Benefits of a low processing delay for first-time hearing aid users in real life

Nadja Schinkel-Bielefeld[1], Frederik Vejby Nielsen[1], Laura Winther Balling[2], Federica Bianchi[2], Hendrik Husstedt[3], Florian Denk[3], Filip Rønne[1]

1) ORCA Labs, 2) WSA, 3) German Institute of Hearing Aids

Abstract

Open-fit hearing aids combine amplified and direct sound at the eardrum. The inevitable processing delay of hearing aids can therefore produce comb-filtering artifacts that may affect sound quality, own-voice perception, and overall hearing aid satisfaction. This field study investigated whether reducing processing delay to 0.5 ms provides benefits for first-time hearing aid users.

In a three-week field trial, participants wore two pairs of hearing aids with 0.5 ms (low delay, LD) and with 8 ms delay (high delay, HD). During an initial familiarization week, participants directly compared the two pairs of devices in selected listening situations. This was followed by a randomized crossover trial in which each pair was worn for one week. Daily ratings of satisfaction, sound quality, and own voice perception were collected. At the end of each week, participants completed the International Outcome Inventory for Hearing Aids (IOI-HA), additional satisfaction questionnaires including assessments of own-voice perception, and the Hearing-Related Lifestyle Questionnaire (HEARLI-Q). Finally, participants indicated their overall device preference.

Results showed higher total IOI-HA scores for LD, with higher scores for both introspection and interaction. HEARLI-Q responses indicated greater hearing aid satisfaction and reduced perceived difficulty for LD processing. While there was no difference in objective speech intelligibility, subjective ratings of sound quality, hearing aid satisfaction, naturalness of sound, own voice perception, and listening effort consistently favored LD processing. These findings suggest that processing delay is an important determinant of real-world hearing-aid experience and hence may support first-time hearing-aid adoption.

Keywords: hearing aids, processing delay, comb filtering, open fit, first-time hearing aid users

Introduction

Despite the high prevalence and many possible negative consequences of hearing loss (e.g. Borre et al., 2023; Dixon et al., 2020; Li et al., 2014; Livingston et al., 2024), hearing aid adoption rates and hearing aid use are unexpectedly low (e.g. Snyder et al., 2025). The reasons for low adoption and non-use are numerous, but one contributing factor is low sound quality (McCormack & Fortnum, 2013). New hearing aid wearers often have mild to moderate sensorineural hearing loss and little hearing loss in the low frequencies. Hence, they are typically treated with hearing aids with open fittings. Advantages of open fittings include a reduction of the occlusion effect and, accordingly an improved sound quality of one's own voice (Denk et al., 2023; Winkler et al., 2016). However, using an open fitting may also introduce sound artifacts as the amplified sound is delayed and therefore out of phase with the direct sound entering through the vent.

Hearing aids inevitably have some processing delay. For modern hearing aids, delays are typically around 5-10 ms (Denk et al., 2024), but can be as low as 0.5 ms (Balling et al., 2020). For open fittings, longer delays can cause a comb-filter effect when the processed sound interferes with the direct sound that enters through the vent in the ear tip or leakage around it, resulting in a change of timbre. These distortions are influenced by the magnitude of the processing delay, the hearing aid gain, and the degree of openness of the provided fitting. Openness and hearing aid gain also interact, as the balanced-armature drivers used in hearing aids are unable to provide high sound pressure levels at low frequencies with open fittings.

In a study with simulated linear vented hearing aids, Denk et al. (2021) showed that for normal hearing listeners, processing delays as low as 0.1 ms can be detectable. In another simulation study, Stiefenhofer (2022) found detection thresholds of 0.3 ms for normal hearing listeners and 0.6 to 1 ms for hearing impaired listeners, depending on gain and hearing loss. While detectable, Stone et al. (2008) found 5 to 6 ms delay acceptable. Goehring et al. (2018) found that delay tolerance was higher for participants with hearing impairment, particularly experienced hearing aid users, than for normal hearing listeners. Bramsløw (2010) found no difference in audio quality for processing delays between 5 and 10 ms. Using simulated hearing aids with delays ranging from 0.5 to 10 ms, Lelic, Stiefenhofer et al. (2022) showed that the shortest delay was preferred by listeners with normal hearing and listeners with mild hearing loss in frequencies below 2 kHz. Súsonnudóttir et al. (2024) found no difference in preference between 0 ms delay and 0.5 ms, but a significant difference between these two pooled delays and a set of longer delays (2 ms, 5 ms, and 10 ms).

The timbre change due to comb-filtering may be particularly perceptible on one's own voice, which is critical for hearing aid satisfaction (Kochkin, 2010) and hearing aid use (Kochkin, 2000). Súsonnudóttir (2026) showed that a 0.5 ms processing delay led to a more favorable own-voice perception than 10 ms of delay. Similarly, Groth & Birkmose (2004) found that individuals with either normal hearing or hearing loss rated their own voice as more disturbing with 10 ms delay than with 2 ms or 4 ms delay. Agnew and Thornton (2000) found that, when listening to their own voice, a processing delay of 3 to 5 ms was noticeable, and a delay of more than 10 ms was objectionable.

The effect of processing delay cannot only be seen in behavioral measures, but also in the neural representation of the temporal envelope. Zhou et al. (2024) demonstrated that a processing delay of 5 or 7 ms disrupts neural phase locking and hence may impact speech-in-noise performance. Also, Küçük & Çetinkaya (2026) show that an ultra-low delay of 0.5 ms can improve central auditory temporal and spectral processing abilities compared to delay decreasing with frequency from approx. 5 to 2 ms.

While the aforementioned studies show the benefit of low delay in the laboratory, there is also one real-life study on low delay (Súsonnudóttir Hansen, 2025). This pilot study includes 14 participants with different levels of hearing aid experience and also found benefits of a 0.5 ms delay compared to 10 ms delay. This is suggestive of a real-world benefit, but the evidence is limited, particularly among first-time hearing-aid users. Because first-time users are at risk of hearing-aid rejection and may be especially sensitive to first impressions, understanding the role of processing delay in this population is of particular clinical relevance.

Here, we investigate in a three-week blinded cross-over field trial whether a low processing delay of 0.5 ms increases hearing aid satisfaction in first-time hearing aid users, compared to a typical 8 ms processing delay. Hereafter, these two conditions will be referred to as low delay (LD) and high delay (HD), respectively. Primary outcome measures were hearing-aid benefit as assessed by the International Outcome Inventory for Hearing Aids (IOI-HA) (Cox et al., 2002) and hearing aid satisfaction as assessed by the modified Hearing-Related Lifestyle Questionnaire (HEARLI-Q) (Lelic, Wolters, et al., 2022). These are standardized questionnaires and were also used in the aforementioned pilot study (Súsonnudóttir Hansen, 2025). Daily questionnaires, questionnaires on own voice, and final ratings of preference, hearing aid satisfaction, sound quality, and own voice perception were added as secondary outcome measures to understand the effect of low delay on various aspects of hearing aid benefit in more detail.

Our research hypotheses were:

H1: Participants are overall more satisfied with LD hearing aids than with HD hearing aids.

H2: The hearing aid outcome as measured using IOI-HA is more positive with LD devices. This includes the participants’ relation to their hearing aids as well as their interaction with the rest of the world.

H3: Participants are more satisfied with the sound of LD devices and perceive it as more natural.

H4: Participants are more satisfied with their own voice using LD hearing aids.

H5: Participants report less listening effort and less hearing difficulty with LD devices.

H6: Participants overall prefer LD devices over HD devices, particularly if they report audio quality as being important to them.

H7: Advantages of LD are not only apparent in direct comparison, but also after a week of real-world use.

Methods

Participants

54 German-speaking first-time hearing aid (HA) users aged 47 to 82 years (mean age 67.5 years, SD: 7.2) participated in the study. A power analysis based on IOI-HA and HA satisfaction data from HEARLI-Q (Súsonnudóttir Hansen, 2025) indicated that a power of 0.8 would be achieved with 34 and 36 participants, respectively. Inclusion criteria were a) first-time HA wearers (never owned HAs and never used them outside of laboratory experiments or trial periods more than 2 years before the study), b) self-reported active interest in obtaining HAs, and c) hearing loss in the better ear of at least 30 dB in at least one frequency between 0.5 and 4 kHz which corresponds to the legal indication for HAs in Germany. An exclusion criterion was a hearing loss exceeding 40, 45, 55, 70, and 80 dB in the frequencies <0.5, 1, 2, 4, and 8 kHz, respectively. Four participants dropped out before the end of the study. Two more participants did not indicate a final preference, as they decided HAs were not beneficial for them. This resulted in 52 sets of responses for analysis for week 1, 50 for weeks 2 and 3, and 48 for the final comparison. None of the participants indicated to have (had) a profession in which audio quality plays a role. Mean and individual audiograms of the 52 participants included in the analysis of week 1 can be seen in Figure 1.

The study was approved by the ethics committee of the University of Lübeck (vote 2025-393) and executed at the German Institute of Hearing Aids, Lübeck, from October 2025 to March 2026. Data collection was performed by trained health care professionals. All participants gave written informed consent prior to participating in the study. They were recruited from an internal database and through public advertisements. Reimbursement of 200 Euros for study participation was provided, which was reduced in case they dropped out voluntarily.

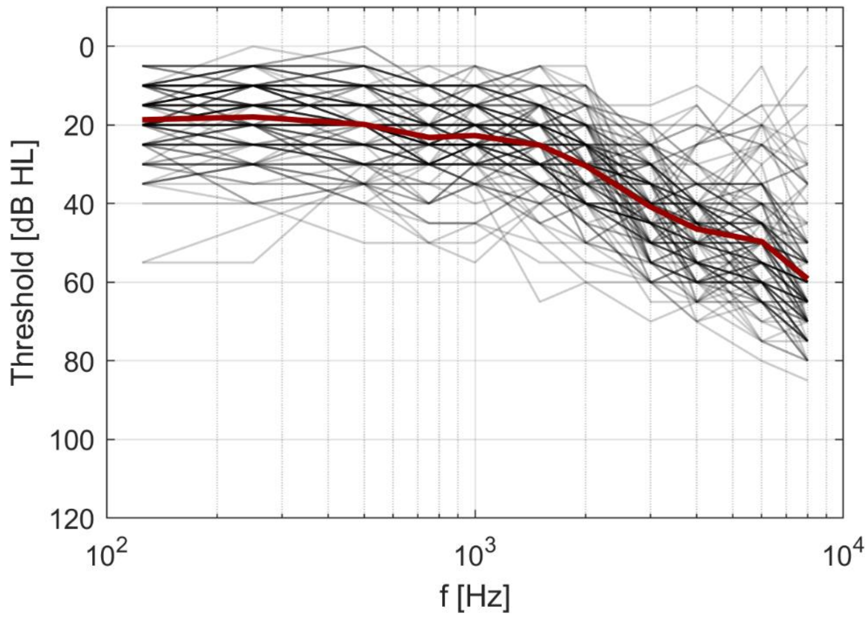

Figure 1: Mean (red line) and individual air conduction audiograms (grey lines) of the 52 participants who contributed sufficient data to be included in the analysis of week 1.

HA and fitting procedure

Participants were fitted with two pairs of Widex Allure RIC R D HAs with different processing delays (LD and HD). The LD HAs were fitted using Widex Compass Cloud and configured to include only one program consisting of the Widex PureSound program, which has a processing delay of 0.5 ms. Ear tips were chosen according to the software recommendation. Most participants used open domes (16), tulip domes (13), or a combination thereof (9). Fewer participants had vented click sleeves on one (11) or both (3) ears. The settings of the LD HAs were copied to the HD HAs, which have a specialized firmware that adds extra samples in the signal chain to bring the processing delay to 8 ms. Also, the feedback canceller of the HD HAs has been fine-tuned to this longer processing delay. As this implementation in real HAs is subject to constraints and may introduce artifacts, we performed a laboratory test to investigate potential differences in sound quality between LD and HD that are not caused by comb-filtering (see Supplemental Material1). Differences between HD and LD HAs are perceptible for completely closed ear canals (where the delay should not create any comb-filtering), but upon repeated exposure do not lead to any systematic preference. Differences between LD and HD HAs other than comb-filtering artifacts are therefore likely negligible in this field study.

Experimental procedure

The study flow is depicted in Figure 2. Participants started with one week of acclimatization in which they were provided with two pairs of HAs, one with LD and one with HD, and asked to perform a direct comparison of the HAs in selected situations. After the acclimatization week, they started a field trial where they wore each pair of HAs for a week and rated it both daily and at the end of it.

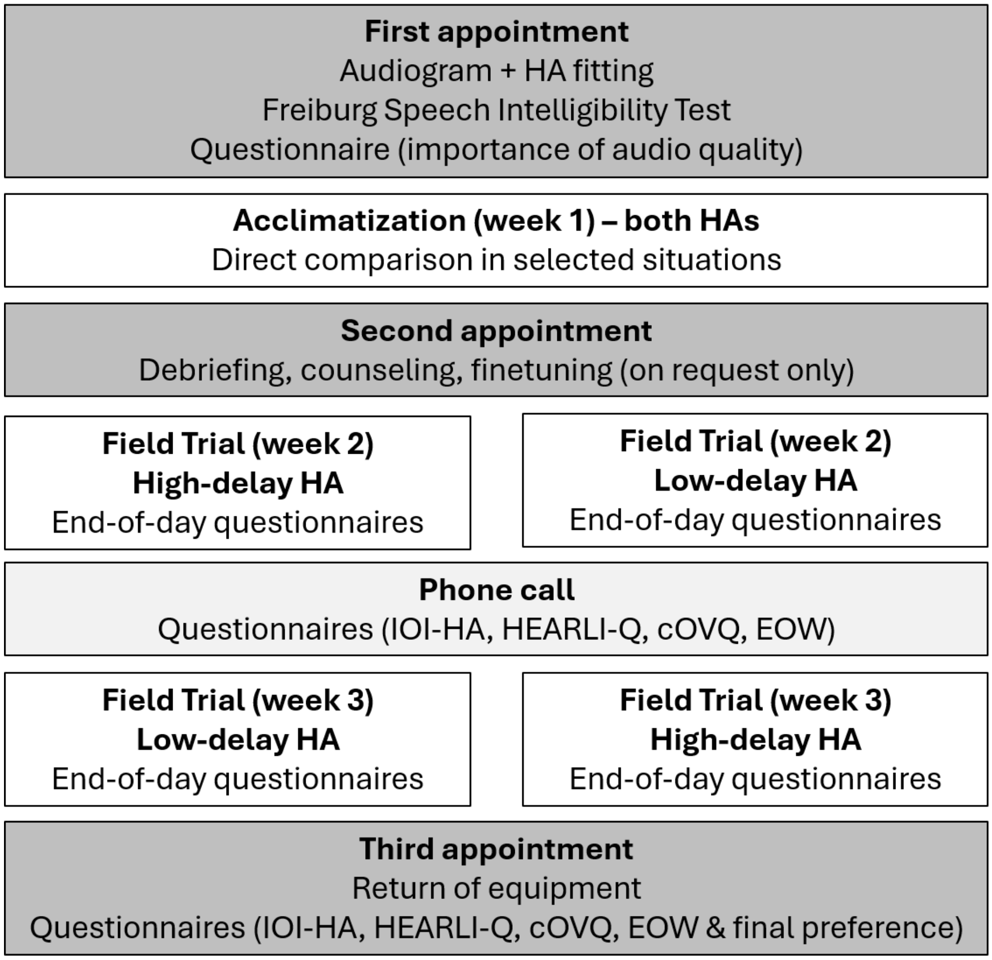

Figure 2: Study flow. Abbreviations: HA – hearing aid, IOI-HA – International Outcome Inventory for Hearing Aids, HEARLI-Q - Hearing Related Lifestyle questionnaire, EOW – end-of-week questionnaire, cOVQ – customized Own Voice Qualities questionnaire

First appointment

At the first appointment, the participant's audiogram was measured, including air conduction and bone conduction thresholds, as well as loudness discomfort levels. Participants were fitted with LD and HD HAs. The wearing order of LD and HD was randomized during the field trial. HAs were marked with red and blue dots (also randomized) and participants were blinded towards which HA they were wearing. The initial hearing aid delay used (LD or HD) was alternated between each participant as they were recruited into the study, and the allocation of colors to delays was changed after every other participant. After every four incoming participants, balancing of the experimental condition order was achieved and repeated. No blinding of the condition to the experimenter was possible. For the participants, the hearing aids were referred to as “blue pair” and “red pair”.

A Freiburg monosyllabic speech test using speech at 65 dB SPL and noise at 60 dB SPL was conducted unaided and with each of the two HAs to confirm adequate device fitting using two lists of 20 words for each test. During the first appointment, participants answered a questionnaire regarding the perceived importance of audio quality. It inquired whether participants sing or play an instrument, have a profession in which audio quality plays a role, own expensive audio equipment, enjoy music in CD quality more than lower-quality streamed music, and enjoy listening to music or audiobooks from a good speaker system more than via mobile phone or radio. We also asked how important music and audio quality are to them.

Acclimatization week and second appointment

During the first week after fitting, participants were encouraged to use both pairs of HAs to familiarize themselves with both the LD and HD conditions. There were instructions to use the HAs as much as possible but no wearing time constraints. Participants had the task to try out both pairs and seek out at least eight situations from a list of ten situations, including street noise, birds singing, newspaper rustling, kitchen sounds, sounds in a public place, listening to music, TV news, conversations on the street, conversation in the garden, and conversation at home. A description of what to expect in these situations (e.g. “newspaper rustling: Rustle with a newspaper or magazine. The paper sound should be sharper than without HAs but not uncomfortable”), taken from Signia Telecare (Froehlich et al., 2020), was handed out to participants. In those situations, and three additional situations of their choice in which they usually encounter hearing problems, participants were asked to answer a questionnaire directly comparing the two HAs. This comprised questions on HA satisfaction, satisfaction with their own voice, and listening effort for each HA, as well as a final preference.

After the acclimatization week, a second laboratory visit was conducted where the participants had the opportunity to discuss and address any encountered problems with the HAs to ensure smoother handling in weeks two and three. Fine-tuning of gain was only performed if changes in sound or loudness were specifically requested by participants, and always mirrored on both HA pairs.

Field study in weeks two and three

In weeks two and three, participants wore the HD and the LD HAs for one week each in a randomized crossover design. They were instructed to wear HAs for at least six hours per day and to answer an end-of-day (EOD) questionnaire every evening. This contained questions on HA satisfaction and satisfaction with sound quality and own voice, as well as naturalness of sound. These were rated on a five-point scale (from not at all satisfied/natural to extremely satisfied/natural) with an additional option to indicate that HAs were not worn. An additional question enquired about the HA use time in hours.

At the end of each week, participants answered a series of questionnaires, namely:

- The German version of the IOI-HA (Cox et al., 2002). This consists of questions regarding HA use, benefit, satisfaction, residual activity limitations, impact on others, residual participation restriction, and quality of life. The total score is often used as a measure of fitting outcome. The factors introspection and interaction, describing the HA users' relation to their HAs and their interaction with others, respectively, can also be computed.
- A modified version of the Hearing-Related Lifestyle Questionnaire (HEARLI-Q)(Lelic, Wolters, et al., 2022). This questionnaire, in its original form, enquires about HA satisfaction, perceived difficulty, importance of the situation, and frequency of the situation in 23 listening situations based on the Common Sound Scenario framework (Wolters et al., 2016). Here, the question on importance was omitted and instead a question on the occurrence of the situation in the last week was added. Due to the rating in 23 different situations, if experienced, it gives a more nuanced picture of HA satisfaction than the equivalent question of the IOI-HA, if ratings differ between situations.
- A customized version of the Own Voice Qualities questionnaire (OVQ) (Laugesen et al., 2011) as previously used by (Súsonnudóttir et al., 2026) .
- Questions regarding HA satisfaction, satisfaction with sound quality, satisfaction with own voice, and naturalness of the sound (referred to as end-of-week (EOW) questionnaire). This was identical to the EOD questionnaire but referred to the complete week.
- Final forced-choice preference for one of the two HAs phrased as a hypothetical buying decision (after the third week only).

All questionnaires were administered using the online survey platform Jotform. A link to the questionnaires, including a pre-filled pseudonym, was set up on the participants' cell phones during the first and second appointments. The possibility to borrow a phone for this purpose was not used by any participant.

Analysis

If not mentioned otherwise, descriptive plots show mean values with 95% confidence intervals obtained using bootstrapping with 20.000 bootstrap samples. Statistical analyses were conducted in R (R Core Team, 2023). Ordinal questionnaire outcomes (i.e., EOD questionnaire, HEARLI-Q, direct comparison in week 1 and final ratings) were analyzed using cumulative link mixed models (CLMMs) with a logit link using the *ordinal* package (Christensen, 2023). Model building followed a hierarchical approach. Participant-specific random slopes for delay by participant were evaluated against random-intercept-only models using likelihood-ratio tests. For EOW-questionnaires, each participant provided a single rating per HA condition. Therefore, EOW-models included participant-specific random intercepts only. Potential covariates (e.g., order, color the HA was marked in, day in the EOD questionnaire, and situation in both, HEARLI-Q and outcomes of the direct HA comparison in week 1) were retained only if they significantly improved model fit, based on likelihood-ratio tests and improvements in AIC. Final models are reported in the Results section.

During the acclimatization week (week 1), participants indicated which HA they preferred in that situation. Preference was analyzed using a mixed-effects logistic regression model with delay as a fixed effect and participant as a random intercept. For this model, the neutral response options “did not hear a difference” and “no preference” were excluded from the analysis. Responses indicating either a weak or strong preference (i.e., “rather LD/HD” and “definitively LD/HD”) were collapsed into a single preference category for the respective device, resulting in a binary outcome (LD vs. HD preference). Order and hearing-aid color assignment were evaluated as potential covariates but dropped as they did not improve model fit.

The composite IOI-HA scores Total Score, Introspection, and Interaction were analyzed using linear mixed-effects models with participant included as a random effect. These were fit using *lme4* (Bates et al., 2015), and p-values were obtained using *lmerTest* (Kuznetsova et al., 2017). Order was evaluated as a fixed effect but was dropped as it did not improve model fit. All model comparisons were based on likelihood-ratio tests and Akaike's Information Criterion (AIC). Odds ratios (OR) and 95% confidence intervals were derived from the final models and reported in the direction favoring the low-delay device.

For IOI-HA single attributes, several item-level cumulative link mixed models exhibited numerical instability due to sparse response distributions and the limited number of observations per participant. Therefore, individual items were evaluated using paired-samples t-tests and Hedges' g effect sizes. These were computed using Matlab and the effect size toolbox (Hentschke & Stüttgen, 2011). A forest plot with all odds ratios was created using *ggplot2* (Wickham, 2016). Generally, in case a questionnaire was answered multiple times, only the last time was evaluated. Results are based on all available data without any imputation of missing data.

Results

Objective speech intelligibility

The Freiburg monosyllabic speech test performed at the first appointment by 52 participants showed that speech understanding increased from 66.3% (CI: 62.1 – 70.0 %) without HAs to 77.9% (CI: 75.3 - 80.2) and 78.3% (CI: 75.4 – 80.8) with LD and HD HAs, respectively. There was no significant difference ($p=0.74$) in speech understanding with LD and HD HAs. However, speech understanding with both LD and HD HAs was significantly better than without HAs ($p<0.001$).

Direct comparisons (week 1)

Figure 3 shows the results of the direct comparisons during the acclimatization week for each situation. 52 participants performed a total of 657 comparisons, of which 124 were self-specified situations. The median number of comparisons per participant was 11 (range: 5 to 63). In 66.8% of all comparisons, the LD devices were rated first.

Ratings for sound quality, listening effort, and satisfaction with own voice (Figure 3a-c) were analyzed using a cumulative link mixed model (logit link). See Table 1 for the results of the statistical models. Fixed effects for the sound quality model included delay, situation, and the order of rating the HAs. Participant was included as a random intercept and delay as random slope. For sound quality, a small effect of presentation order was observed, with more favorable ratings for the HA that was rated second. For listening effort, rating order did not improve model fit and was therefore excluded from the model. In both models, inclusion of participant-specific random slope for delay significantly improved model fit compared with a random-intercept-only model. Delay significantly affected sound-quality ratings as well as listening-effort ratings, over and above the effects of other variables. Participants reported significantly greater listening effort and lower satisfaction with sound quality with the HD devices than with the LD devices. The corresponding odds ratios were 2.29 for sound quality and 1.82 for listening effort.

Satisfaction with own voice was rated in 450 situations. Fixed effects in the cumulative link mixed model included delay and situation. Participant was included as a random intercept and random slope for delay by participant. Inclusion of the random slope significantly improved model fit compared with a random-intercept-only model. Neither rating order nor color improved model fit and were therefore excluded from the final model.

Participants reported significantly lower satisfaction with their own voice when using the HD devices than when using the LD devices. The odds of reporting a more negative own-voice rating were 1.86 times higher in the HD condition than in the LD condition.

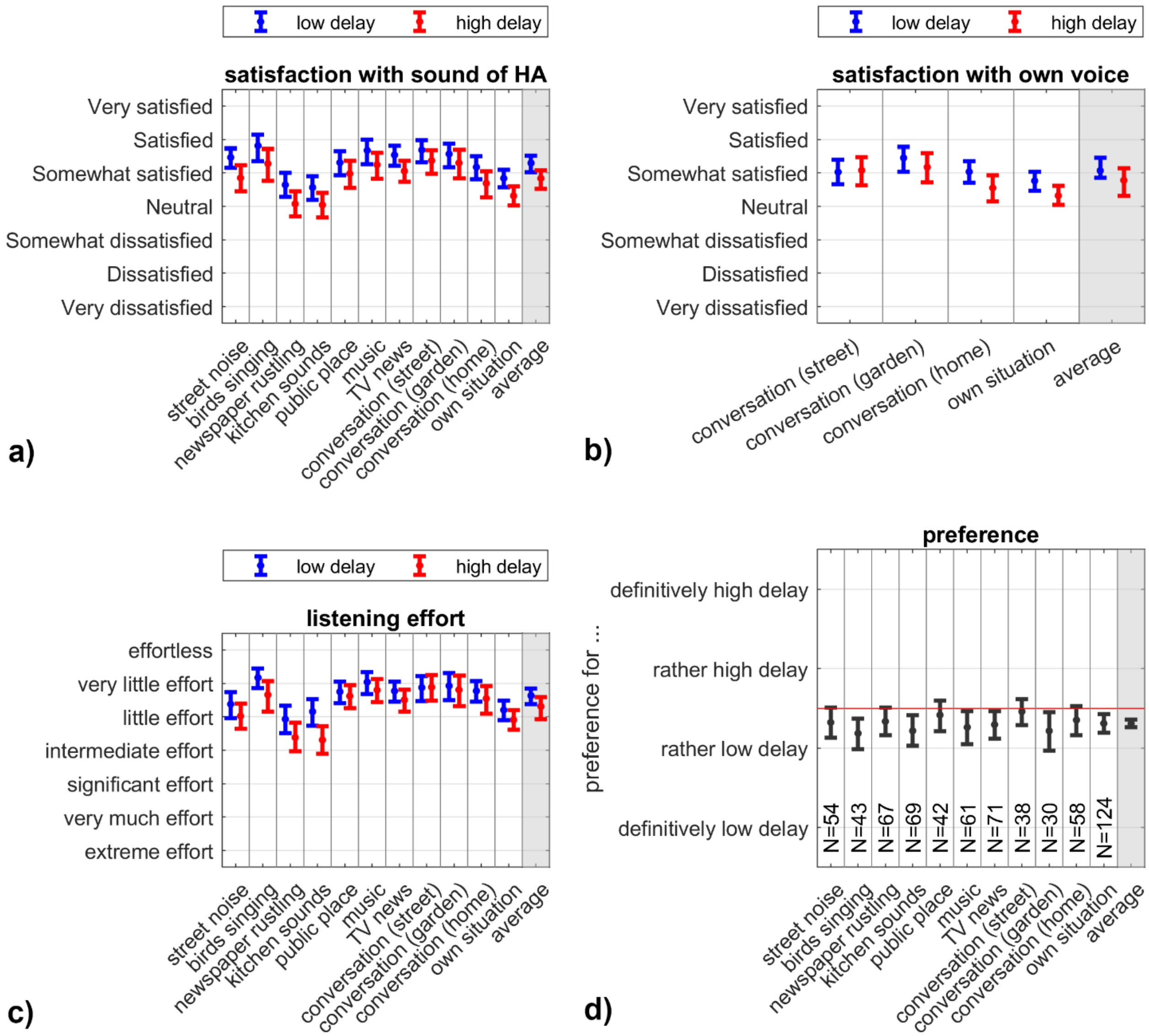

Figure 3: Mean ratings for each situation in the acclimatization week for a) satisfaction with the sound of the hearing aids, b) satisfaction with own voice (where applicable), c) listening effort, and d) preference. N denotes the number of responses obtained for each listening situation. Error bars denote 95% confidence intervals computed via bootstrapping. The average across situations is plotted in the grey shaded area to the right of each panel. P-Values and odds ratios are reported in Table1.

Mean preference ratings are shown in Figure 3d. Response options contained the two neutral response options 'I don't hear a difference' and 'I like both equally'. Pooled across all participants, these made up 30.0% and 9.4% of ratings, respectively. Preference judgments were analyzed using a generalized linear mixed model with a binomial link. Responses indicating a preference for either LD or HD were included, whereas responses indicating no preference were excluded, as described in more detail above. Participants showed a significant overall preference for the LD devices. When a preference was expressed, the odds of preferring the LD devices were 5.60 times greater than the odds of preferring the HD devices. Including situation category did not significantly improve model fit, suggesting that the preference for LD devices was broadly consistent across listening situations. As shown in Figure 3d, the strongest preference for LD devices was observed for music, TV news, kitchen sounds, conversation in the garden, and the self-selected situations, whereas preference was less pronounced for public-place and street-noise situations. Overall, these findings show that LD is advantageous compared to HD across a wide range of situations and for all four outcome measures (satisfaction with sound and own voice, listening effort and preference).

Outcome	β	SE	z	p	OR (LD vs HD)	95% CI
Direct comparison (week 1)						
Sound quality	0.83	0.27	3.10	0.002	2.29	[1.48, 4.16]
Own voice	0.62	0.30	2.06	0.040	1.86	[1.03, 3.36]
Listening Effort	0.60	0.22	2.72	0.007	1.82	[1.18, 2.78]
Preference	1.72	0.56	3.09	0.002	5.60	[2.05, 21.50]
EOD						
HA satisfaction	2.10	0.41	5.15	< 0.001	8.19	[3.67, 18.22]
Sound quality	2.19	0.43	5.14	< 0.001	9.00	[3.89, 20.70]
Naturalness	2.00	0.40	4.98	< 0.001	7.37	[3.36, 16.15]
Own voice	1.75	0.42	4.16	< 0.001	5.77	[2.53, 13.16]
EOW						
HA satisfaction	1.62	0.49	3.32	<0.001	5.06	[1.94; 13.20]
Sound quality	1.60	0.51	3.14	0.002	4.95	[1.82;13.45]
Naturalness	1.37	0.49	2.78	0.005	3.95	[1.50;10.40]
Own voice	2.22	0.79	2.82	0.005	9.21	[1.97; 43.01]
HEARLI-Q						
Satisfaction	1.07	0.33	3.21	0.001	2.92	[1.52, 5.62]
Difficulty	0.57	0.26	2.18	0.029	1.76	[1.06, 2.93]
Own Voice Questionnaire						
I like my voice	0.91	0.39	2.34	0.019	2.47	[1.16, 5.27]

Table 1: Outcomes of the statistical models for direct comparison in week 1, EOD and EOW questionnaires, HEARLI-Q and a selected question from the cOVQ questionnaire.

IOI-HA

Results of the IOI-HA are shown in Figure 4. In addition to single attributes (left panel), composite measures like the total score and the two factors, introspection ("Me and my hearing aids") and interaction ("Me and the rest of the world"), were computed and are shown in the right panel. Introspection, i.e., the relationship of the HA wearers with their HAs, is based on use time, benefit, satisfaction, and quality of life. Interaction is based on activity limitation, participation restriction, and impact on others.

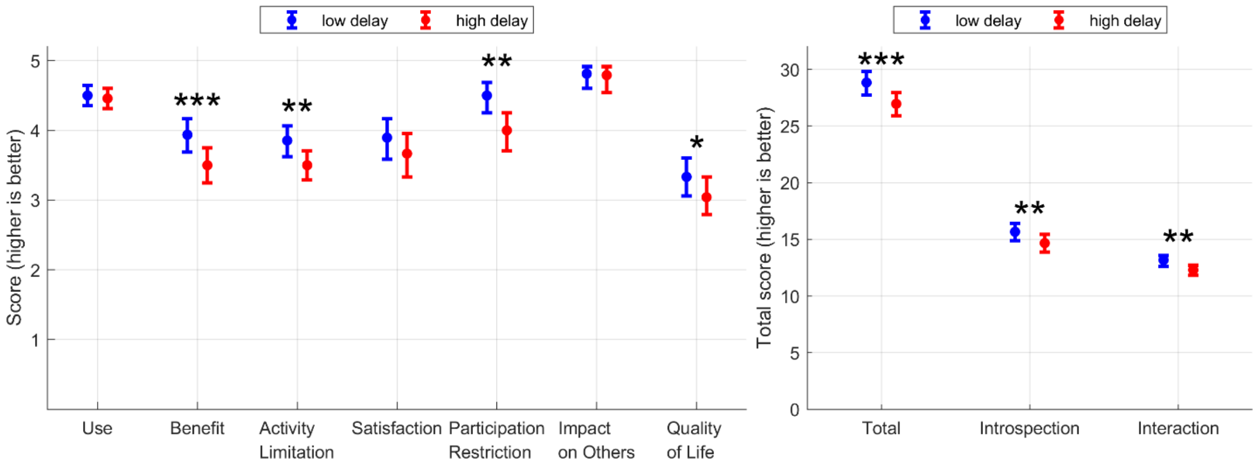

Figure 4: Results of IOI-HA: Results of single attributes are shown on the left; composite measures like total score and the two factors Introspection ("Me and my HA") and Interaction ("Me and the rest of the world") are shown in the right plot. Error bars denote 95% confidence intervals obtained using bootstrapping. Asterisks indicate significant differences between LD and HD (p<0.05; ** p <0.01; *** p < 0.001).*

The primary IOI-HA outcomes (total score and factor scores) were analyzed using linear mixed-effects models with participant included as a random effect. Presentation order did not improve the model fit and was therefore excluded from the final models. Outcomes are summarized in Table 1. The LD devices yielded significantly higher IOI-HA total scores than the HD devices, indicating a more favorable outcome of HA fit for LD. Similarly, scores on both the introspection and the interaction factor were higher for the LD devices than for the HD devices, suggesting that LD improved not only the relationship of the hearing aid users with their hearing aids, but also their social interaction.

Table 2: Coefficients, p-values, and effect sizes for all IOI-HA measures

IOI-HA Outcome	β	95% CI	d.f.	t-value	p	Hedges' g ± 95% CI
Composite measures						
Total score	1.88	[0.98, 2.77]	47	4.17	< 0.001	0.50 ± 0.27
Introspection	1.00	[0.40, 1.60]	47	3.33	0.002	0.36 ± 0.23
Interaction	0.88	[0.38, 1.37]	47	3.47	0.001	0.54 ± 0.33
Single attributes						
Use	-	-		-	0.420	0.08 ± 0.20
Benefit	-	-		-	< 0.001	0.49 ± 0.27
Activity Limitation	-	-		-	0.002	0.46 ± 0.30
Satisfaction	-	-		-	0.102	0.21 ± 0.26
Participation Restriction	-	-		-	0.002	0.56 ± 0.37
Impact on Others	-	-		-	0.860	0.04 ± 0.41
Quality of Life	-	-		-	0.012	0.30 ± 0.24

Item-level analyses were considered exploratory and were evaluated using two-sided paired-samples t-tests with Hedges' g effect sizes. Intermediate-sized, significant effects in favor of LD were observed for benefit, activity limitation, and participation restriction (see Table 1). A somewhat smaller but significant effect was shown for quality of life, also indicating a higher quality of life for LD. No significant difference was observed for use, satisfaction, and impact on others. Note that the attribute use is likely influenced by the instruction to use HAs at least six hours per day.

HEARLI-Q

HEARLI-Q enquires about HA satisfaction and perceived difficulty in 23 everyday situations. Pooled results of the HEARLI-Q questionnaire are shown in Figure 5. Participants reported a comparable number of experienced situations for the LD (mean = 15.4) and HD (mean = 15.6) devices.

HEARLI-Q satisfaction and difficulty ratings were analyzed using cumulative link mixed models with a logit link. Participant-specific random slopes for delay significantly improved model fit, whereas rating order did not. Situation significantly improved model fit. A delay × situation interaction did not improve fit and hence was not included. The final model included delay and situation as fixed effects and participant-specific random intercepts and random slopes for delay. In the final model, delay significantly affected satisfaction ratings as well as perceived difficulty (see Table 2), with higher HA satisfaction and lower perceived difficulty being reported for the LD HAs. The odds of receiving a more favorable satisfaction rating were nearly three times higher for the LD devices than for the HD devices, and participants were nearly twice as likely to report greater listening difficulty with the HD devices than with the LD devices.

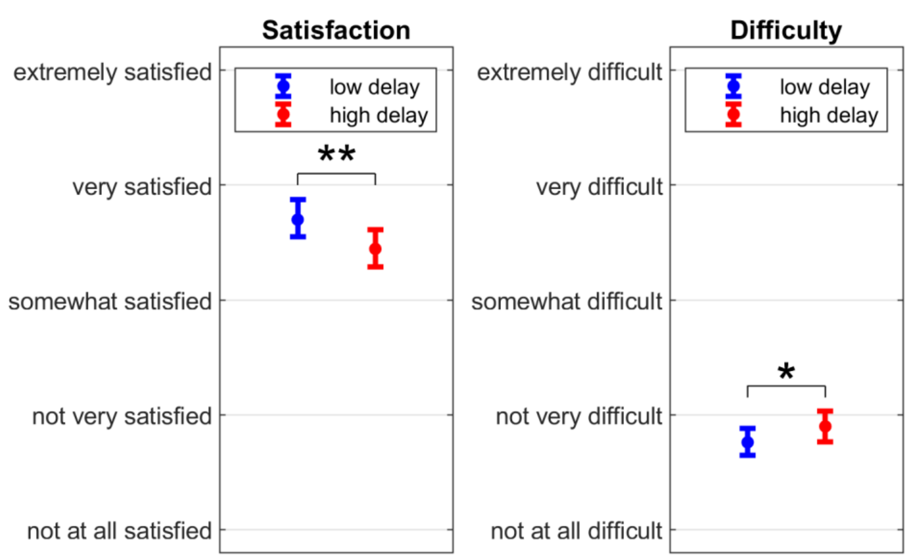

Figure 5: HA satisfaction and perceived difficulty ratings from the HEARLI-Q questionnaire. Error bars denote 95% confidence intervals obtained by bootstrapping.

End-of-day questionnaires

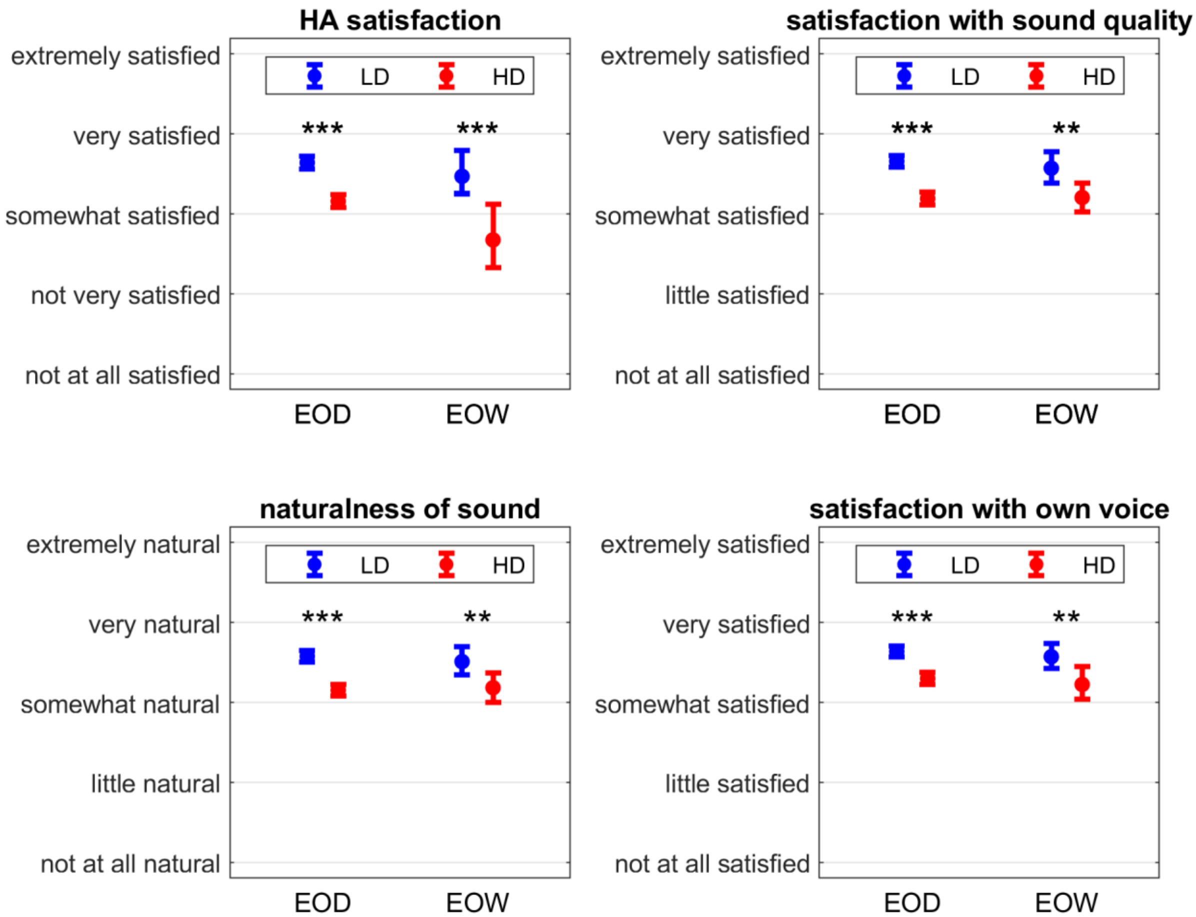

Figure 6: Ratings in the end-of-day (EOD) questionnaire and the end-of-week (EOW) questionnaire for hearing aid satisfaction, satisfaction with sound quality, naturalness of sound, and satisfaction with own voice. Error bars depict 95% confidence intervals computed using bootstrapping.

On average, participants completed 13.6 questionnaires over the total field-trial period of 14 days. Self-reported wearing time was on average 9.1 h (CI: 8.8 – 9.5) for LD and 8.9 h (CI: 8.6 – 9.3) for HD. EOD questionnaire ratings are shown in Figure 6. They were analyzed using cumulative link mixed models with a logit link. For all four questionnaire outcomes, inclusion of a participant-specific random slope for delay significantly improved model fit compared with random-intercept-only models (all $p < 0.001$). Neither rating order nor study day improved model fit for any outcome and were therefore excluded from the final models. Consequently, all final models included delay as a fixed effect and participant-specific random intercepts and random slopes for delay.

Delay significantly affected all end-of-day ratings, with lower ratings for the HD than for the LD devices for hearing aid satisfaction, sound quality, naturalness, and own-voice satisfaction. The odds of giving a more favorable rating were between approximately six-and nine-fold greater for the LD devices than for the HD devices. There was no evidence that these effects changed over the observation week, as study day did not improve model fit for any outcome.

End-of-week questionnaires

EOW ratings, collected at the end of weeks 2 and 3, are shown in Figure 6 along with the EOD ratings. Because the EOW questionnaire contained only one rating per HA and participant, EOW ratings were analyzed using cumulative link mixed models with participant-specific random intercepts only (no random slope). Delay significantly affected all four outcome measures. When using the HD devices, participants reported lower hearing aid satisfaction, lower satisfaction with sound quality, lower naturalness ratings, and lower satisfaction with their own voice than when using the LD devices (see Table 1). The odds of receiving a more favorable rating were approximately four- to five-fold higher for the LD devices across HA satisfaction, sound quality, and naturalness. For satisfaction with own voice, they were nine times as high.

Own Voice Qualities questionnaire

Responses to the shortened Own-Voice Qualities questionnaire were generally consistent with the own-voice outcomes in week 1 and in the EOD questionnaire. Participants were significantly more likely to agree with the statement *"I like my own voice"* when using the LD devices than when using the HD (see Table 1). Detailed item-level results are shown in Supplementary Figure S4.

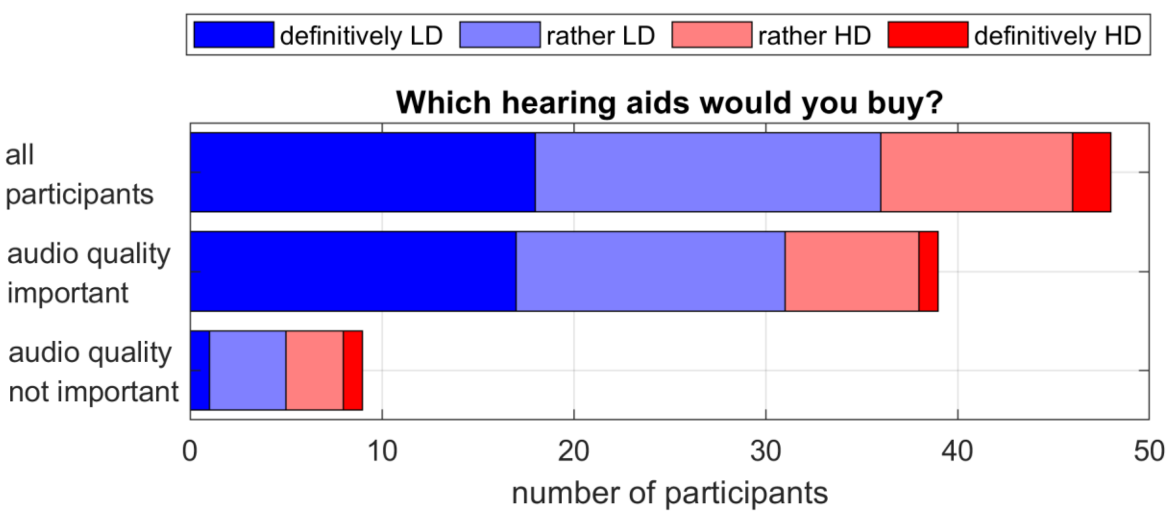

Figure 7: Final preference at the end of the study based on responses from 48 participants (two participants did not provide an answer to this question as they decided against buying a HA).

Final preference

The final preference question at the end of the study was phrased as a buying decision question to facilitate forced choice. Results are shown in Figure 7. Thirty-six of 48 participants (75 %) preferred the LD HAs. Among the 39 participants who indicated that audio quality was important to them (collapsing the categories 'rather important' and 'very important'), 79% preferred the LD devices. In contrast, LD preference for the remaining nine participants was 56% and hence close to chance level (50%).

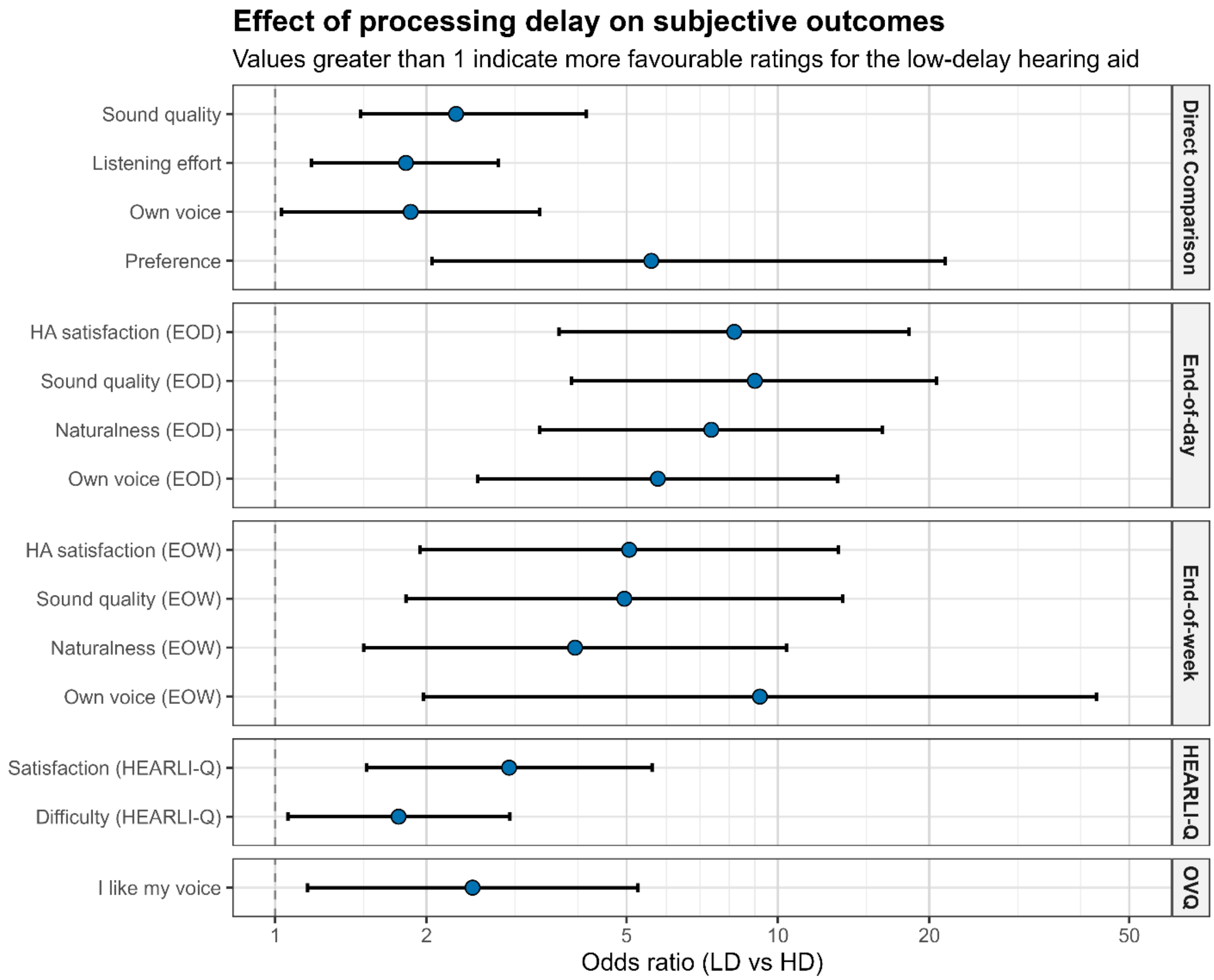

Figure 8: Overview plot with odds ratios for all ratings evaluated in the study (composite measures of IOI-HA were analyzed differently). All measures favor LD, with the strongest effects for EOD.

Comparison across measures

As summarized in Figure 8, all questionnaire outcomes favored the LD devices. Odds ratios ranged from 1.76 for HEARLI-Q difficulty to 9.21 for the final own-voice rating. Effects were generally larger for the EOD and EOW ratings than for the direct-comparison ratings and HEARLI-Q outcomes, indicating that the advantage of the LD devices was evident not only during immediate comparisons but also during prolonged real-world use. Importantly, the advantage was observed across multiple independent measures of satisfaction, sound quality, naturalness, own-voice perception, perceived benefit, and hearing-related difficulty.

Although effect sizes from different questionnaires are not directly comparable because of differences in questionnaire design, scaling, and sampling, it is interesting that the largest odds ratios were observed for the EOD and EOW ratings. The advantage of the LD devices was therefore evident not only in situation-specific assessments such as the direct-comparison task and HEARLI-Q, but also in retrospective evaluations summarizing experiences across an entire day or week.

Reliability of questionnaire responses

Subjective outcome measures are potentially susceptible to inattentive or insufficient-effort responding. It is not possible to find suitable measures for careless responses for each questionnaire employed in this study. However, several observations suggest that at least some responses may contain an elevated level of careless responses.

One indication was obtained from the HEARLI-Q questionnaire. Participants reported the occurrence of a range of listening situations in their daily lives twice, once for each HA condition. Because the occurrence of these situations should be independent of processing delay, substantial differences between repeated reports may indicate inconsistent responding. Small differences (i.e., by one response category) are expected because the response scale (never, a few times per year, monthly, weekly, daily) is relatively coarse and the true answer may fall in between two given response options. However, five participants showed repeated discrepancies of three or more categories, suggesting reduced response consistency.

A second indication comes from the direct comparisons during week 1, where several participants completed the questionnaires for all situations within a relatively short time interval and in exactly the order in which the situations appeared in the questionnaire. This pattern may reflect retrospective entry in the online form of responses which were previously recorded elsewhere, which some participants reported they had done. However, given the extra effort this causes, it may also indicate that some responses were provided without carefully evaluating each listening situation individually. For four participants, the longest time interval between any two completed questionnaires was less than five minutes.

As quantification of careless responses is difficult and exclusion of participants is prone to introducing bias, we did not exclude any participants from the analysis. Participants were blinded to the HA condition throughout the study. Consequently, any careless responding would be expected to introduce random variability rather than systematically favor one delay condition over the other. The presence of such measurement noise would be expected to attenuate true effects. Therefore, if inattentive responding occurred, the reported differences between the LD and HD devices are more likely to represent conservative estimates of the underlying effects than overestimates.

Discussion

In this study, we analyzed the satisfaction with and benefit of a low processing delay of 0.5 ms compared to a typical 8 ms delay in first-time HA users. The lower processing delay was advantageous across a multitude of measures, such as direct-comparison ratings, EOD and EOW assessments, HEARLI-Q outcomes, IOI-HA measures, and overall preference decisions. Participants reported higher satisfaction, better sound quality, greater naturalness, lower listening difficulty, and more favorable own-voice perception with the LD devices. There were no measurable differences in speech intelligibility between the two hearing-aid conditions.

The direct comparison in selected situations during week 1 was added to the study design, as direct comparison usually is more sensitive to small changes in sound quality than indirect comparison (see e.g. Carbajo Benito et al., 2026). However, due to practical constraints, participants here had to change HAs, which made the comparison less quick and direct than switching hearing programs. Also, first-time HA users may still need to get used to the handling of HAs. Nevertheless, the first week of direct comparison doubled as an acclimatization time, making weeks two and three more comparable and hence may have contributed to the clear results for those two weeks.

Participants preferred LD devices over HD devices already during the acclimatization week. They were more than five times as likely to prefer LD than HD and reported better sound quality, lower listening effort, and greater satisfaction with their own voice with LD devices. The advantage of a low processing delay was also apparent in EOD and EOW ratings, HEARLI-Q responses, and IOI-HA outcomes. In fact, the largest odds ratios were observed for the EOD and EOW ratings. Although effect sizes obtained from different questionnaires cannot be directly compared because of differences in questionnaire design and scaling, this suggests that the subjective benefit of reduced delay continues during extended real-world use. Despite two weeks of real-world use, there was no evidence that participants adapted to the higher processing delay, as study day did not improve the EOD models and the EOW ratings continued to favor the low-delay devices.

In the IOI-HA, participants obtained higher total scores with the LD devices as well as higher scores on the factors Introspection and Interaction. The Introspection factor reflects the user's relationship with the HA, including benefit, satisfaction, and quality of life. The Interaction factor reflects hearing-related participation and communication in everyday life. The presence of effects on both factors suggests that reduced delay influences not only perceived sound quality but also broader aspects of hearing-aid benefit and participation. The largest item-level effects were observed for benefit, activity limitation, and participation restriction. This also suggests that improvements in sound quality and own-voice perception translate into meaningful improvements in everyday functioning.

Total scores of the IOI-HA, as well as factor scores, found here are comparable to the results of other recent studies with first-time HA users (Cañete et al., 2024; Houmøller et al., 2022; Laakso et al., 2023). While the total scores obtained here may be influenced by the instruction to wear HAs at least 6 hours per day, factor 2 ("interaction") should be independent of this. For LD, the interaction factor is 13.2 (or 4.4 when averaging instead of summing the contributing single attributes) and hence slightly higher than the 12.7 reported by Houmøller et al. (2022) or the 4.3 average across attributes found by Cañete et al. (2024) pooled across two different fitting procedures.

Interestingly, there were significant differences in both listening effort during the acclimatization week as well as perceived difficulty in HEARLI-Q, despite no significant difference in the Freiburg speech test. This highlights that clinically used speech intelligibility tests in the laboratory do not necessarily capture potential differences in subjective listening difficulty, which, however, is very relevant in real life.

There was no significant interaction of situation and processing delay in HEARLI-Q, which could indicate that the advantage of LD is stable across a multitude of situations. This would be in line with the finding that situation had no significant effect on preference during the acclimatization week. However, in both direct comparison in week 1 as well as for HEARLI-Q in weeks 2 and 3, the frequency of situations differed widely, and especially some of the more difficult situations might have been underpowered. The number of experienced situations in HEARLI-Q was, however, comparable with both processing delays. So, there is no evidence of situations being avoided with one processing delay.

During direct comparisons, end-of-day ratings, final ratings, and the dedicated own-voice questionnaire, participants consistently preferred their own voice when using the LD devices. This is consistent with a laboratory study by Súsonnudóttir et al. (2026) which showed a higher quality of own voice with 0.5 ms than with 10 ms.

Across multiple analyses (i.e., direct-comparison ratings, EOD ratings, and HEARLI-Q outcomes), inclusion of participant-specific random slopes for delay substantially improved model fit. This finding indicates substantial inter-individual variability in sensitivity to hearing-aid processing delay. While most participants preferred the LD devices, the magnitude of this preference varied considerably across individuals. For counselling help-seeking individuals with hearing impairment on the suitable technology, it would hence be beneficial to know which factors predict strong LD preference. Potential contributing factors are hearing loss and gain (Bramsløw, 2010; Lelic, Stiefenhofer, et al., 2022), as well as spectral processing ability (Súsonnudóttir et al., 2024).

In the current study, participants who reported that audio quality was important to them showed a stronger preference for the LD devices than participants who assigned lower importance to audio quality. Although the sample size of the latter group was small, this observation suggests that individual expectations and priorities may play a role in shaping preferences for reduced processing delay. While this seems contradictory to Súsonnudóttir et al. (2024), that study used questions regarding the perception and willingness to pay for audio equipment as a proxy for audio quality importance. Here, similar questions were included but were not analyzed further, as the question "How important is a good sound quality to you?" predicted LD preference better. If the relationship between audio importance and LD preference can be confirmed by larger studies, a simple question regarding the importance of audio quality may be useful when counselling new hearing-aid users and selecting hearing-aid technologies.

Generally, all measures collected in this study jointly paint a clear picture that low delay is preferable for first-time HA users. This is in line with previous research in the laboratory showing that a 0.5 ms processing delay is preferred over higher processing delays (Lelic, Stiefenhofer, et al., 2022; Súsonnudóttir et al., 2024, 2026) and leads to improved neural representation (Zhou et al., 2024) and better auditory temporal and spectral processing abilities (Küçük & Çetinkaya, 2026). While we did not study HA uptake here, we would expect that the increase in benefit and satisfaction shown here will have a positive impact on HA adoption and use of HAs for help-seeking individuals. However, further research is necessary to investigate this further.

Potential limitations

The duration of this study was limited to three weeks. Although benefits of reduced delay were already evident, given that HA acclimatization typically lasts several months (e.g., Wentzel et al., 2025), longer-term studies are required to determine whether these advantages persist over months of hearing-aid use.

Implementation of the HD condition required modifications to the hearing-aid signal processing chain that introduced artifacts that were small but detectable in a laboratory test (see Supplemental Material 1). This test comprised recording with HD and LD HAs using a head and torso simulator (HATS) with either open or closed ear canals. The hypothesis was that for the open fitting condition the direct and amplified sound would interact and create comb-filtering in the HD condition. For the closed fitting condition, the HATS ear canals were completely sealed to remove the comb-filter effect. Normal hearing participants listened to the recordings via headphones and rated their preference for either LD or HD. Preferences for LD were smaller for sealed ear canal condition than for the open condition. More importantly, preference disappeared with repeated exposure in the closed fitting condition, while it stayed stable for the open-fit recordings. While we cannot rule out that some artifacts in the HD condition affected the study results, these findings suggest that artifacts alone are unlikely to explain the consistent pattern of results in this study, which showed advantages of LD also for extended real-world use.

Some signs of insufficient effort responding were observed in various questionnaire responses. Indicators included unusually rapid completion of direct-comparison questionnaires, despite requiring a change of location, and inconsistent reports of the normal occurrence of HEARLI-Q situations, which should be independent of the HA tested. As first-time HA users are mostly new recruits, it may not be surprising to see varying degrees of motivation and more insufficient effort responses than in studies with experienced participants, who often already participated in many other studies. Because participants were blinded to the HA condition, there is no obvious mechanism by which careless responding would systematically favor one delay condition over the other. Such responding may, however, have increased variability in the questionnaire outcomes and hence decreased Hedges' g effect sizes reported for IOI-HA.

Conclusion

The present study demonstrates that reducing HA processing delay from 8 ms to 0.5 ms provides measurable real-world benefits for first-time HA users. The LD devices were preferred overall and yielded higher ratings of sound quality, HA satisfaction, naturalness, own-voice perception, perceived benefit, and hearing-related participation. The consistency of these findings across a wide range of outcome measures suggests that processing delay is an important determinant of user experience in open-fit HAs and may contribute to improved acceptance of HA technology.

Acknowledgments

We thank Lea Hackenberg, Pia Stoppe, Florian Mitteldorf, Carolin Schnetzer, and Tobias Babel for data collection and Borgny Súsonnudóttir Hansen for many fruitful discussions in planning and interpreting this study.

Statements and declarations

AI-assisted writing statement.

Portions of the manuscript text were drafted and edited with assistance from Microsoft Copilot. All statistical analyses were performed by the authors. The authors reviewed, revised, and approved all AI-assisted text and take full responsibility for the content of the manuscript.

Ethical considerations

The study was approved by the ethics committee of the University of Lübeck (vote 2025-393).

Consent to participate

All participants gave written informed consent prior to participating in the study.

Consent for publication

Not applicable.

Declaration of Conflicting Interests

Nadja Schinkel-Bielefeld, Frederik Vejby Nielsen, Laura Winther Balling, Federica Bianchi, and Filip Rønne are employees of WS Audiology.

Funding

The study was funded by WS Audiology.

Data availability statement

Fully anonymized datasets from this study are available upon reasonable request.

References

Agnew, J., & Thornton, J. M. (2000). Just Noticeable and Objectionable Group Delays in Digital Hearing Aids. *Journal of the American Academy of Audiology*, *11*(06), 330–336. https://doi.org/10.1055/s-0042-1748062

Balling, L. W., Townend, O., Stiefenhofer, G., & Switalski, W. (2020). Reducing hearing aid delay for optimal sound quality: A new paradigm in processing. *The Hearing Review*, *Vol 27*(No 4), 20–26.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Borre, E. D., Kaalund, K., Frisco, N., Zhang, G., Ayer, A., Kelly-Hedrick, M., Reed, S. D., Emmett, S. D., Francis, H., Tucci, D. L., Wilson, B. S., Kosinski, A. S., Ogbuoji, O., & Sanders Schmidler, G. D. (2023). The Impact of Hearing Loss and Its Treatment on Health-Related Quality of Life Utility: A Systematic Review with Meta-analysis.

Journal of General Internal Medicine, *38*(2), 456–479. https://doi.org/10.1007/s11606-022-07795-9

Bramsløw, L. (2010). Preferred signal path delay and high-pass cut-off in open fittings. *International Journal of Audiology*, *49*(9), 634–644. https://doi.org/10.3109/14992021003753482

Cañete, O. M., Loquet, G., Sánchez-López, R., Hougaard, D. D., Schnack-Petersen, R., Gaihede, M., Schmidt, J. H., Hammershøi, D., & Neher, T. (2024). Auditory Profile-Based Hearing Aid Fitting: Self-Reported Benefit for First-Time Hearing Aid Users. *Audiology Research*, *14*(1), 183–195. https://doi.org/10.3390/audiolres14010017

Carbajo Benito, R., Lelic, D., & Schinkel-Bielefeld, N. (2026). Comparing hearing aid settings with ecological momentary assessment using direct and indirect approaches. *The Journal of the Acoustical Society of America*, *159*(3), 2810–2821.

Christensen, R. H. B. (2023). *ordinal—Regression Models for Ordinal Data* [R package version 2023.12-4.1]. https://CRAN.R-project.org/package=ordinal

Cox, R. M., Stephens, D., & Kramer, S. E. (2002). Translations of the international outcome inventory for hearing aids (IOI-HA). *International Journal of Audiology*, *41*(1), 3–26.

Denk, F., Hieke, T., Roberz, M., & Husstedt, H. (2023). Occlusion and coupling effects with different earmold designs—All a matter of opening the ear canal? *Int J Audiol*, *62*(3), 227–237. https://doi.org/10.1080/14992027.2022.2039966

Denk, F., Ohlmann, K., & Kollmeier, B. (2021). Detection mechanisms for processing delays in simulated vented hearing devices. *JASA Express Letters*, *1*(1), 014402. https://doi.org/10.1121/10.0003064

Denk, F., Wiederschein, L., Kemper, M., & Husstedt, H. (2024). (Why) Do Transparent Hearing Devices Impair Speech Perception in Collocated Noise? *Trends Hear*, *28*, 23312165241246597. https://doi.org/10.1177/23312165241246597

Dixon, P. R., Feeny, D., Tomlinson, G., Cushing, S., Chen, J. M., & Krahn, M. D. (2020). Health-related quality of life changes associated with hearing loss. *JAMA Otolaryngology–Head & Neck Surgery*, *146*(7), 630–638.

Froehlich, M., Branda, E., & Apel, D. (2020). User Engagement with Signia TeleCare: A Way to Facilitate Hearing Aid Acceptance. *AudiologyOnline*, *Article 26463*.

Goehring, T., Chapman, J. L., Bleeck, S., & Monaghan*, J. J. M. (2018). Tolerable delay for speech production and perception: Effects of hearing ability and experience with hearing aids. *International Journal of Audiology*, *57*(1), 61–68. https://doi.org/10.1080/14992027.2017.1367848

Groth, J., & Birkmose, M. (2004). Disturbance caused by varying propagation delay in non-occluding hearing aid fittings. *International Journal of Audiology*, *43*(10), 594–599. https://doi.org/10.1080/14992020400050076

Hentschke, H., & Stüttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *European Journal of Neuroscience*, *34*(12), 1887–1894.

Houmøller, S. S., Wolff, A., Möller, S., Narne, V. K., Narayanan, S. K., Godballe, C., Hougaard, D. D., Loquet, G., Gaihede, M., Hammershøi, D., & Schmidt, J. H. (2022). Prediction of successful hearing aid treatment in first-time and experienced hearing aid users: Using the International Outcome Inventory for Hearing Aids. *International Journal of Audiology*, *61*(2), 119–129. https://doi.org/10.1080/14992027.2021.1916632

ITU-T. (1996). *P.800: Methods for subjective determination of transmission quality*. https://www.itu.int/rec/t-rec-p.800-199608-i

Kochkin, S. (2000). MarkeTrak V: Why my hearing aids are in the drawer: The consumer's perspective. *The Hearing Journal*, *53*(2), 34–41.

Kochkin, S. (2010). MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing. *Hear J*, *63*(1), 19–32.

Küçük, S., & Çetinkaya, Ü. C. (2026). Impact of hearing aid processing delay on temporal and spectral auditory processing. *International Journal of Audiology*, *0*(0), 1–10. https://doi.org/10.1080/14992027.2026.2642750

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Laakso, M., Lipsanen, J., Pajo, K., Salmenlinna, I., Aaltonen, T., Ruusuvuori, J., & Aarnisalo, A. (2023). Working-age first-time hearing aid users' self-reported outcomes. *Int J Audiol*, *62*(9), 877–885. https://doi.org/10.1080/14992027.2022.2106454

Laugesen, S., Jensen, N. S., Maas, P., & Nielsen, C. (2011). Own voice qualities (OVQ) in hearing-aid users: There is more than just occlusion. *International Journal of Audiology*, *50*(4), 226–236. https://doi.org/10.3109/14992027.2010.547991

Lelic, D., Stiefenhofer, G., Lundorff, E., & Neher, T. (2022). Hearing aid delay in open-fit devices: Preferred sound quality in listeners with normal and impaired hearing. *JASA Express Letters*, *2*(10), 104803. https://doi.org/10.1121/10.0014950

Lelic, D., Wolters, F., Herrlin, P., & Smeds, K. (2022). Assessment of Hearing-Related Lifestyle Based on the Common Sound Scenarios Framework. *American Journal of Audiology*, *31*(4), 1299–1311. https://doi.org/https://doi.org/10.1044/2022_AJA-22-00079

Li, C.-M., Zhang, X., Hoffman, H. J., Cotch, M. F., Themann, C. L., & Wilson, M. R. (2014). Hearing Impairment Associated With Depression in US Adults, National Health and Nutrition Examination Survey 2005–2010. *JAMA otolaryngology-- head & neck surgery*, *140*(4), 293–302. https://doi.org/10.1001/jamaoto.2014.42

Livingston, G., Huntley, J., Liu, K. Y., Costafreda, S. G., Selbaek, G., Alladi, S., Ames, D., Banerjee, S., Burns, A., Brayne, C., Fox, N. C., Ferri, C. P., Gitlin, L. N., Howard, R., Kales, H. C., Kivimaki, M., Larson, E. B., Nakasujja, N., Rockwood, K., … Mukadam, N. (2024). Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet*, *404*(10452), 572–628. https://doi.org/10.1016/S0140-6736(24)01296-0

McCormack, A., & Fortnum, H. (2013). *Why do people fitted with hearing aids not wear them?* Taylor & Francis. https://doi.org/10.3109/14992027.2013.769066

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Snyder, N., Bixler, A., Belcastro, K., Hunter, J. B., & Middleton, I. L. (2025). Hearing Aid Adoption Rates: Perceptions, Attitudes, and Influencing Factors of Prospective Hearing Aid Users. *Journal of the American Academy of Audiology*, *36*(3), 188–195. https://doi.org/10.3766/jaaa.240024

Stiefenhofer, G. (2022). Hearing aid delay in open-fit devices–coloration-pitch discrimination in normal-hearing and hearing-impaired. *International Journal of Audiology*, 1–9.

Stone, M. A., Moore, B. C., Meisenbacher, K., & Derleth, R. P. (2008). Tolerable hearing aid delays. V. Estimation of limits for open canal fittings. *Ear and Hearing*, *29*(4), 601–617.

Súsonnudóttir, B., Kowalewski, B., Stiefenhofer, G., & Neher, T. (2024). Individual Differences Underlying Preference for Processing Delay in Open-Fit Hearing Aids. *Trends in Hearing*, *28*, 23312165241298613. https://doi.org/10.1177/23312165241298613

Súsonnudóttir, B., Mosgaard, L. D., Stiefenhofer, G., Souza, P., & Neher, T. (2026). Own-voice perception with different processing delays in open-fit hearing aids. *Trends Hear*, *submitted*, 23312165261435260. https://doi.org/10.1177/23312165261435260

Súsonnudóttir Hansen, B. (2025). *The Effects of Short Processing Delay on Sound Quality in Open-Fit Hearing Aids: Candidacy, Preference, and Real-Life Experiences*. https://doi.org/10.21996/92e441ce-f90e-4387-aeea-0905b359db1e

Wentzel, C., Swanepoel, D. W., Mahomed-Asmail, F., Beukes, E., Dawes, P., Munro, K., Almufarrij, I., & Manchaiah, V. (2025). Auditory Acclimatization in New Adult Hearing Aid Users: A Registered Systematic Review of Magnitude, Key Variables, and Clinical Relevance. *Journal of Speech, Language, and Hearing Research : JSLHR*, *68*(7), 3445–3479. https://doi.org/10.1044/2025_JSLHR-24-00856

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Winkler, A., Latzel, M., & Holube, I. (2016). Open versus closed hearing-aid fittings: A literature review of both fitting approaches. *Trends Hear*, *20*, 2331216516631741.

Wolters, F., Smeds, K., Schmidt, E., Christensen, E. K., & Norup, C. (2016). *Common Sound Scenarios – A context-driven categorization of everyday sound environments for application in hearing-device research*. IHCON, Lake Tahoe, CA, USA.

Zhou, M., Soleimanpour, R., Mahajan, A., & Anderson, S. (2024). Hearing Aid Delay Effects on Neural Phase Locking. *Ear and Hearing*, *45*(1), 142–150. https://doi.org/doi:%2010.1097/AUD.0000000000001408

Supplemental Material 1: Hearing aid validation study

Motivation

Here we performed a real-life study where participants wore identical hearing aids that only differed in the processing delay. To achieve this, for the longer processing delay the auditory signal was loaded into the memory of the hearing aid for 7.5 ms, and the feedback canceler was optimized for this processing delay. Implementing both very low and high processing delays in the same device while keeping the remaining signal processing identical can be challenging. To implement the high-delay condition, we used the low-delay processing configuration and introduced the additional delay by buffering the audio signal in memory. However, the bit depth available for this buffering was insufficient to preserve the full dynamic range of the audio signal. The signal level therefore had to be selected as a compromise between minimizing the risk of clipping at high signal levels and minimizing self-noise at low signal levels. As a result, the high-delay implementation involved a trade-off between occasional faint clipping artifacts at high levels and slightly increased self-noise. To evaluate whether our implementation of the longer processing delay introduced any artifacts that could have influenced the results of the study, we performed an audio quality test in the laboratory with recordings of both hearing aids.

For hearing aids where nothing but the processing delay varies, the only perceivable difference is in the interaction between direct and processed sound, i.e. the "comb filtering" effect. If the ear canal is completely sealed, no direct sound enters the ear canal, and recordings with LD and HD hearing aids should be identical if no artifacts are introduced by the implementation. Consequently, a listening test should not find any systematic differences between them. In case there are detectable artifacts, we would like to know how big their impact on perceived audio quality is compared to the impact of the comb-filtering effect. Hence, we also included recordings with the same fitting used in the study.

Methods

Sound recordings and audio material

The recordings for this listening test were done in an acoustically treated listening room where the audio was delivered over 45 loudspeakers placed in a 3D array. The sound scenes were reproduced using mixed-order ambisonics and presented at their natural level. Additionally, a recording of an audiobook was rendered through a realistic spatial room impulse response and presented at a high and low level, e.g., 65 and 86 dB SPL.

For each sound scene, approx. 10-second segments are chosen to include either acoustic crackling, comb filtering effects, or both. The full list of sound scenes and their overall presentation level can be found in Table S1.

Table S1: Stimuli used in the listening test

	Duration [s]	Presentation Level [dB SPL]
Audio Book Loud 1	10	85
Audio Book Loud 2	12	85
Audio Book Quiet	12	65
Cafe	10	68
Children playing	10	71
Conference room	10	55
Hair drying	10	62
Rain	10	70

The Widex Allure RIC R D hearing aids were configured to two different hearing loss profiles (see Table S2) and two different processing delays (LD and HD) and placed on a KEMAR using custom-made closed couplings. Afterwards, the ear canals on the KEMAR were acoustically sealed using Blu Tack, ensuring both a completely closed coupling and that the hearing aid receivers did not move between configurations.

Subsequently, similar recordings were done with the recommended open or semi-open hearing aid ear tips, according to the hearing loss, to include both forms of acoustic coupling.

Table S2: Hearing loss used for fitting the hearing aids for the recordings of sound samples in the validation test.

	250 Hz	500 Hz	1000 Hz	1500 Hz	2000 Hz	4000 Hz	6000 Hz
Minimum	5	0	10	15	20	25	25
Maximum	30	40	50	45	45	70	75

Listening test

16 individuals (5 male) between the ages of 20 and 50 years with self-reported normal hearing participated in the test. All participants gave informed consent before beginning the listening test.

As our focus was the impact on audio quality, rather than detectability of any potential artifacts, we chose an ITU-P.800 comparison category rating test design (ITU-T, 1996).

Tests were implemented through SenseLabOnline, an online listening test facility provided by Force Technology. See Figure S1 for a screenshot of the user interface. The tests were performed in an acoustically treated listening room, using a laptop with a Focusrite Scarlett 2i2 sound card and Beyer Dynamics DT 770 Pro (250 Ohm) headphones.

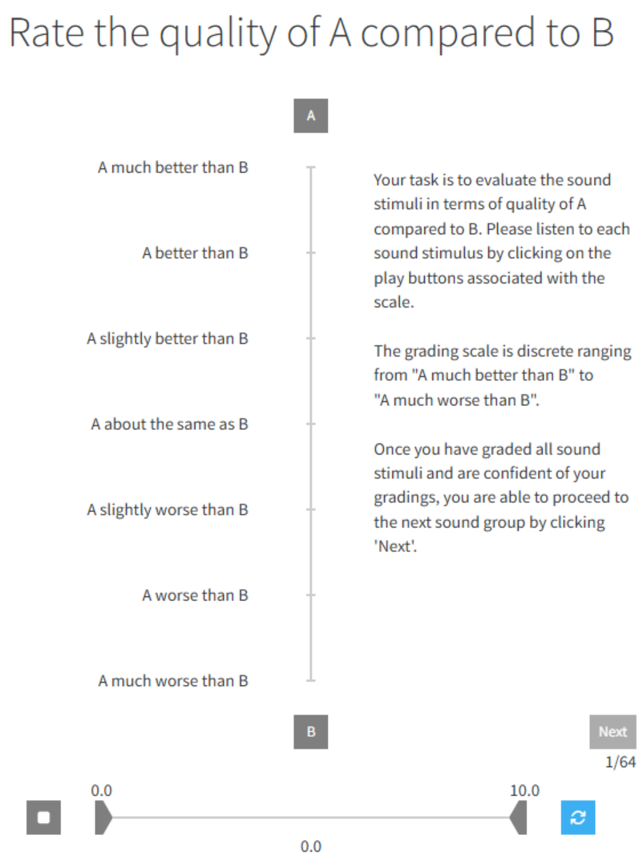

Figure S1: user interface of the data validation test: Comparison category rating test implemented through SenseLabOnline.

The test contained 64 trials, resulting from 8 audio samples, 2 hearing aid fittings (for maximum and minimum hearing loss included in the field study), 2 conditions (open and closed fitting, and two repetitions of the same stimuli. Assignment of HD and LD to A and B was randomized but reversed in the second repetition. All stimuli were presented in random order, with open and closed fittings presented interleaved.

Results

The median duration of the test (excluding instructions) was 27.5 min (range: 14.1-44.3 min). As there was a significant difference between the first and second repetition of each sample, they were analyzed separately. Ratings were mapped to a scale from -3 to 3. Mean ratings can be seen in Figure S2. Effect sizes and p-values from two-sided paired sample t-tests are shown in Table S3. While the median duration of the trials for the second repetition was reduced to 84% (CI: 78 % – 90 %), there was no significant difference in the trial duration of open or closed coupling recordings, neither in the first nor in the second repetition. However, generally, trial duration decreased throughout the listening test (see Figure S3).

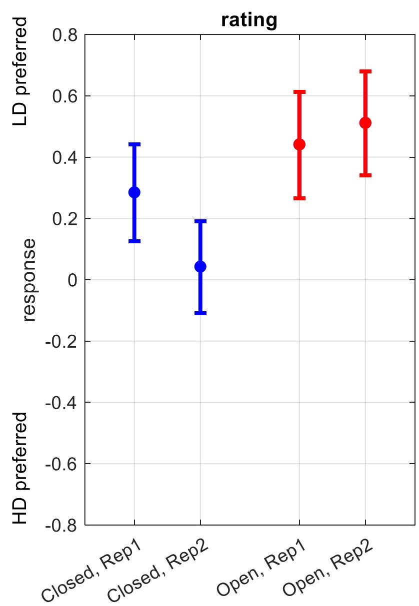

Figure S2: Mean ratings for the closed and the open condition by repetition.

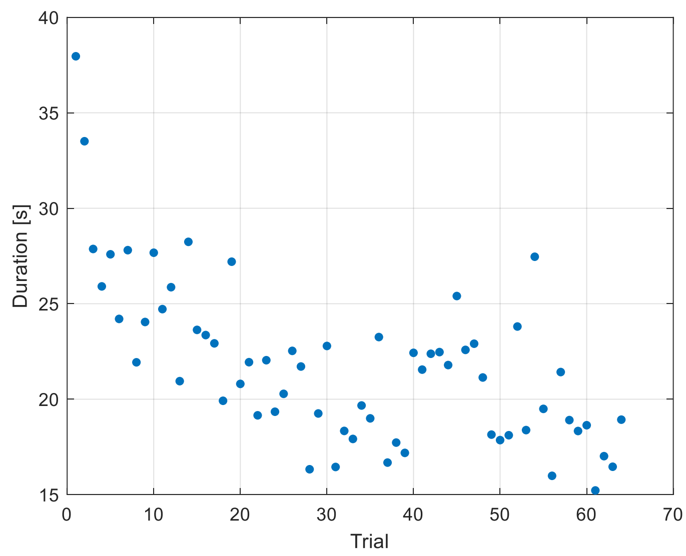

Figure S3: Trial duration averaged across audio samples and conditions.

Table S3: Effect sizes and p-values for the open and the closed conditions.

Condition		Mean	CI	Cohens' d	CI	p-value
CLOSED	Rep1	0.29	0.13; 0.44	0.22	0.10; 0.35	P<0.001
	Rep2	0.04	-0.11; 0.20	0.03	-0.09; 0.16	P=0.59
	Rep2 – Rep1	-0.24	-0.44; -0.05	-0.15	-0.27; -0.03	P=0.02
OPEN	Rep1	0.44	0.27; 0.61	0.31	0.18; 0.44	P<0.001
	Rep2	0.51	0.34; 0.68	0.37	0.24; 0.50	P<0.001
	Rep2 – Rep1	0.07	-0.11; 0.26	0.05	-0.8; 0.17	P=0.46

Discussion

For the open condition, there was a small effect for the difference between HD and LD, preferring the reduced comb-filtering of the LD condition. This effect was stable across repetitions, meaning participants were able to perceive comb-filtering also with the reduced effort in the second repetition. This would be consistent with a slight training effect.

For the closed condition, there was a small but significant effect in the first repetition. This means there have been perceivable artifacts introduced when implementing the added delay. However, as there are no significant differences any longer in the second repetition, it seems that these artifacts either are so hard to perceive that most participants did not perceive them anymore with the reduced effort spent in the second repetition or they got used to it so quickly that they did not mind those artifacts anymore. To the extent that it is possible to draw conclusions from a laboratory test for real life, this indicates that the artifacts introduced during the implementation of HD are not the main driver of the results found in the main study – though we cannot exclude that they contributed to the effects found.

Supplemental Material 2: Results of the cOVQ questionnaire

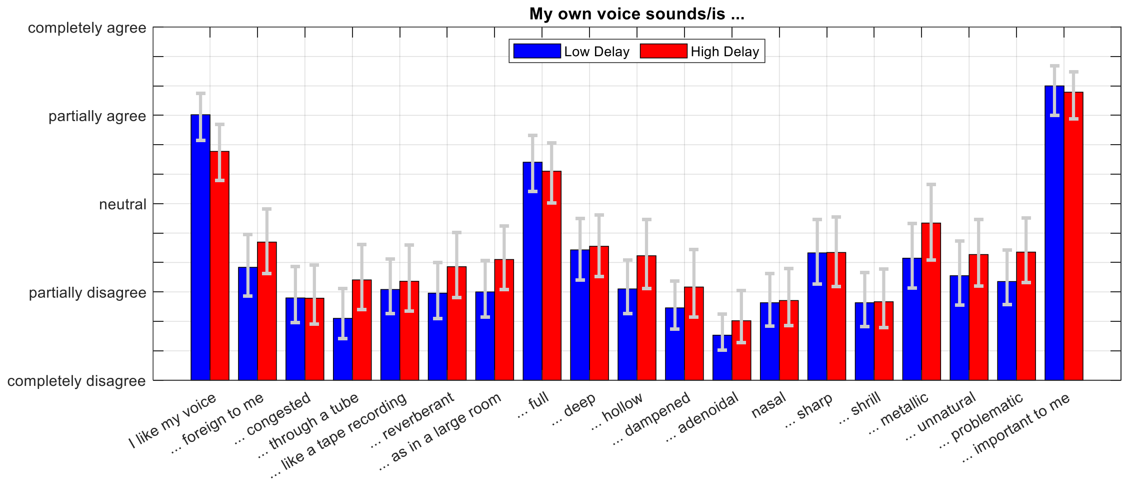

Figure S4: Results of the customized Qualities of Own Voice questionnaire. Error bars denote 95% confidence intervals.