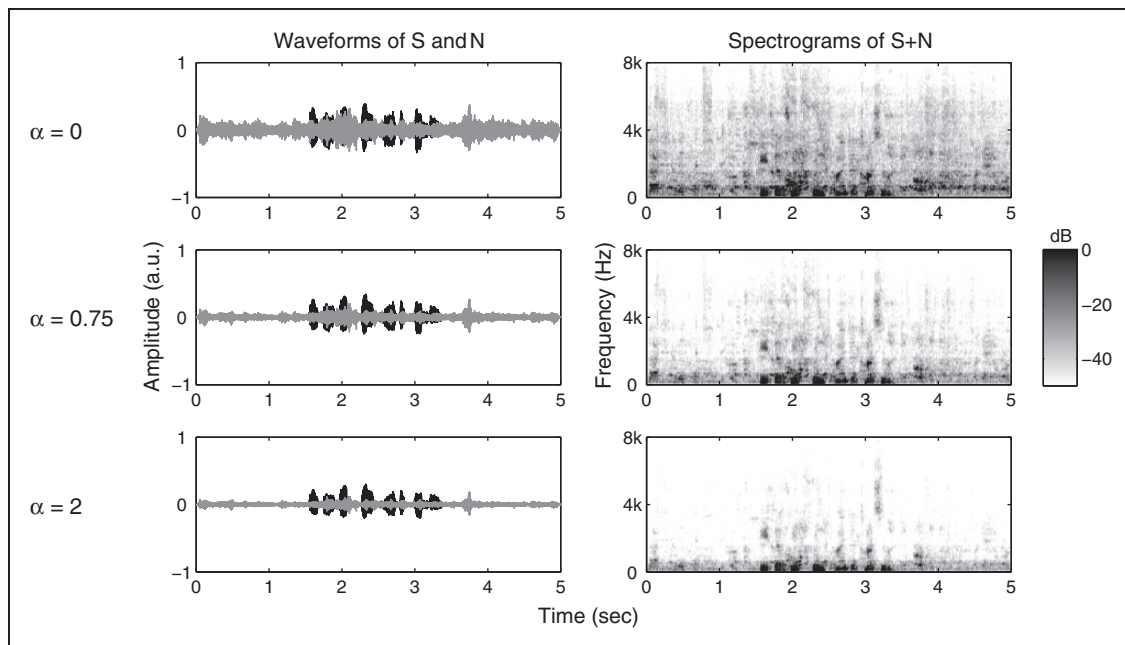


Figure 2. Graphical illustration of the effects of inactive ($\alpha = 0$), moderate ($\alpha = 0.75$), and strong ($\alpha = 2$) binaural coherence-based NR processing on (one channel of) an example stimulus with an input SNR of +4 dB. Panels on the left-hand side show time waveforms of the target speech, S (black) and the cafeteria noise, N (grey). Panels on the right-hand side show corresponding spectrograms for the signal mixtures, S + N. a.u. denotes arbitrary units.

components with a given level of binaural coherence. Figure 2 illustrates the physical effects of the inactive, moderate, or strong NR settings for an example stimulus with an input SNR of 4 dB. The panels on the left-hand side show, for each NR setting, the waveforms of the speech and noise signals at the HA output. The panels on the right-hand side show the spectrograms of the corresponding signal mixtures. As can be seen, the dominant effect of moderate and especially strong NR is to suppress incoherent signal components above about 1 kHz. The speech-weighted SNR improvements due to moderate and strong NR amounted to 1.7 and 2.8 dB for an input SNR of 0 dB, and to 2.3 and 3.8 dB for an input SNR of 4 dB (cf., Table 2 in Neher, 2014). Thus, greater NR strength led to an increase in output SNR, especially at higher input SNRs. However, greater NR strength also resulted in greater distortion of the target speech, especially at lower input SNRs (cf., Table 3 in Neher, 2014). As is typical of NR processing, the amount of noise attenuation achieved, therefore, covaried with the amount of speech distortion introduced concurrently.

Measurements

The measurements described below were distributed across two visits with a maximum duration of 1.5 h each. At the beginning of the study, the sound personality questionnaire was sent out to the participants who completed it in their own time. Upon returning the

questionnaire, they went through their responses with an experimenter to resolve any open issues.

Self-adjusted NR strength. To confirm the basic group difference (and in this way assess long-term consistency) with respect to NR preference, we asked our participants to imagine being inside the cafeteria and wanting to communicate with the target talker. They then had to adjust the strength of the NR algorithm such that they would be willing to listen to the result for a prolonged time. Participants could make these adjustments in real-time using a large slider arranged vertically on a GUI displayed on the touch screen. The slider, which allowed for the adjustments to be made with a step size of less than 0.01, was labeled “Less noise suppression” at the bottom and “More noise suppression” at the top; no other labels or markers were used. Positioning the slider at the bottom resulted in inactive ($\alpha = 0$) NR; positioning it at the top resulted in very strong ($\alpha = 4$) NR. To force the participants to adjust the slider anew on each run, we randomized the initial slider position (and hence α -value) across runs. Furthermore, we applied a non-linear mapping between the slider scale and the underlying α -values (e.g., small α -increments at the bottom end and large α -increments at the top end of the scale for a given slider displacement and vice versa), which we also varied across runs. In this way, we forced our participants to change the slider position across a range of α -values on each run in order to find their preferred setting.

At the beginning of a given run, 20 randomly chosen sentences from the Oldenburg sentence test were concatenated with 1.5 s of silence between consecutive sentences. The resultant signal was then mixed with a randomly chosen extract from the cafeteria recording, and the speech-in-noise mixture was played back in a loop until the measurement was completed. The measurements were carried out at two input SNRs: 0 and 4 dB. Participants initially completed two training runs (one per input SNR), followed by six test runs (three per input SNR) in randomized order.

Acceptable noise level. To assess noise acceptance, we made use of the ANL measure. In the original ANL procedure, participants initially have to adjust the level of the target speech to their most comfortable level, which is kept fixed during all subsequent measurements. Background noise is then added, and participants are asked to adjust its level three times in a row: (a) so they no longer can follow the target speech, (b) so they can follow the target speech very easily, and (c) so they are just about able to tolerate the noise while trying to follow the target speech for a prolonged time (the “maximal ANL”). The difference between the most comfortable speech level and the maximal ANL is then taken as the ANL estimate, with lower values indicating greater noise acceptance. Essentially, the ANL can, therefore, be interpreted as the lowest SNR that a listener is willing to accept for prolonged listening.

In the current study, we presented the target speech at a fixed, nominal level of 65 dB SPL, that is, our participants only adjusted the level of the cafeteria noise. For that purpose, they used a GUI which included six horizontally arranged buttons: three for attenuating the noise and three for amplifying it. From left to right, these buttons were labeled “---,” “--,” “-,” “+,” “++,” and “+++.” Pressing the buttons resulted in changes to the background noise level of ± 6 , ± 3 , and ± 1 dB for the outermost, intermediate, and innermost buttons, respectively. Participants could change the noise level as long as they needed to reach a decision. They then had to confirm their adjustment by pressing an “OK” button located at the bottom of the GUI, after which the next run was automatically started.

The stimuli for the ANL measurements were identical to those used for measuring self-adjusted NR strengths (see above), except that the SNR was determined by the noise level adjustments made by the participants. The noise level adjustments occurred at the input of our simulated pair of HAs. The HAs were programmed to provide inactive ($\alpha=0$), moderate ($\alpha=0.75$), or strong ($\alpha=2$) NR. The measurements made with inactive NR served as estimates of general noise acceptance (“baseline ANL”). The measurements made with moderate and strong NR served to verify the expected benefit from active NR with respect to (greater) noise acceptance.

Initially, we carried out six training runs (two per NR setting) followed by nine test runs (three per NR setting) in randomized order. Despite additional training, one participant was unable to carry out the ANL measurements according to the instructions and was thus excluded from the analyses. For a given test run, we obtained the ANL estimate by taking the difference between the nominal speech level (i.e., 65 dB SPL) and the maximal ANL from that run.

Detectability of speech distortions. To assess detectability of distortions caused by NR processing, we followed the approach of Brons et al. (2014). That is, we measured detection thresholds for speech distortions using an adaptive three-interval two-alternative forced-choice paradigm. On each trial, the task of the participant was to choose which of two sound samples (“A” or “B”) was different from a reference sound sample (“Ref”). The reference sound sample, which was always presented in the first interval, was an unprocessed sentence without noise from the Oldenburg sentence test. The target sound sample was the same sentence without noise processed with the NR gains computed for the signal mixture at +4 dB SNR. On each trial, the target sound sample was randomly allocated to interval A or B. During stimulus presentation, each interval was visually highlighted on a GUI that consisted of three large buttons arranged left to right and labeled Ref, A, and B. Following stimulus presentation, participants responded by pressing on A or B, after which the correct interval was visually highlighted for feedback purposes.

Each measurement started with a very large NR strength ($\alpha=4$). Following a correct (or incorrect) response, α was halved (or doubled) until the first lower reversal occurred (one-up one-down procedure). Subsequently, it was divided (or multiplied) by 1.5 until the second reversal occurred, and then by 1.25 until the minimum step size of 0.125 was reached. Following three lower reversals, the measurement phase started and the adaptive procedure changed to a one-up three-down procedure that allowed us to estimate the 79.4% detection threshold (Levitt, 1971). A measurement was completed once five additional lower reversals had occurred. Two such measurements were carried out per participant.

The reference sound sample was presented at a nominal level of 69 dB SPL and thus an input SNR of +4 dB, broadly consistent with the +5 dB(A) used by Brons et al. (2014). In general, one would expect the input SNR to affect absolute detection thresholds, with higher SNRs leading to higher thresholds. This is because, for a given NR strength, speech distortions will decrease with input SNR (see HA processing section). In contrast, the input SNR is unlikely to affect inter-individual threshold differences, which the current

study focused on. The target sound sample was equated with the reference sound sample in terms of its root-mean-square level. To prevent the participants from relying on any potentially remaining loudness differences, we applied level roving of 0, ± 1 , or ± 2 dB during intervals A and B and also instructed them to concentrate on differences other than loudness to complete the task. For both the target and reference sound samples, we randomized the five possible roving levels and applied them in a blockwise manner (i.e., to five consecutive trials). We then repeated these steps until the end of the measurement sequence.

The measurements started with one training run that included three lower reversals with the one-up one-down procedure followed by one lower reversal with the one-up three-down procedure. Afterwards, the two test runs were carried out. As our threshold estimates, we used the median of the last eight upper and lower reversals per measurement and participant. If, for a given measurement, the standard deviation of these eight reversals exceeded two times the minimum step size of the corresponding threshold value, we discarded that estimate (and thereby rejected threshold estimates with large tracking excursions). As a consequence, we excluded six (out of 54) threshold estimates, that is, one threshold each of two NR haters and four NR lovers.

Self-reported sound personality. To assess self-reported characteristics related to sound personality traits, such as noise sensitivity and importance of sound quality, we used a recently developed questionnaire intended to predict preference for, and thus usage of, different types of HA technology (Meis, Huber, Fischer, Schulte, & Meister, 2015). In its original form, this questionnaire consists of 46 items that were derived based on expert interviews as well as focus groups and in-depth interviews with both normal-hearing and hearing-impaired listeners. In analyzing the data from 622 predominantly older participants with different degrees of hearing loss who had been given the questionnaire to investigate its basic properties, Meis et al. (2015) uncovered an underlying structure with seven factors: (F1) annoyance/distraction by background noise, (F2) importance of sound quality, (F3) noise sensitivity, (F4) avoidance of unpredictable sounds, (F5) openness towards loud/new sounds, (F6) preference for warm sounds, and (F7) detail in environmental sounds/music. Appendix A provides an overview of the 7 factors and 23 questionnaire items loading onto them.

As part of the current study, we explored the predictive power of these factors with respect to NR preference. Given our focus on factors related to response to noise and processing artifacts, we were particularly interested in the predictive power of F1, F2, and F3. Furthermore, given the low-pass filter-like effects of the NR algorithm

tested here (see HA processing section), we were also interested in the predictive power of F6.

Speech intelligibility. As mentioned earlier, previous research has shown that NR processing can lead to speech intelligibility impairments. In our earlier study (Neher, 2014), we had, therefore, assessed speech intelligibility with the inactive, moderate, and strong NR settings also tested here. More specifically, we had carried out measurements at SNRs of -4 and 0 dB using stimuli essentially identical to the ones described above (see Speech stimuli section). For each measurement, we had used one test list from the Oldenburg sentence test consisting of 20 five-word sentences each (Wagener et al., 1999). As a supplement to the outcomes considered in the current study, we reanalyzed the data of the 27 participants tested here. That is, for each participant and NR setting, we calculated the corresponding speech recognition rate (in percent correct).

Results

Self-Adjusted NR Strength

To assess the consistency of the participants' NR adjustments across the three test runs per input SNR, we calculated six pairwise Pearson's correlation coefficients, which were all high (all $r > 0.71$, all $p < .0001$). Since six corresponding paired t tests showed no changes in mean self-adjusted α -values across test runs (all $t_{26} < 0.9$, all $p > .4$), we used the median of the three self-adjusted α -values per input SNR and participant for all subsequent analyses.

At 0 dB SNR, self-adjusted α -values ranged from 0.1 to 2.2 among the NR haters and from 0.6 to 2.2 among the NR lovers; at 4 dB SNR, these ranges were virtually unchanged (NR haters: 0.1 to 2.3 ; NR lovers: 0.6 to 2.3). Thus, the two groups overlapped somewhat in terms of self-adjusted NR strengths. To check if individual differences in self-adjusted NR strength were correlated across the two input SNRs, we calculated Pearson's correlation coefficient for the two sets of α -values, which we found to be high ($r = 0.74$, $p < .0001$).

Figure 3 shows mean self-adjusted α -values and corresponding 95% confidence intervals for the two groups of participants and input SNRs (for illustrative purposes, the α -values corresponding to the inactive, moderate, and strong NR settings are also indicated). Consistent with our expectations, the NR haters set the algorithm to provide weaker NR processing than the NR lovers (grand average α -values: 0.8 and 1.4 , respectively). Also consistent with our expectations, both groups set the algorithm to provide stronger NR processing at 4 than at 0 dB SNR (grand average α -values: 1.3 and 1.0 , respectively). To check the statistical significance of

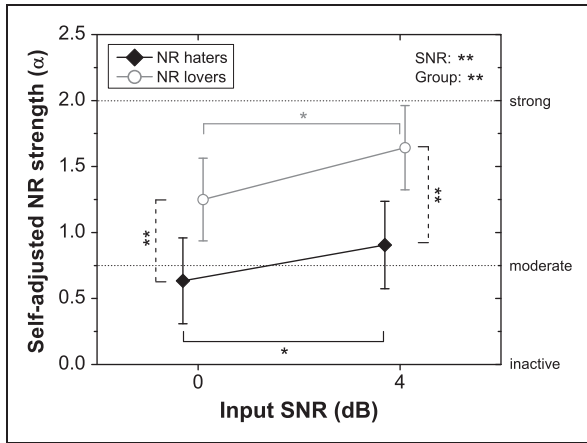


Figure 3. Mean self-adjusted NR strengths and corresponding 95% confidence intervals for the two groups of participants and input SNRs. α -values corresponding to the inactive, moderate, and strong NR settings are also indicated. * $p < .05$. ** $p < .01$.

these observations, we performed a repeated-measures analysis of variance (ANOVA) with SNR as within-subject factor and participant group as between-subject factor. This revealed strongly significant effects of SNR ($F(1, 25) = 12.5$, $p < .01$, $\eta_p^2 = 0.33$) and participant group ($F(1, 25) = 11.4$, $p < .01$, $\eta_p^2 = 0.31$), but no interaction between these factors ($p > .5$).

Acceptable Noise Level

To assess the consistency of the ANL estimates across the three test runs per NR setting, we calculated nine pairwise Pearson's correlation coefficients, which were all rather high (all $r > 0.66$, all $p < .001$). Since nine corresponding paired t tests showed no changes in mean ANLs across test runs (all $|t|_{25} < 1.3$, all $p > .2$), we used the median of the three ANL estimates per NR setting and participant for all subsequent analyses.

Baseline ANLs ranged from -5 to 13 dB among the NR haters and from -6 to 15 dB among the NR lovers. With moderate (or strong) NR, the corresponding ranges were -5 to 12 dB (or -5 to 11 dB) and -3 to 10 dB (or -3 to 8 dB), respectively. Thus, the two groups also overlapped in terms of their ANLs. To check if individual differences in ANL were correlated across the three NR settings, we calculated Pearson's correlation coefficients for the three sets of scores, which were all high (all $r > 0.75$, all $p < .00001$).

Figure 4 shows mean ANLs and corresponding 95% confidence intervals for the two groups of participants and three NR settings. Consistent with our expectations, the NR lovers tended to have higher baseline ANLs than the NR haters (mean ANLs: 7.0 and 4.8 dB, respectively). Also consistent with our expectations, active NR processing resulted in lower ANLs than inactive NR

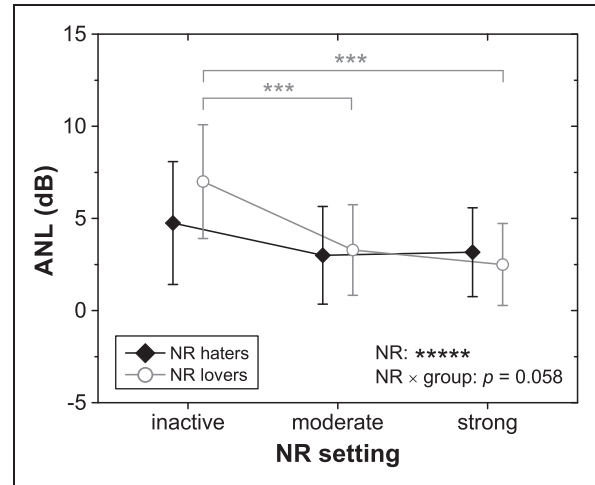


Figure 4. Mean ANLs and corresponding 95% confidence intervals for the two groups of participants and three NR settings. *** $p < .001$. **** $p < .00001$.

processing (mean ANLs: 6.0 , 3.2 , and 2.8 dB for inactive, moderate, and strong NR, respectively). To check the statistical significance of these observations, we performed a repeated-measures ANOVA with NR setting as within-subject factor and participant group as between-subject factor. This revealed a highly significant effect of NR setting ($F(2, 48) = 15.3$, $p < .00001$, $\eta_p^2 = 0.39$), a non-significant effect of participant group ($p > .7$), and an interaction between NR setting and participant group that just failed to reach significance ($F(2, 48) = 3.0$, $p = .058$, $\eta_p^2 = 0.11$). To further examine the effect of NR setting, we performed a post hoc analysis that revealed significant differences between inactive NR and both moderate and strong NR (both $p < .0001$), but not between moderate and strong NR ($p = .6$). Closer inspection of the (marginally significant but potentially interesting) interaction with listener group showed that for the NR lovers ANLs decreased by 3.7 and 4.5 dB with moderate and strong NR, respectively (both $p < .001$). In contrast, no improvements in ANL due to active NR were observable for the NR haters (both $p > .075$).

Detectability of Speech Distortions

To assess the consistency of the detection thresholds for speech distortions, we calculated Pearson's correlation coefficient for the data from the 21 participants with two reliable threshold estimates (see Measurements section). This revealed a reasonably strong test-retest correlation ($r = 0.67$, $p = .001$). Given that a paired t test revealed no difference in mean thresholds between the two sets of measurements ($t_{20} = 1.7$, $p = .1$), we used the arithmetic mean of the two threshold estimates of these participants for all subsequent analyses. For the other six

participants, we used the single remaining threshold estimate. Because the threshold estimate of one participant (i.e., NR lover) was disproportionately high (α -value at threshold = 0.85; test and retest thresholds: 0.94 and 0.76, respectively), we excluded that data point to normalize the variance in our dataset.

The α -value detection thresholds of the remaining (2×13) participants ranged from 0.21 to 0.56 among the NR haters and from 0.25 to 0.75 among the NR lovers (data not shown). Thus, detection thresholds for speech distortions also overlapped somewhat across the two groups. Although the NR lovers had on average somewhat higher detection thresholds for speech distortions than the NR haters (mean α -values at threshold: 0.46 and 0.36, respectively), this difference failed to reach statistical significance in a one-way ANOVA with participant group as between-subject factor ($F(1, 23) = 3.9$, $p = .060$, $\eta_p^2 = 0.15$).

Self-Reported Sound Personality

For the analysis of the sound personality data, we calculated, for each participant, the mean score across all questionnaire items belonging to a given factor (cf., Appendix A). Figure 5 shows boxplots of the scores for the seven factors separated by participant group. As can be seen, with the exception of F1 (“annoyance/distraction by background noise”) and F7 (“detail in environmental sounds/music”), the spread in the scores was large for both groups. Furthermore, the data of the two groups showed considerable overlap. Performing a series of two-tailed Mann-Whitney U -tests on these data revealed no significant group differences (all $p > .05$).

Speech Intelligibility

Grand average speech recognition rates at -4 and 0 dB SNR were 37% and 76%-correct, respectively. Grand average speech recognition rates with inactive, moderate, and strong NR were 60%, 57%, and 52%-correct, respectively. Performing a repeated-measures ANOVA on the rationalized arcsine unit-transformed (Studebaker, 1985) speech scores with SNR and NR setting as within-subject factors and participant group as between-subject factor confirmed highly significant effects of SNR ($F(1, 25) = 300.7$, $p < .00001$, $\eta_p^2 = 0.92$) and NR setting ($F(2, 50) = 16.3$, $p < .00001$, $\eta_p^2 = 0.39$). The effect of participant group was non-significant, as were all the interactions (all $p > .1$). A post hoc analysis revealed significant differences between strong NR and both inactive and moderate NR (both $p < .001$), but not between inactive and moderate NR ($p = .058$). Taken together, these results imply that for SNRs above 0 dB speech intelligibility was generally high and that for α -values larger than 0.75 (corresponding to the moderate NR setting) speech intelligibility impairments likely occurred.

Correlations Among Measures

To assess the long-term consistency of NR preference, we correlated the self-adjusted α -values averaged across 0 and 4 dB SNR with the aggregate preference scores for inactive and strong NR that we had derived based on the pairwise preference judgments collected previously at 0 and 4 dB SNR (see Participants section). In support of the hypothesis that preferred NR strength is a stable

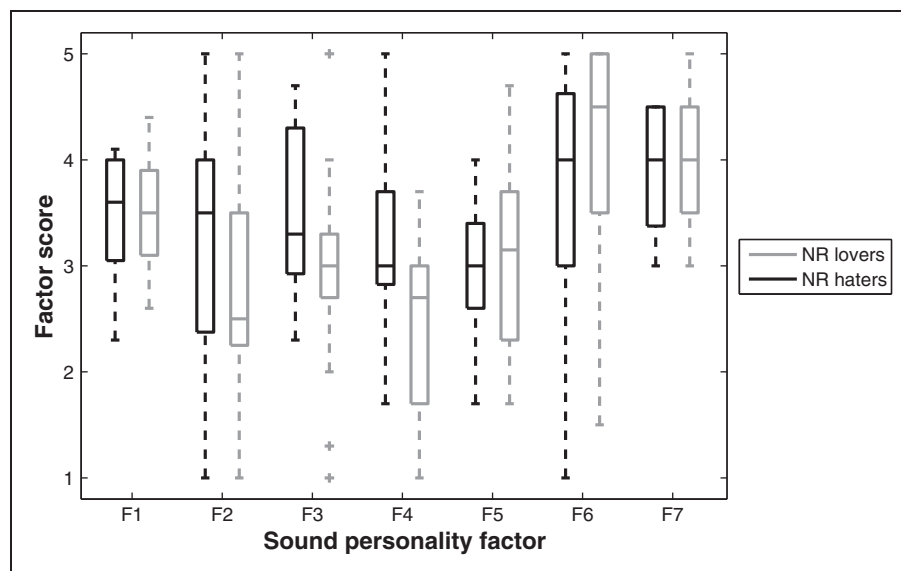


Figure 5. Boxplots of the scores for the seven factors from the sound personality questionnaire for the two groups of participants.

trait, we observed relatively strong correlations (preference scores for inactive NR: $r = -0.64$, $p < .001$; preference scores for strong NR: $r = 0.62$, $p < .001$). Figure 6 shows a scatter plot of aggregate preference scores for strong NR against mean self-adjusted α -values. As can be seen, the self-adjusted α -values of the NR lovers exceeded the moderate NR setting ($\alpha = 0.75$), consistent with a general liking of strong NR. Concerning the NR haters, there were seven participants whose self-adjusted α -values fell clearly below the moderate NR setting, consistent with a general dislike for strong NR. However, there were also six participants (including the two “borderline cases”; see Participants section) whose self-adjusted α -values clearly exceeded the moderate NR setting and thus fell within the range of the NR lovers.

To find out if individual differences in response to noise and processing artifacts can account for NR outcome, we correlated the self-adjusted α -values at 0 and 4 dB SNR with the baseline ANLs, detection thresholds for speech distortions, and the F1, F2, F3, and F6 questionnaire scores. Consistent with the lack of clear across-group differences in terms of the latter measures or factors (see above), we found no significant correlations (all $|r| < 0.27$, all $p > .15$). (The same was true for the speech scores, for which we observed no correlations either.)

Finally, because working memory capacity has recently received considerable attention as a potential predictor of HA outcome (cf., Souza, Arehart, & Neher, 2015), we also explored potential correlations between reading span performance and self-adjusted

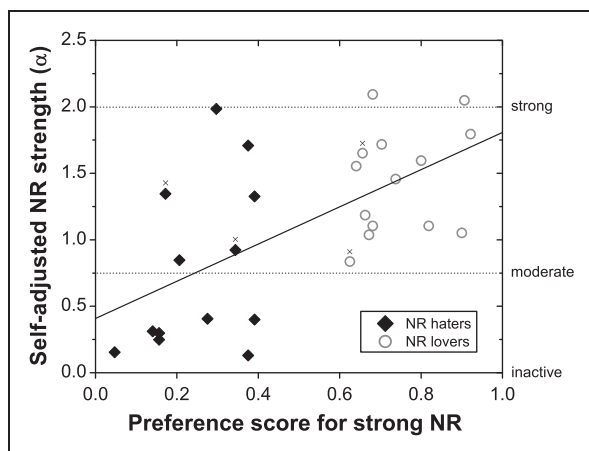


Figure 6. Scatter plot of aggregate preference scores for strong NR derived from the data from the two previous studies against self-adjusted NR strengths averaged across 0 and 4 dB SNR from the current study. The black solid line shows the least-squares linear fit. Data points marked by the \times symbols correspond to the four participants with “borderline” preference scores (see Participants section for details). α -values corresponding to the inactive, moderate, and strong NR settings are also indicated.

α -values, baseline ANLs, detection thresholds for speech distortions, and F1, F2, F3, and F6 questionnaire scores. When adjusting for multiple comparisons, none of the correlations was significant (which could be due to a lack of statistical power).

Discussion

The aims of the current study were (a) to assess the long-term consistency as well as the SNR dependence of NR preference and (b) to investigate if a number of psychoacoustic, audiological, and self-report measures of distortion sensitivity, noise acceptance, and sound personality traits are able to explain (or predict) group membership.

Concerning the first aim, the NR lovers set the strength of the algorithm tested here to almost twice the value chosen by the NR haters (Figure 3), thereby confirming the group difference observed previously. Furthermore, the self-adjusted NR strengths reported here were clearly correlated with the preference scores from our previous studies ($|r| > 0.6$). Given that we had collected the previous set of data about 1 year earlier, this finding indicates that, for experienced HA users at least, NR preference is generally stable across time. Nevertheless, there were also a few NR haters whose self-adjusted NR settings fell well within the range of the NR lovers (Figure 6). In other words, some participants who previously had favored fairly weak NR processing favored a much stronger setting this time, thereby effectively changing groups. It is also worth recalling that inter-individual differences in preferred NR strength were generally large. This variability, which is in agreement with other literature data (see Introduction section), suggests that when fitting HAs, it could be helpful to be able to adjust the NR strength over a wide range of levels in order to find the individually optimal setting.

Also consistent with our earlier results, we found that at 4 dB SNR, our participants preferred stronger NR processing than at 0 dB SNR. This finding can be traced back to the fact that at higher input SNRs the adverse effects of NR processing (i.e., speech distortions) decrease while its positive effects (i.e., noise attenuation) increase, as also confirmed by some technical measurements (see HA processing section). Thus, with increasing input SNR, the positive effects of NR processing will increasingly outweigh any unwanted side effects. In principle, HA users can, therefore, be expected to experience benefit from NR processing at positive SNRs where speech intelligibility is at ceiling and where at least some HA manufacturers have chosen to restrict the efficacy of their NR algorithms (cf., Smeds, Bergman, Hertzman, & Nyman, 2010).

Furthermore, it is worth noting that the self-adjusted α -values generally clearly exceeded the detection

thresholds for speech distortions (e.g., by a factor of 3.4 at 4 dB SNR across all participants). As a matter of fact, there were only four participants (i.e., NR haters) who set the NR strength to be near their individual detection thresholds. This is broadly consistent with Brons et al. (2014) who found that their participants preferred NR settings that were stronger than the individual detection thresholds for distortions in noisy speech. Taken together, these results indicate that when listening to speech in noise for a prolonged time the preferred NR setting generally exceeds those settings for which speech distortions become detectable and for which speech intelligibility impairments start to occur.

Concerning the second research aim, neither baseline ANLs nor detection thresholds for speech distortions proved to be effective at explaining preference for either weak or strong NR. That is, although there was a tendency for the NR lovers to exhibit higher (poorer) detection thresholds for speech distortions than the NR haters, this difference was only marginally significant. Essentially, the same was true for the baseline ANLs, that is, there was a tendency for the NR lovers to exhibit higher baseline ANLs—and thus less tolerance toward noise—than the NR haters (Figure 4). As a consequence, neither of these measures correlated with self-adjusted NR strength (nor with speech intelligibility). A possible reason for this could be that the two measures were not sufficiently reliable, thus leading to the non-significant findings. In this context, however, it should be recalled that we had generally observed strong test–retest correlations. Furthermore, in the aforementioned study of Brons et al. (2014), no significant correlation between preferred NR strength and detectability of signal distortions was observed either. Together with our results, this seems to suggest that sensitivity to processing artifacts on its own is not an effective predictor of NR preference.

Regarding the ANL measure, we had made an effort to control intra-individual variability by carrying out two training and three test runs per condition to obtain reliable ANL estimates. These choices were motivated by two recent studies that recommended that at least one ANL measurement be carried out prior to actual testing (Brännström, Olsen, Holm, Kastberg, & Ibertsson, 2014) and that proposed that noise acceptance changes gradually over a *range* of levels (Brännström, Holm, Kastberg, & Olsen, 2014), thereby requiring multiple measurements to obtain reliable ANL data. In spite of our efforts, however, we could not confirm the putative difference in terms of noise acceptance among our NR haters and lovers. This suggests that, as applied here, the ANL measure is not an effective predictor of NR preference either.

Under the assumption that preferred NR strength is governed by a trade-off between noise attenuation and signal distortion, a more promising approach for future

work might be to develop a measure that somehow combines susceptibility to background noise and processing artifacts. In this way, participants would be able to weigh these conflicting factors according to the importance that they have for them. In a recent study, Brons, Houben, and Dreschler (2013) asked a group of young normal-hearing listeners to perform a comprehensive assessment of various NR-processed stimuli in terms of overall preference, noise annoyance, and speech naturalness. Using linear regression analysis, these authors then estimated how much the latter two factors contributed to overall preference. Interestingly, although some participants weighted noise annoyance and speech naturalness equally, for others one of these factors was clearly more dominant. Future research would have to examine if this approach is equally effective for revealing such differences among HA users and if it can be turned into a clinically feasible (and thus more time-efficient) method for fitting NR schemes. An alternative approach for future work could be to explore the predictive power of quantitative distortion metrics such as the Hearing Aid Sound Quality Index (Kates & Arehart, 2014) or PEMO-Q (Huber & Kollmeier, 2006), which were designed to respond to both additive noise and signal distortions.

Interestingly, whereas for the NR lovers, we observed statistically significant ANL improvements due to active NR processing, the same was not true for the NR haters (Figure 4). This could imply that for listeners who inherently are rather insensitive to background noise active NR processing will offer less or even no benefit compared with HA users who experience strong noise discomfort. An alternative explanation for this across-group difference could be that, because of the SNR dependency of the NR processing that we applied (see HA processing section), participants with higher baseline ANLs (and thus input SNRs) benefited more from active NR processing because of the larger SNR improvements and weaker speech distortions that they experienced. To resolve this issue, future research would have to separate the physical from the perceptual effects, for example by exposing groups of listeners with low and high baseline ANLs to the same changes in SNR and speech distortion.

Regarding the sound personality questionnaire, this measure did not reveal any differences among our two groups (Figure 5). Broadly speaking, this is consistent with other research failing to link self-reported characteristics to an individual's response to noisy speech. For instance, Freyaldenhoven, Nabelek, and Tampas (2008) found no relation between self-reported hearing problems as measured using the Abbreviated Profile of Hearing Aid Benefit questionnaire and ANLs for 191 hearing-impaired listeners. Consistent with this, Recker et al. (2011) found that their results from a custom

noise-tolerance questionnaire failed to accurately predict the ANLs of 86 individuals with normal hearing and 53 individuals with impaired hearing. In the current study, we explored the ability of the sound personality questionnaire to explain differences in HA outcome for the first time using a small set of processing conditions. Because this inventory was designed with a relatively broad range of personal listening habits in mind, it would be worthwhile to apply it to a wider range of HA conditions with a broader range of acoustical and perceptual effects to test it more fully.

Finally, it is worth noting that we found the reading span measure to be uncorrelated with self-adjusted NR strength (and several other measures). While this is in line with our two earlier studies as part of which the reading span data also used here were collected (Neher, 2014; Neher et al., 2016), it is inconsistent with another study of ours in which we had found an association between shorter reading span and preference for strong NR (Neher et al., 2014). In general, the predictive power of measures of reading span has received a great deal of attention in HA research lately. However, the obtained results are not always easily reconcilable across studies, particularly so with respect to NR outcome. This has led some researchers to advocate the use of more well-controlled and comparable test conditions in future work (Souza et al., 2015).

In conclusion, the results from the current study suggest that preferred NR strength is an individual trait that, at least for experienced HA users, is generally stable across time and that is not easily captured by psychoacoustic, audiological, or self-report measures intended to index susceptibility to background noise and processing artifacts. From a practical point of view, if the aim is to address differences in response to NR processing among HA users, an effective way of doing so could be to include some form of self-adjustment in the fitting process.

Appendix A: Sound Personality Questionnaire

The sound personality questionnaire of Meis et al. (2015) consists of 46 items covering different listening habits. For scoring purposes, a 5-point rating scale ranging from *disagree completely* to *agree completely* is used. Below, the seven underlying factors (F1 to F7) together with the individual items loading onto them are summarized. (In the actual questionnaire, the items appear in randomized order.)

F1: Annoyance/distraction by background noise

1. I find background sounds very disturbing when on the phone

2. I find background sounds very disturbing during conversations in restaurants
3. Mobile phones should suppress disturbing sounds, even at the cost of poorer sound quality
4. I get readily distracted by background sounds during conversations
5. Conversations are no fun when the radio is switched on
6. I could do without music or background sounds in movies in order to better understand the actors
7. Background sounds like a dishwasher are very disturbing for me during conversations
8. I switch off the radio when in noisy vehicles

F2: Importance of sound quality

1. I would be willing to invest a lot of money in a good sound system
2. I would be able to hear differences between expensive and cheap loudspeakers

F3: Noise sensitivity

1. I am more sensitive to loud sounds than most other persons
2. I am more sensitive to sharp/shrill sounds than other persons
3. I am generally sensitive to sounds

F4: Avoidance of unpredictable sounds

1. I avoid uncertain listening situations
2. I don't like unpredictable sounds such as a suddenly approaching car
3. I prefer acoustically familiar situations to unfamiliar ones

F5: Openness towards loud/new sounds

1. I like "bombastic" sounds when watching movies at home
2. I am open to new sound impressions and settings in audio devices
3. I enjoy music even when it is loud

F6: Preference for warm sounds

1. I prefer a warm sound quality when on the phone
2. I prefer a warm sound quality when listening to music

F7: Details of environmental sounds/music

1. I would like to hear all sounds when going for a walk in the woods
2. I would like to perceive all details when listening to music

Acknowledgments

The authors thank Giso Grimm (HörTech gGmbH) for support with the test setup and Markus Meis (Hörzentrum Oldenburg GmbH) for input regarding the sound personality questionnaire. Parts of this research were presented at the 5th International Symposium on Auditory and Audiological Research (ISAAR), Nyborg, Denmark, August 26–28.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the DFG Cluster of Excellence EXC 1077/1 “Hearing4all” and by Siemens Audiologische Technik (now Sivantos) GmbH, Erlangen, Germany.

References

- 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.
- Brännström, K. J., Holm, L., Kastberg, T., & Olsen, S. O. (2014). The acceptable noise level: The effect of repeated measurements. *International Journal of Audiology, 53*, 21–29.
- Brännström, K. J., Olsen, S. O., Holm, L., Kastberg, T., & Ibertsson, T. (2014). The effect of repeated measurements and working memory on the most comfortable level in the ANL test. *International Journal of Audiology, 53*, 787–795.
- Brons, I., Dreschler, W. A., & Houben, R. (2014). Detection threshold for sound distortion resulting from noise reduction in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 136*, 1375–1384.
- Brons, I., Houben, R., & Dreschler, W. A. (2013). Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort. *Ear & Hearing, 34*, 29–41.
- Carroll, R., Meis, M., Schulte, M., Vormann, M., Kiessling, J., & Meister, H. (2015). Development of a German reading span test with dual task design for application in cognitive hearing research. *International Journal of Audiology, 54*, 136–141.
- Dillon, H. (2012). *Hearing aids* (2nd ed.). Sydney, Australia: Boomerang Press.
- Fredelake, S., Holube, I., Schlueter, A., & Hansen, M. (2012). Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms. *International Journal of Audiology, 51*, 299–308.
- Freyaldenhoven, M. C., Nabelek, A. K., & Tampas, J. W. (2008). Relationship between acceptable noise level and the abbreviated profile of hearing aid benefit. *Journal of Speech, Language, and Hearing Research, 51*, 136–146.
- Gatehouse, S., Naylor, G., & Elberling, C. (2006). Linear and nonlinear hearing aid fittings – 2. Patterns of candidature. *International Journal of Audiology, 45*, 153–171.
- Grimm, G., Herzke, T., Berg, D., & Hohmann, V. (2006). The master hearing aid: A PC-based platform for algorithm development and evaluation. *Acta Acustica United With Acustica, 92*, 618–628.
- Grimm, G., Hohmann, V., & Kollmeier, B. (2009). Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme. *IEEE Transactions on Audio, Speech, and Language Processing, 17*, 1408–1419.
- Houben, R., Dijkstra, T. M. H., & Dreschler, W. A. (2012a). The influence of noise type on the preferred setting of a noise reduction algorithm. In T. Dau, M. L. Jepsen, T. Poulsen, & J. C. Dalsgaard (Eds.), *Speech perception and auditory disorders*. Nyborg, Denmark: The Danavox Jubilee Foundation.
- Houben, R., Dijkstra, T. M. H., & Dreschler, W. A. (2012b). Analysis of individual preferences for tuning noise-reduction algorithms. *Journal of the Audio Engineering Society, 60*, 1024–1037.
- Huber, R., & Kollmeier, B. (2006). PEMO-Q – A new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech, and Language Processing, 14*, 1902–1911.
- Kates, J. M. (2008). *Digital hearing aids* (2nd ed.). San Diego, CA: Plural Publishing Inc.
- Kates, J. M., & Arehart, K. H. (2014). The hearing-aid speech quality index (HASQI) version 2. *Journal of the Audio Engineering Society, 62*, 99–117.
- Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of multi-channel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal of Advanced Signal Processing, 298605* (1), 1–10.
- Keidser, G., Dillon, H., Convery, E., & Mejia, J. (2013). Factors influencing individual variation in perceptual directional microphone benefit. *Journal of the American Academy of Audiology, 24*, 955–968.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467–477.
- Loizou, P. C., & Kim, G. (2011). Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Transactions on Audio, Speech, and Language Processing, 19*, 47–56.
- Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology, 42* (Suppl. 1): S49–S58.
- Marzinzik, M. (2000). *Noise reduction schemes for digital hearing aids and their use for the hearing impaired* (PhD thesis). Medical Physics Section, Oldenburg University, Shaker Verlag.
- Meis, M., Huber, R., Fischer, R. -L., Schulte, M., & Meister, H. (2015). Development of a questionnaire for the assessment of sound preferences and hearing habits of people with different degrees of hearing impairment. *12th Congress of the European Federation of Audiology Societies (EFAS), Istanbul, Turkey, May 27–30*.
- Mueller, H. G., Weber, J., & Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends in Amplification, 10*, 83–93.

- Nabelek, A. K., Tucker, F. M., & Letowski, T. R. (1991). Toleration of background noises: Relationship with patterns of hearing aid use by elderly persons. *Journal of Speech and Hearing Research, 34*, 679–685.
- 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. doi:10.3389/fnins.2014.00391.
- Neher, T., Grimm, G., Hohmann, V., & Kollmeier, B. (2014). Do hearing loss and cognitive function modulate benefit from different binaural noise-reduction settings? *Ear & Hearing, 35*, e52–e62.
- Neher, T., Wagener, K. C., & Fischer, R.-L. (2016). Directional processing and noise reduction in hearing aids: Individual and situational influences on preferred setting. *Journal of the American Academy of Audiology, 27* (8). doi: 10.3766/jaaa.15062.
- Peeters, H., Kuk, F., Lau, C. C., & Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. *Journal of the American Academy of Audiology, 20*, 89–98.
- Recker, K., McKinney, M. F., & Edwards, B. (2011). Can acceptable noise levels be predicted from a noise-tolerance questionnaire? *Canadian Hearing Report, 6*, 31–38.
- Smeds, K., Bergman, N., Hertzman, S., & Nyman, T. (2010). Noise reduction in modern hearing aids – Long-term average gain measurements using speech. In J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, & T. Poulsen (Eds.), *Binaural processing and spatial hearing* (pp. 445–452). Elsinore, Denmark: The Danavox Jubilee Foundation.
- Souza, P., Arehart, K., & Neher, T. (2015). Working memory and hearing aid processing: Literature findings, future directions, and clinical applications. *Frontiers in Psychology, 6*, 1894. doi:10.3389/fpsyg.2015.01894.
- Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research, 28*, 455–462.
- Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I-III: Design, optimierung und evaluation des Oldenburger satztests [Development and evaluation of a sentence test for the German language. I-III: Design, optimization and evaluation of the Oldenburg sentence test]. *Zeitschrift für Audiologie (Audiological Acoustics), 38*, 44486–155695.
- Wu, Y. H., & Stangl, E. (2013). The effect of hearing aid signal-processing schemes on acceptable noise levels: Perception and prediction. *Ear & Hearing, 34*, 333–341.
- Zimmermann, P., & Fimm, B. (2012). *Testbatterie zur aufmerksamkeitsprüfung-version mobilität* (Version 1.3 edn.) [Test battery for the assessment of attentional skills-Mobility version]. Herzogenrath, Germany: Psytest.