Measuring hearing aid satisfaction in everyday listening situations: Retrospective and in-situ assessments complement each other

Dina Lelic, Florian Wolters, Nadja Schinkel-Bielefeld.

Affiliations below.

DOI: 10.1055/a-2265-9418

Please cite this article as: Lelic D, Wolters F, Schinkel-Bielefeld N. Measuring hearing aid satisfaction in everyday listening situations: Retrospective and in-situ assessments complement each other. Journal of the American Academy of Audiology 2023. doi: 10.1055/a-2265-9418

Conflict of Interest: All authors are employees of WS Audiology.

Abstract:
Background: Recently, we developed a hearing-related lifestyle questionnaire (HEARLI-Q) which asks respondents to rate their hearing aid (HA) satisfaction in 23 everyday listening situations. It is unknown how HA satisfaction on the retrospective HEARLI-Q scale compares to HA satisfaction measured on the same scale implemented in Ecological Momentary Assessments (EMA).
Purpose: To learn how retrospective (HEARLI-Q) and in-situ (EMA) assessments can complement each other.
Research Design: An observational study.
Study Sample: Twenty-one experienced HA users.
Data Collection and Analysis: The participants first filled out the HEARLI-Q questionnaire, followed by a one-week EMA trial using their own hearing aids. HA satisfaction ratings were compared between the two questionnaires and the underlying drivers of discrepancies in HA satisfaction ratings were evaluated.
Results: HA satisfaction scores were significantly higher in EMA for speech communication with one or several people. Hearing difficulty in these situations was rated higher in HEARLI-Q than in EMA, but occurrence of those difficult listening situations was also rated to be lower. When comparing only the situations that occur on daily or weekly basis, the two questionnaires had similar HA satisfaction ratings.
Conclusions: Lower occurrence of difficult listening situations seems to be the key driver of discrepancies in HA satisfaction ratings between EMA and HEARLI-Q. The advantage of EMA is that it provides insight into an individual’s day-to-day life and is not prone to memory bias. HEARLI-Q, on the other hand, can capture situations which occur infrequently or are inconvenient to report in the moment. Administering HEARLI-Q and EMA in combination could give a more holistic view of HA satisfaction.

Corresponding Author:
Dr. Dina Lelic, WS Audiology, Lynge, Denmark, dina.lelic@wsa.com, dina.lelic@wsa.com

Affiliations:
Dina Lelic, WS Audiology, Lynge, Denmark
Florian Wolters, ORCA Europe, WS Audiology, Stockholm, Sweden
Nadja Schinkel-Bielefeld, WS Audiology, Erlangen, Germany
Measuring hearing aid satisfaction in everyday listening situations: Retrospective and in-situ assessments complement each other

Dina Lelic¹*, Florian Wolters², Nadja Schinkel-Bielefeld³

1. WS Audiology, Lynge, Denmark; 2. ORCA EU, WS Audiology, Stockholm, Sweden; 3. WS Audiology, Erlangen, Germany

*Corresponding author:
Address: Nymøllevej 6, 3540 Lynge, Denmark
Email: dina.lelic@wsa.com

Abstract

Background: Recently, we developed a hearing-related lifestyle questionnaire (HEARLI-Q) which asks respondents to rate their hearing aid (HA) satisfaction in 23 everyday listening situations. It is unknown how HA satisfaction on the retrospective HEARLI-Q scale compares to HA satisfaction measured on the same scale implemented in Ecological Momentary Assessment (EMA).

Purpose: To learn how retrospective (HEARLI-Q) and in-situ (EMA) assessments can complement each other.

Research Design: An observational study.

Study Sample: Twenty-one experienced HA users.
Data Collection and Analysis: The participants first filled out the HEARLI-Q questionnaire, followed by a one-week EMA trial using their own hearing aids. HA satisfaction ratings were compared between the two questionnaires and the underlying drivers of discrepancies in HA satisfaction ratings were evaluated.

Results: HA satisfaction scores were significantly higher in EMA for speech communication with one or several people. Hearing difficulty in these situations was rated higher in HEARLI-Q than in EMA, but occurrence of those difficult listening situations was also rated to be lower. When comparing only the situations that occur on daily or weekly basis, the two questionnaires had similar HA satisfaction ratings.

Conclusions: Lower occurrence of difficult listening situations seems to be the key driver of discrepancies in HA satisfaction ratings between EMA and HEARLI-Q. The advantage of EMA is that it provides insight into an individual’s day-to-day life and is not prone to memory bias. HEARLI-Q, on the other hand, can capture situations which occur infrequently or are inconvenient to report in the moment. Administering HEARLI-Q and EMA in combination could give a more holistic view of HA satisfaction.

Key words: Hearing aid satisfaction; EMA; Retrospective report

Abbreviations: HA – Hearing Aid; HEARLI-Q – Hearing-Related Lifestyle Questionnaire; CoSS – Common Sound Scenarios; EMA – Ecological Momentary Assessment; GHABP - Glasgow Hearing Aid Benefit Profile
Introduction

Overall hearing aid (HA) satisfaction depends upon the individually weighted improvements one perceives with HA treatment. An individual’s interpretation of HA satisfaction could considerably vary, and people may use different criteria when judging whether they are satisfied with the HA\(^1\). For example, one might think about comfort of fit, streaming possibilities / connectivity, speech understanding, or different aspects of sound quality. Moreover, certain situations may be given more weight than others based on frequency of occurrence, importance, severity of hearing difficulty, recency or relationship to the present\(^2\) which can bias the overall HA satisfaction rating.

We developed the hearing-related lifestyle questionnaire (HEARLI-Q)\(^3\) that asks participants to rate 23 everyday listening situations on frequency of occurrence, importance to hear well, difficulty to hear and HA satisfaction. The HEARLI-Q situations are grouped in seven listening task categories: speech communication (two people, more than two people, and through device), focused listening (live sounds, through media device), and nonspecific (monitoring surroundings and passive listening), as defined by the Common Sound Scenarios (CoSS) framework\(^4\). In a study where the HEARLI-Q was administered four times (days 1, 2, 15 and 29), we found that experienced HA users’ responses were reliable across short (day 1 vs day 2) and longer (day 1 vs day 15 or 29) time spans\(^3\). Nonetheless, HEARLI-Q is a retrospective questionnaire, and as such, is at risk of being subject to memory bias\(^5,6\). Ecological Momentary Assessment (EMA), on the other hand, relies on participants repeatedly reporting in their momentary listening environment, reducing memory bias\(^7\). Because of this, EMA has gotten recent traction in audiology community to evaluate HAs in real-life\(^8\). However, to achieve a balance between getting high quality data and minimizing participant burden, EMA in hearing research is normally not administered for longer periods than four weeks,
although on most occasions, the period is shorter and more likely to fall within 1-2 weeks. This means that situations that occur less frequently than the EMA period may not be represented. Moreover, since EMA requires participants to answer in the moment, some situations may be unsafe (e.g., when driving) or inappropriate (e.g., in the middle of an important meeting) to report in. Consequently, certain situations are underrepresented in EMA\textsuperscript{9, 10}. Participants may also be less likely to interact with the EMA application in a moment that is already challenging\textsuperscript{11} and these situations are often the most interesting when evaluating HAs.

One way to include assessments of those challenging situations in EMA is to ask participants to rate HA performance based on a short retrospective period\textsuperscript{12-15}. For example, Wu et al\textsuperscript{15} asked their participants to answer EMA-prompted Glasgow Hearing Aid Benefit Profile (GHABP) questionnaire every 1.5 hours based on the last 1.5 hours for one week. The authors compared these EMA ratings to the ones obtained by the standard retrospective GHABP to assess outcomes of two different HAs. While both EMA-based (in-situ) GHABP and retrospective GHABP showed a significant difference between the two HAs on the satisfaction sub-scale, only the EMA-based GHABP was significantly different on benefit and residual disability sub-scales.

Recently, we conducted a study where we utilized both HEARLI-Q and EMA based on the same questions & response alternatives to investigate the effect of a “positive focus” intervention on HA satisfaction and benefit\textsuperscript{16}. Similar to Wu et al.\textsuperscript{15}, we observed comparable effects on HA satisfaction between retrospective and in-situ questionnaires, where participants who were asked to focus on positive listening experiences for two weeks after HA fitting had higher ratings relative to the control group. Although it was not the aim of the study to compare HA satisfaction ratings between questionnaires, EMA ratings tended to be higher than HEARLI-Q ratings. In other words, although both questionnaires were sensitive to detect a difference due to the intervention, when evaluating
HA performance, the satisfaction rating itself may be over/under-estimated depending on the type of questionnaire utilized.

In the current study, we set out to investigate how HA satisfaction ratings on the HEARLI-Q scale compare to HA satisfaction measured on the same scale implemented in an EMA trial. The goal was to compare both, the overall HA satisfaction as well as CoSS listening task specific HA satisfaction. In case any significant discrepancies were observed, we investigated what drove these inconsistencies in ratings. The overarching aim was to learn how the two types of assessments can complement each other to give a holistic view of HA experience.

**Methods**

Ethical clearance for conducting the study was obtained from the Research Ethics Committee of the Capital Region of Denmark (case no. H-18056647). The data in the current work were collected during August-November 2021.

**Participants**

Twenty-one HA users with mild-moderate hearing losses were enrolled in the study (6 females, 15 males with an average age of 66 years (SD: 7 years)). All the participants were experienced HA users (> 1 year), smartphone users and fluent in Danish. Nineteen participants were Widex HA users and two used Oticon HAs. Twenty participants reported to use their HAs all day, whereas one reported to use it sporadically. Ten participants were retired at the time of the study. Exclusion
criterium was severe cognitive impairment that would preclude the ability to perform the necessary tasks, as judged by the audiologist who did the recruiting. The participants were recruited through an internal database of participants via phone or e-mail. They were informed about the study orally and in writing. Before the trial commenced, the participants gave their informed consent in writing.

**Study design**

The participants were asked to download and install the MyHearingExperience app (Lenox UG, Herrsching, Germany), which is available for iOS and Android, on their own smartphones. Each participant was provided with a unique study log-in code for the app and instructed to fill out the HEARLI-Q available in the app, on the day of the initial log-in. Starting the day after, the participants were prompted to answer an EMA questionnaire every two hours between 9:00 and 21:00 over seven days. That is, the total number of prompts was 49. The EMA questionnaire remained available from the initial prompt until the participant completed it, or the next questionnaire was prompted (two hours later or next morning for the 21.00 o’clock questionnaire). The EMA questionnaire asked the participants to indicate their location, listening task (same categories as in HEARLI-Q) and noisiness of the situation. Further, the participants were asked to rate the frequency of occurrence, importance to hear well, hearing difficulty and HA satisfaction in the situation on the same scale as HEARLI-Q. The specific EMA questions and response alternatives are outlined in Supplemental Table 1. The full HEARLI-Q questionnaire can be found in the Supplemental Materials of Lelic et al.³

The participants wore their own HAs throughout the trial period and all the ratings were based on their experiences with own devices.
Data analysis

Based on 1000 simulations of HEARLI-Q HA satisfaction data with mean=3.4 and standard deviation = 0.6, twenty-one participants would enable difference detection of 0.5 scale points with power of approximately 80% using mixed-effects linear regression. The mean and standard deviation were taken from existing internal HEARLI-Q data.

Mixed-effects linear regression with random intercept for participant was conducted to analyze predictors of HA satisfaction ratings within each of the two methods. The dependent variable was HA satisfaction, and the covariates were frequency of occurrence, importance to hear well and difficulty to hear.

For comparison between methods, the overall HEARLI-Q HA satisfaction score was calculated by averaging the satisfaction scores of individual situations, as described in Lelic et al.\textsuperscript{3} HEARLI-Q HA satisfaction scores were further averaged for each of the seven CoSS task categories. To ensure that the results of EMA and HEARLI-Q could be directly compared and analyzed in the same statistical model, EMA overall and CoSS task specific HA satisfaction scores were also calculated by averaging the scores within each individual.

Repeated measures analysis of variance (RM-ANOVA) was conducted to compare the overall and CoSS task specific HA satisfaction scores between HEARLI-Q and EMA. For those CoSS task categories where there were significant differences between HEARLI-Q and EMA, further analyses were done to understand what these differences can be attributed to. Specific details of these analyses are reported in results / Supplemental Material. Correlation between the HEARLI-Q and EMA HA satisfaction ratings was analyzed using Pearson’s correlation.
The dependent variables subjected to ANOVA were visually inspected to ensure that they fit a normal distribution. If not, the ladder of powers was applied to transform the data. The residuals for all the mixed-effects linear regression analyses were visually inspected for normality.

All the statistical analyses were done in Stata (v. 15, StataCorp, College Station, TX, USA).

Results

EMA compliance and location / listening task distribution across reports

The total number of collected reports was 845. EMA compliance was on average 82% (SD: 15%) (submitted reports / prompted reports). Participants responded on average within 21 minutes of the prompt (see Supplemental Figure 1). Most of the reports were filled in home environment without conversation / focused listening activity, followed by 1-1 conversation and focused listening (media) (see Supplemental Figure 2).

HA satisfaction, HA use and predictors of HA satisfaction

The entire range of the HA satisfaction scale was used in both EMA and HEARLI-Q, although most responses were in the moderate-very satisfied range (87% responses in EMA and 72% of responses in HEARLI-Q). In HEARLI-Q, two participants indicated they do not wear their hearing aids when “Passive listening - Hearing sounds of nature.” In the EMA questionnaire, 82 reports attributed to 16 participants were submitted while they were not wearing their HAs – 63 of those reports were in “situation without a conversation or focused listening”, eight reports were in “conversation with one person”, two reports were in “conversation with several people”, one report was in “conversation
over phone or another technical device” and two reports were in “focused listening (media)” situation.

Hearing difficulty was an independent predictor of HA satisfaction in both questionnaires, where satisfaction decreased with increased hearing difficulty (see Table 1).

Table 1. Mixed-effects linear regression results for HEARLI-Q and EMA. Significant $P$-values are shown in bold.

<table>
<thead>
<tr>
<th>HEARLI-Q</th>
<th>EMA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable</strong></td>
<td><strong>Independent variable</strong></td>
</tr>
<tr>
<td>HA Satisfaction</td>
<td>Occurrence</td>
</tr>
<tr>
<td></td>
<td>(reference: ~2 times a year)</td>
</tr>
<tr>
<td></td>
<td>Monthly</td>
</tr>
<tr>
<td></td>
<td>Weekly</td>
</tr>
<tr>
<td>Importance</td>
<td>(reference: not important)</td>
</tr>
<tr>
<td></td>
<td>A little important</td>
</tr>
<tr>
<td></td>
<td>Moderately important</td>
</tr>
<tr>
<td></td>
<td>Very important</td>
</tr>
<tr>
<td></td>
<td>Extremely important</td>
</tr>
<tr>
<td>Hearing difficulty</td>
<td>(reference: not difficult)</td>
</tr>
<tr>
<td></td>
<td>A little difficult</td>
</tr>
<tr>
<td></td>
<td>Moderately difficult</td>
</tr>
<tr>
<td></td>
<td>Very difficult</td>
</tr>
<tr>
<td></td>
<td>Extremely difficult</td>
</tr>
</tbody>
</table>

Comparison of HA satisfaction ratings

The overall HA satisfaction was rated 3.4 ± 0.6 in HEARLI-Q and 3.5 ± 0.4 in EMA ($F_{1,20}=2.06$, $P=.17$). The overall HA satisfaction ratings were positively correlated between the two
questionnaires ($r = 0.58$, $P < .01$). Comparison of HA satisfaction ratings for individual CoSS task categories is presented in Figure 1. Not all the participants contributed with ratings for “Speech Communication – Through Device” and “Focused Listening – Live Sounds” in either of the questionnaires, although the number of contributing participants was lower in EMA than HEARLI-Q for both CoSS task categories. Eighteen participants contributed with HA satisfaction ratings for “Focused Listening – Through Media” in EMA. HA satisfaction was significantly higher in EMA for “Speech Communication – One Person” ($F_{1,20}=7.88$, $P=.01$) and “Speech Communication – Several People” ($F_{1,20}=5.38$, $P=.03$). There was not a significant difference between the two questionnaires for “Speech Communication – Through Device” ($F_{1,9}=0.04$, $P=.84$), “Focused Listening – Live Sounds” ($F_{1,5}=0.31$, $P=.59$), “Focused Listening – Through Media” ($F_{1,17}=3.85$, $P=.07$), or “No Conversation / No Focused Listening” ($F_{1,20}=3.93$, $P=.06$).

![Hearing Aid Satisfaction by CoSS Task](image)

This article is protected by copyright. All rights reserved.

Accepted Manuscript

This article is protected by copyright. All rights reserved.

Accepted Manuscript
Figure 1. Mean and 95% confidence intervals for hearing aid satisfaction ratings in HEARLI-Q and EMA by CoSS task category. The number of contributing participants is indicated below each error bar. Significant differences are denoted by asterisk (*).

Reasons for higher HA satisfaction ratings in EMA

1 - How do occurrence and hearing difficulty compare between HEARLI-Q and EMA?

Figure 2 shows occurrence, hearing difficulty and HA satisfaction ratings for the two CoSS task categories where HA satisfaction was rated higher in EMA than HEARLI-Q. For both categories, speech communication in quiet in HEARLI-Q was rated similarly to EMA on occurrence, hearing difficulty and HA satisfaction. On the other hand, when comparing the EMA ratings to HEARLI-Q ratings for situations where there was some background noise, occurrence was lower, hearing difficulty was higher and HA satisfaction was lower in HEARLI-Q (all \( P < .001 \); see Supplemental Table 2 for detailed statistics). The situations in the “noise” category for HEARLI-Q were combined based on visual inspection of the data showing similar contrasts to EMA for the individual situations.
Figure 2. Mean and 95% confidence intervals of occurrence, hearing difficulty and HA satisfaction ratings in HEARLI-Q and EMA for “Speech Communication – One Person” and “Speech-Communication – Several People” CoSS task categories. The X axes show the questionnaire type: HEARLI-Q (quiet) indicates speech communication in quiet and HEARLI-Q (noise) is a group of the remaining situations in that category with some background noise. The left Y-axis shows the labels for occurrence ratings and the right Y-axis shows the labels for hearing difficulty and HA satisfaction ratings.

2 – Are difficult listening situations with disturbing background noise captured in EMA?

To assess whether difficult listening situations with disturbing background noise are captured in EMA, EMA ratings for situations where there was no noise were compared to the ones where non-disturbing and disturbing background noise was present (see Figure 3). Hearing difficulty was rated significantly higher in situations where disturbing background noise was present (\(P < .001\) for communication with one and \(P = .03\) for communication with several), and HA satisfaction was lower in these situations (all \(P < .001\)) (see Supplemental Table 3 for detailed statistics). Neither hearing difficulty nor HA satisfaction ratings in situations with disturbing background noise were
significantly different from the ratings in “HEARLI-Q (noise)” presented in Figure 2 (detailed statistics can be seen in Supplemental Table 4). However, although represented in EMA, these difficult and less satisfactory situations with disturbing background noise were reported by only 10 participants in “Speech Communication – One Person” and 12 participants in “Speech Communication – Several People”.

Figure 3. Mean and 95% confidence intervals of occurrence, hearing difficulty and HA satisfaction ratings in EMA according to the indicated level of noisiness for “Speech Communication – One Person” and “Speech Communication – Several People” categories. The left Y-axis shows the labels for occurrence ratings and the right Y-axis shows the labels for hearing difficulty and HA satisfaction ratings. For plotting purposes, an average score for occurrence, hearing difficulty and HA satisfaction for each participant was taken, such that each person contributes equally to the summary statistic. N indicates the number of participants that contributed to the summary statistics.

3 – How do HA satisfaction ratings compare between HEARLI-Q and EMA when only frequently occurring situations are analyzed?
When comparing only those situations where participants indicated the occurrence to be daily or weekly, there was not a significant difference in HA satisfaction ratings between the two questionnaires ($\beta = -0.06$ [95% CI: -0.24, 0.12], $P = .53$, mixed-effects linear regression). Hearing difficulty was, however, still lower in EMA ($\beta = -0.26$ [95% CI: -0.44, -0.07], $P < .01$, mixed-effects linear regression). These effects were mainly attributed to “Speech Communication – One Person”, “Speech Communication – Several People”, and “No Conversation / No Focused Listening”. Detailed statistics for individual CoSS task categories are presented in Supplemental Table 5. In Figure 4, it can be seen that hearing difficulty ratings were well below moderate.

Figure 4. Comparison of hearing difficulty and HA satisfaction ratings within individual CoSS task categories between HEARLI-Q and EMA when only those HEARLI-Q situations where Occurrence was daily / weekly were considered. Means and 95% confidence intervals are plotted. N indicates the number of participants that contributed to the summary statistics.
4 – How do hearing difficulty ratings in HEARLI-Q compare between frequently and infrequently occurring situations?

HEARLI-Q hearing difficulty ratings for daily / weekly situations were significantly lower than HEARLI-Q hearing difficulty ratings when including all the listening situations or when comparing to listening situations that occur more seldom than on a weekly basis (frequent vs all situations: β = 0.24 [95% CI: 0.06, 0.43], \( P = .01 \); frequent vs infrequent situations: β = 0.54 [95% CI: 0.34, 0.74], \( P < .001 \), mixed-effects linear regression). See Figure 5 and Supplemental Table 6 for detailed statistics within individual CoSS task categories.

**Figure 5.** Comparison of HEARLI-Q hearing difficulty ratings between frequent situations (occurring on daily or weekly basis), all situations and infrequent situations (occurring more seldom than on weekly basis) within individual CoSS task categories. Means and 95% confidence intervals are plotted. N indicates the number of participants that contributed to the summary statistics.
Discussion

The results of the current study indicate that the overall HA satisfaction ratings are similar between the retrospective HEARLI-Q and in-situ EMA. On the other hand, when comparing the HA satisfaction ratings within individual CoSS task categories, it is evident that ratings are higher in EMA for in-person conversations.

Although EMA has clear advantages over retrospective questionnaires, it is also known that certain situations might be underrepresented, specifically those pertaining to social interactions and noisy environments\textsuperscript{9,10}. In fact, the distribution of listening environments where participants filled out EMA questionnaires points to easier / more familiar listening situations, such as in home environments while not participating in a conversation or focused listening, followed by 1-1 conversations and media listening. This is in line with auditory reality patterns shown in previous EMA studies\textsuperscript{9,17-19}. Moreover, the EMA data in the current work show that difficult situations with disturbing background noise, while represented, are seen less frequently. From EMA data alone, it is unclear whether those more challenging listening situations occurred less frequently, or participants did not fill out the questionnaire in those moments. However, when comparing the situations that only occur on daily or weekly basis, the HA satisfaction ratings between the two questionnaires were comparable. Additional analysis revealed that hearing difficulty ratings
increased with decreased occurrence in HEARLI-Q. Hence, lower occurrence of difficult listening situations seems to be driving the discrepancy in HA satisfaction ratings between the HEARLI-Q and EMA.

Then, if more difficult listening situations occur less frequently, the likelihood of capturing them in time-limited EMA is lower. And, if it is true that people spend most time in easier / more familiar listening situations, and this is what is reflected in EMA, then EMA alone may well be sufficient to accurately model the dynamics of one’s daily life. It has been shown that when a question pertains to a frequent behavior, participants are less likely to have detailed representation of each event in their memory, but rather these are grouped into one global representation without specific traits related to those events. In this case, EMA has a clear advantage over HEARLI-Q because nuances of those frequent episodes would be detected. However, better understanding of the less frequent and more difficult listening situations could give relevant insights into where improvements in hearing care can be made. This in turn could lead to people spending time in situations that they otherwise avoid.

Two previous studies compared retrospective questionnaire responses about HA satisfaction and benefit to EMA. In a study investigating real-life benefit from “noise management” processing, Andersson et al. compared SSQ-12 ratings with EMA. The authors found that EMA data were able to provide insights into more specific listening environments where participants experienced benefit. Wu et al., on the other hand, compared retrospective GHABP scores to those captured by EMA for two HAs. They found that GHABP-EMA was significantly different between the two HAs on all the subscales, in comparison to the retrospective questionnaire where only the satisfaction subscale showed a significant difference. Authors of both studies discuss their findings in light of EMA being more sensitive to detect differences in HA performance. We have shown that both HEARLI-Q and EMA are sensitive to detect effect of the applied intervention on HA
satisfaction\textsuperscript{16} and as such, we do not consider one method to be superior to the other in this aspect, but rather want to discuss their potential of complementing each other. For example, “Focused Listening – Live Sounds” was rated by 13 participants in EMA, whereas only one participant in HEARLI-Q indicated such situations to occur on daily or weekly basis. This is potentially a consequence of the trial period when the country was just starting to open again during COVID-19 and the participants presumably started to attend more live events, which is reflected in EMA. When rating in HEARLI-Q, on the other hand, they reflected on the past year which did not have many such events. Similarly, all the 21 participants indicated that they experience “Speech Communication – Over Device” on daily / weekly basis in HEARLI-Q, but this situation was reported by only nine participants in EMA. This is a situation that typically does not last very long and hence is less likely to be captured by a 2-hour sampling scheme or even if the prompt does align with the event, it may be considered inappropriate to fill out an EMA questionnaire. As such, it is not surprising that speech communication over device is underrepresented in EMA. In the two examples of live sounds and communication over device, we captured complementary information about participants’ everyday lives that we would not have with either of the two questionnaires alone.

It is promising that HEARLI-Q and EMA paint a similar picture: HA satisfaction ratings are highly correlated between the two questionnaires and hearing difficulty is the key predictor of HA satisfaction in both questionnaires. It is also noteworthy that the two questionnaires offer balancing information. When rating HA satisfaction in HEARLI-Q, two participants indicated that they do not wear their HAs when “Passive listening – Hearing sounds of nature.” In EMA, conversely, there were many more situations where participants indicated they do not wear their HAs. In this aspect, EMA offers a more refined insight into day-to-day dynamics and how people use their devices. While it is likely that people do not always take their HAs off in those situations, with EMA, we
can see how often people tend to take their HAs off and when. It is also true that participants are likely to give a more accurate rating of the current situation in EMA, as they are basing it on their experience right-now and not relying on their memory. In contrast, when filling out the HEARLI-Q, one gets a description of the situation to assess, but still must remember a real-life situation which matches the questionnaire item. That process might overemphasize negative experiences as those are more memorable and more salient than positive experiences when doing retrospective reports. For example, we observed a borderline significant contrast in HA satisfaction ratings for “No Conversation / No Focused Listening” category, where there was a tendency of HA satisfaction to be higher in EMA. These types of situations without conversation / focused listening could include vacuum cleaning, washing dishes, operating machinery, etc. In other words, noisy situations where one would need to stop the activity to answer an EMA questionnaire. As such, these situations might be underrepresented. On the other hand, when filling out the HEARLI-Q, the participant is likely to think back to situations that stand out (e.g., of higher hearing difficulty). That is, we potentially have a slight overestimation of hearing difficulty in HEARLI-Q and a slight underrepresentation of hearing difficulty in EMA, with the truth lying somewhere in between.

It may not always be possible to administer both in-situ and retrospective questionnaires to get insights from both angles. One way to overcome the limitation of EMA not capturing infrequently occurring or especially difficult situations can be overcome by asking participants to seek out and report in those types of situations that are relevant for the research questions, as done in Lelic et al. or ask for EMA ratings based on a short retrospective period. HEARLI-Q can also be administered more frequently in a longitudinal fashion, such that participants do not need to think too far back, and this to an extent can reduce the amount of guessing and estimation related to long reference periods.
While providing meaningful insights into how retrospective and in-situ satisfaction ratings compare, it should be noted that the results here most likely depend on the EMA sampling scheme. In the current study, we allowed participants to answer the EMA questionnaire anytime between two consecutive prompts and this potentially allowed for more selection bias than studies that may keep the prompt open for e.g., up to 15 minutes. That is, some situations may appear to “not occur” because they were inconvenient to report in the moment of the prompt and when the participant was able to answer, they were in a completely different situation. The sampling scheme employed in this study is also a likely reason for the high compliance rate. Further, the results of this study are based on experienced HA users. Although it has been previously shown that HEARLI-Q and aggregated EMA responses are stable over time\textsuperscript{3, 25} in this population, expectedly HA satisfaction ratings may vary with time in new HA users as they get used to their devices\textsuperscript{26}. Thus, we may see a different contrast in satisfaction ratings within and between the two questionnaires depending on the time of administration, relative to HA fitting, when considering new HA users.

### Conclusion

The overall satisfaction ratings are similar between HEARLI-Q and EMA. The satisfaction ratings for 1-1 and group conversations are, on the other hand, higher in EMA. Lower occurrence of difficult listening situations seems to be the key driver of discrepancies in HA satisfaction ratings between the two questionnaires. The advantage of EMA is that it provides insight into an individual’s day-to-day life and is less prone to memory bias than HEARLI-Q. HEARLI-Q, on the other hand, can capture those difficult situations which occur infrequently or are inconvenient to report in the moment. The high correlation between the two questionnaires is promising because it may not always be feasible to administer both types of questionnaires. In this case, HEARLI-Q can
provide a reasonable assessment of HA satisfaction. Otherwise, when possible, administering HEARLI-Q and EMA in combination has the potential to give a complementary and more holistic view of HA satisfaction.

References

<table>
<thead>
<tr>
<th>Item</th>
<th>Instruction / Question</th>
<th>Response Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Choose the category that best fits the place where you are now:</td>
<td>1. Inside (home environment)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Inside (public place)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Outside</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Transport (car, train, etc.)</td>
</tr>
<tr>
<td>Listening Task</td>
<td>Choose the category that best fits the situation you are in:</td>
<td>1. Conversation with one person</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Conversation with several people</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Conversation over a phone or another technical device</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Focused listening (live)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Focused listening (media)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Situation without a conversation or focused listening</td>
</tr>
<tr>
<td>Noisiness</td>
<td>Is there background noise?</td>
<td>1. No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Yes - but not disturbing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Yes – it is disturbing</td>
</tr>
<tr>
<td>Occurrence</td>
<td>How often are you in this situation?</td>
<td>1. A couple of times a year</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Monthly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Weekly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Daily</td>
</tr>
<tr>
<td>Importance</td>
<td>How important is it for you to hear well in this situation?</td>
<td>1. Not important</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Slightly important</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Moderately important</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Very important</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Extremely important</td>
</tr>
<tr>
<td>Hearing Difficulty</td>
<td>How difficult is it for you to hear in this situation?</td>
<td>1. Not difficult</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Slightly difficult</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Moderately difficult</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Very difficult</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Extremely difficult</td>
</tr>
<tr>
<td>Hearing Aid Satisfaction</td>
<td>How satisfied are you with your hearing aids in this situation?</td>
<td>1. I am not wearing my hearing aids</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Not satisfied</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Slightly satisfied</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Moderately satisfied</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Very satisfied</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Extremely satisfied</td>
</tr>
</tbody>
</table>

Table 1. EMA questions and response alternatives. The questions and response alternatives are translated from Danish.
Table 2. Mixed-effects linear regression results for HEARLI-Q vs EMA ratings in “Speech Communication – One Person” and “Speech Communication – Several People” categories. HEARLI-Q ratings for quiet and noisy environments were compared to EMA ratings. Random intercept: participant. EMA responses were averaged within each participant for the analysis. Significant $P$-values are shown in bold.
Table 3. Mixed-effects linear regression results showing the effect of noisiness on occurrence, hearing difficulty and HA satisfaction outcomes in EMA. Random intercept: participant. Significant *P*-values are shown in bold.
<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Coefficient (95% CI)</th>
<th>P</th>
<th>Coefficient (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hearing Difficulty</td>
<td>(reference: HEARLI-Q (noise))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMA no noise</td>
<td>-1.13 (-1.48, -0.79)</td>
<td>&lt;.001</td>
<td>-0.65 (-1.00, -0.30)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>EMA noise - not disturbing</td>
<td>-1.09 (-1.42, -0.76)</td>
<td>&lt;.001</td>
<td>-1.14 (-1.48, -0.80)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>EMA noise - disturbing</td>
<td>-0.34 (-0.76, 0.07)</td>
<td>.10</td>
<td>-0.37 (-0.76, 0.02)</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>HA Satisfaction</td>
<td>(reference: HEARLI-Q (noise))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMA no noise</td>
<td>0.80 (0.50, 1.09)</td>
<td>&lt;.001</td>
<td>0.75 (0.35, 1.14)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>EMA noise - not disturbing</td>
<td>0.64 (0.35, 0.93)</td>
<td>&lt;.001</td>
<td>0.74 (0.37, 1.11)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>EMA noise - disturbing</td>
<td>0.26 (-0.11, 0.63)</td>
<td>.17</td>
<td>-0.17 (-0.59, 0.26)</td>
<td>.44</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Mixed-effects linear regression results comparing hearing difficulty and HA satisfaction ratings between EMA at different noisiness levels and the non-quiet HEARLI-Q situations (termed HEARLI-Q (noise) in Figure 2 of the article). Significant differences are indicated in bold.
Table 5. Mixed-effects linear regression results comparing hearing difficulty and hearing aid satisfaction between HEARLI-Q and EMA for individual CoSS categories for situations that occur up to weekly basis. Independent variable: questionnaire type (HEARLI-Q or EMA), random intercept: participant. The within person average hearing difficulty score and hearing aid satisfaction score were the dependent variables. No comparison was done for “Focused Listening – Live Sounds” because data from only one participant were available in HEARLI-Q. Note that hearing difficulty outcome for “Speech Communication – Through Device” shows the opposite trend from other categories or what Figure 4 in the article shows – this is because the analysis is based on only nine individuals who contributed with data to this category. Significant differences are indicated in bold.

<table>
<thead>
<tr>
<th>Category</th>
<th>Hearing Difficulty</th>
<th>HA Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>Speech Communication – One person</td>
<td>-0.41 (-0.73, -0.09)</td>
<td>.01</td>
</tr>
<tr>
<td>Speech Communication – Several People</td>
<td>-0.24 (-0.45, -0.03)</td>
<td>.03</td>
</tr>
<tr>
<td>Speech Communication – Through Device</td>
<td>0.46 (0.08, 0.85)</td>
<td>.02</td>
</tr>
<tr>
<td>Focused Listening – Live Sounds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focused Listening – Media</td>
<td>-0.01 (-0.29, 0.26)</td>
<td>.91</td>
</tr>
<tr>
<td>No Conversation / Focused Listening</td>
<td>-0.67 (-0.99, -0.36)</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>
Dependent variable: Hearing Difficulty
Reference: frequent situations

<table>
<thead>
<tr>
<th>Situation</th>
<th>All situations</th>
<th></th>
<th>Infrequent situations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (95% CI)</td>
<td>P</td>
<td>Coefficient (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>Speech Communication – One person</td>
<td>0.35 (0.08, 0.61)</td>
<td>.01</td>
<td>0.68 (0.41, 0.96)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Speech Communication – Several People</td>
<td>0.32 (0.05, 0.59)</td>
<td>.02</td>
<td>0.45 (0.17, 0.73)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Speech Communication – Through Device</td>
<td>0.31 (-0.05, 0.66)</td>
<td>.09</td>
<td>0.82 (0.40, 1.23)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Focused Listening – Live Sounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focused Listening – Media</td>
<td>0.00 (-0.14, 0.13)</td>
<td>.96</td>
<td>-0.04 (-0.27, 0.19)</td>
<td>.75</td>
</tr>
<tr>
<td>No Conversation / Focused Listening</td>
<td>0.23 (-0.06, 0.52)</td>
<td>.11</td>
<td>0.69 (0.36, 1.03)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Table 6. Mixed-effects linear regression results comparing hearing difficulty measured on HEARLI-Q when only frequently occurring situations are considered to hearing difficulty ratings when all the situations / infrequently occurring situations are considered. Here, frequent situations are those occurring daily or weekly and infrequent situations are the ones occurring more seldomly (rated monthly or couple of times a year). Infrequently occurring situations are generally rated as more difficult than the frequent situations, except for the “Focused Listening – Media” situation, but here, it should also be noted that only 6 participants rated a sub-category as occurring less frequently than weekly basis. No comparison was done for “Focused Listening – Live Sounds” because data from only one participant were available in HEARLI-Q. Significant differences are indicated in bold.
Hearing Aid Satisfaction by CoSS Task

Speech Communication - One Person
- Extremely
- Very
- Moderately
- A little
- Not at all

N=21

HEARLI-Q
EMA

Speech Communication - Several People
- Extremely
- Very
- Moderately
- A little
- Not at all

N=21

HEARLI-Q
EMA

Speech Communication - Through Device
- Extremely
- Very
- Moderately
- A little
- Not at all

N=20

HEARLI-Q
EMA

Focused Listening - Live Sounds
- Extremely
- Very
- Moderately
- A little
- Not at all

N=19

HEARLI-Q
EMA

Focused Listening - Through Media
- Extremely
- Very
- Moderately
- A little
- Not at all

N=15

HEARLI-Q
EMA

No Conversation / No Focused Listening
- Extremely
- Very
- Moderately
- A little
- Not at all

N=21

HEARLI-Q
EMA

Speech Communication - One Person
- Daily
- Weekly
- Monthly
- Yearly

HEARLI-Q (well)
EMA

Speech Communication - Several People
- Extremely
- Very
- Moderately
- A little

HEARLI-Q (well)
EMA

HEARLI-Q (noise)
EMA

Occurrence
Difficulty
Satisfaction
Situations Occurring Daily or Weekly

- EMA Ratings
  - Speech Communication - One Person
  - N=19
  - Speech Communication - Several People
  - N=15
  - Speech Communication - Through Device
  - N=21

- Situations Occurring Daily or Weekly
  - Speech Communication - One Person
    - HEARLI-Q
    - EMA
    - N=21
  - Speech Communication - Several People
    - HEARLI-Q
    - EMA
    - N=19
  - Speech Communication - Through Device
    - HEARLI-Q
    - EMA
    - N=21

- Focused Listening - Live Sounds
  - HEARLI-Q
  - EMA
  - N=1
  - N=13

- Focused Listening - Media
  - HEARLI-Q
  - EMA
  - N=21
  - N=18

- No Conversation / Focused Listening
  - HEARLI-Q
  - EMA
  - N=21
  - N=21

- Occurrence Difficulty Satisfaction

- HEARLI-Q EMA
- HEARLI-Q EMA
- HEARLI-Q EMA

Legend:
- Green square: Difficulty
- Orange circle: Satisfaction
HEARLI-Q Hearing Difficulty Ratings – Comparison Based on Occurrence

Speech Communication - One Person
- Extremely: N=19
- Very: N=21
- Moderately: N=21
- A little: not at all
- Frequent: All
- Infrequent

Speech Communication - Several People
- Extremely: N=19
- Very: N=21
- Moderately: N=20
- A little: not at all
- Frequent: All
- Infrequent

Speech Communication - Through Device
- Extremely: N=21
- Very: N=21
- Moderately: N=21
- A little: not at all
- Frequent: All
- Infrequent

Focused Listening - Live Sounds
- Extremely: N=21
- Very: N=21
- Moderately: N=1
- A little: not at all
- Frequent: All
- Infrequent

Focused Listening - Media
- Extremely: N=21
- Very: N=21
- Moderately: N=6
- A little: not at all
- Frequent: All
- Infrequent

No Conversation / Focused Listening
- Extremely: N=14
- Very: N=21
- Moderately: N=21
- A little: not at all
- Frequent: All
- Infrequent

Response Time to EMA Prompts

Density

Time of Prompt

Response Time (minutes)