

Individual Hearing Aid Benefit in Real Life Evaluated Using Ecological Momentary Assessment

Trends in Hearing
Volume 25: 1–18
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216521990288
journals.sagepub.com/home/tia



Petra von Gablenz¹ , Ulrik Kowalk¹, Jörg Bitzer¹, Markus Meis² and Inga Holube¹

Abstract

Ecological momentary assessment (EMA) was used in 24 adults with mild-to-moderate hearing loss who were seeking first hearing-aid (HA) fitting or HA renewal. At two stages in the aural rehabilitation process, just before HA fitting and after an average 3-month HA adjustment period, the participants used a smartphone-based EMA system for 3 to 4 days. A questionnaire app allowed for the description of the environmental context as well as assessments of various hearing-related dimensions and of well-being. In total, 2,042 surveys were collected. The main objectives of the analysis were threefold: First, describing the “auditory reality” of future and experienced HA users; second, examining the effects of HA fitting for individual participants, as well as for the subgroup of first-time HA-users; and third, reviewing whether the EMA data collected in the unaided condition predicted who ultimately decided for or against permanent HA use. The participants reported hearing-related disabilities across the full range of daily listening tasks, but communication events took the largest share. The effect of the HA intervention was small in experienced HA users. Generally, much larger changes and larger interindividual differences were observed in first-time compared with experienced HA users in all hearing-related dimensions. Changes were not correlated with hearing loss or with the duration of the HA adjustment period. EMA data collected in the unaided condition did not predict the cancelation of HA fitting. The study showed that EMA is feasible in a general population of HA candidates for establishing individual and multidimensional profiles of real-life hearing experiences.

Keywords

hearing-aid uptake, ecological momentary assessment, self-reported hearing, hearing impairment, nonoverlap of all pairs

Received 9 March 2020; Revised 23 December 2020; accepted 5 January 2021

Introduction

The overall goal of any hearing-aid (HA) fitting is to enhance hearing experiences and the associated hearing-related quality of life. Standardized laboratory measurements cannot represent the “auditory reality” of an individual, which Noble (2008) defined as “a product of the person’s engagement with the world plus what the world provides” (p. 118). As laboratory tests only give limited information about the extent to which rehabilitative goals have been achieved in natural environments, questionnaires are conventionally used for capturing the individual’s perspective. Against this background, ecological momentary assessment (EMA) has been found to be a well-adapted method for surveying the broad spectrum of hearing experiences under real-life conditions

with a maximum degree of individualization. Stone et al. (2007) defined EMA as “real-time collection of data about momentary states, collected in the natural environment, with multiple repeated assessments over time” (p. 3). The technical developments of recent years have promoted the use of this method and led to EMA being understood as an electronically based

¹Institute of Hearing Technology and Audiology, Jade University of Applied Sciences and Cluster of Excellence “Hearing4all,” Oldenburg, Germany
²Hörzentrum Oldenburg GmbH, Oldenburg, Germany

Corresponding author:

Petra von Gablenz, Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Ofener Straße 16/19, 26121 Oldenburg, Germany.

Email: petra.vongablenz@jade-hs.de



collection of real-world and real-time outcome data. After Galvez et al. (2012) had shown the applicability of EMA in hearing research, a number of EMA projects using high-tech devices were conducted. To name but a few, Hasan et al. (2014) described the auditory reality of HA users; Wu et al. (2018) examined the performance of HA features, and Probst et al. (2017) studied time-of-day changes of tinnitus loudness and tinnitus distress. Most recently, Burke and Naylor (2020) studied fatigue in hearing-impaired adults using EMA. A detailed overview of the use of EMA in hearing research is given in the review by Holube et al. (2020).

This EMA study involved 24 adults who came to HA acousticians. They used a smartphone-based EMA device for several days at two milestones of the aural rehabilitation process, before HA fitting and after an average 3-month HA adjustment period. On the basis of about 2,000 surveys, this article describes the auditory reality of adults on the cusp of HA uptake and examines the change in real-life hearing experiences following HA fitting.

Two previous studies are of particular interest in this context. Over three data-collection periods, Timmer et al. (2018) used EMA to investigate the effect of HA amplification in ten participants that had mild-to-moderate hearing loss. More specifically, EMA data were collected for 1 week without HA (baseline phase), 2 weeks with HA (intervention phase), and again 1 week without HA (withdrawal phase). The participants had no previous amplification experience, but were familiar with the research method, having already volunteered for an earlier EMA study. Timmer et al. (2018) found considerable interindividual differences in reported speech intelligibility and reported listening effort, with listening effort being possibly a better indicator for hearing difficulties in challenging listening situations than speech intelligibility.

The second research work of particular interest for this study addressed conceptual topics of real-world listening situations. Wolters et al. (2016) established the three-level framework Common Sound Scenarios (CoSS) to categorize auditory reality. Specifically, CoSS allows for differentiating everyday hearing experiences according to listening intentions and tasks. It was applied by Jensen et al. (2019) for evaluating auditory reality and HA use and by Smeds et al. (2019) for the development of a more realistic laboratory test procedure.

As sample sizes are still small in EMA hearing studies, replicative and collaborative approaches are recommended. We therefore converted our EMA app categories to the CoSS classification to facilitate the comparison of results.

In concrete terms, the following research questions are addressed:

1. Which listening intentions, tasks, and hearing-related disabilities characterize the auditory reality of adults seeking HA fitting?
2. How strong are the effects of HA fitting on self-rated hearing abilities and related dimensions on the individual level and in a heterogeneous group of first-time HA users?
3. Does the EMA data collected in the unaided condition predict ultimate decisions for or against permanent HA use?

Methods

Study Design

This interventional field study was conducted from 2018 to 2019 in Oldenburg, Germany, as a part of the project “Individual Hearing Aid Benefit in Real Life” (IHAB-RL). Using a single-subject design, the study took place in two phases (AB design). Adults who were medically advised for HA fitting performed EMA both before being fitted with HA (phase A, hereinafter: EMA1) and after an individual HA adjustment period (phase B, hereinafter: EMA2). The participants had four visits to the university’s facilities. The study flow is shown in Figure 1. Participants were recruited by local HA acousticians during the counseling interview carried out before HA fitting. Thus, they were *real-world clients* who were seeking HA fitting on the advice of an ENT specialist. The only inclusion criterion was wearing glasses all day because the EMA device required attaching miniature microphones on the frame of the glasses (see “EMA Hardware” section). If a client was willing to participate in the study, an appointment at the university was made, usually for the following day, so as not to unduly postpone HA fitting. The participants received the EMA system on their first visit and returned it after 3 to 4 days on the second visit. Subsequently, HA were fitted by the local acousticians. The study site did not intervene in the process of HA selection and fitting at any time. When study participants had decided on the purchase and considered the HA fitting to be completed, they contacted us to make an appointment for the third visit and thus the postfitting EMA2. The HA adjustment period, a term adapted from Ricketts et al. (2019, p. 730), was calculated as the time elapsed from the day of the first HA fitting to the day of the third visit at the university.

Study participation for the four visits to the university was remunerated on an hourly basis (12 Euros per hour) and in a lump sum for EMA periods (20 Euros per day) after the last visit. In addition, a concise summary of the individual EMA results was given to the participants when they returned the EMA system after both EMA1 and EMA2. This feedback summary included easily

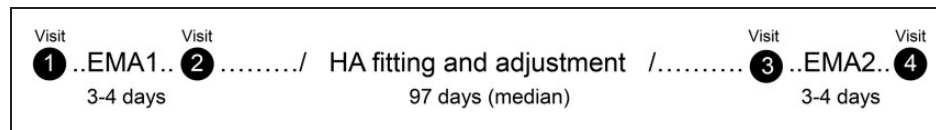


Figure 1. Design of IHAB Study. The participants had two EMA phases and four visits at the university. HA fitting was conducted by local acousticians. The duration of HA adjustment was individually different. EMA = ecological momentary assessment; HA = hearing aid.

understandable graphics, with basic information on the number of surveys, broken down into situations and activities, and the corresponding assessment mean score for speech understanding and listening effort. Following EMA2, this feedback summary compared data collected in the unaided or formerly aided condition with the aided or renewed condition. The research design and procedures passed examination by the ethics commission of the Carl von Ossietzky University in Oldenburg. Written informed consent was obtained from all participants.

Additional Measures. The study protocol also included speech audiometry in quiet and in noise, diverse questionnaires, a standardized interview on sociodemographics and health-related characteristics, an assessment of communication behavior, and—for participants who completed EMA1 and EMA2—cognitive screening and probe-microphone measurements. The pure-tone audiogram was usually provided by the HA acousticians. The examinations took place at the university's facilities. As they were of interest for the above-stated research questions, global results were reported for pure-tone thresholds, the duration of the HA adjustment period, sociodemographic details, the Hearing Handicap Inventory for the Elderly (HHIE, Ventry & Weinstein, 1982), and the Montreal Cognitive Assessment (MoCA, Nasreddine et al., 2005). The participants received the HHIE questionnaire from the HA acousticians and filled them out at home. The MoCA test was performed by trained experimenters during the fourth visit to assess whether cognitive conditions might have impacted compliance with the instructions and thus data reliability.

Study Population. In total, 24 adults (9 females) with mild-to-moderate hearing loss participated in the study. Among these, 20 adults were first-time HA users and 4 adults were experienced HA users. The sample was diverse with regard to hearing loss, hearing-related disability, and in social terms, including low, intermediate, and high professional skill levels, and different employment status (14 retired, 9 working, 1 without employment). The 24 adults' individual characteristics are shown in Table 1.

Sixteen adults decided for HA acquisition. Eight participants left the study after conducting EMA1 (seven canceled the HA fitting, one left the study for personal reasons); these are hereinafter referred to as nonreturner participants (no. 17–24 in Table 1). Sixteen participants conducted EMA1 and EMA2, hereinafter referred to as returner participants (no. 1–16 in Table 1). An experienced clinical audiologist compared the probe-microphone measurements at an input level of 65 dB SPL for speech with the NAL-NL2 (Keidser et al., 2011) targets of the returner participants. The probe-microphone measurements averaged across the frequency range from 0.5 to 4 kHz, both ears, and all participants were 3 dB below target. For each participant, the difference was less than 10 dB. The difference was not related to the reported HA benefit and therefore disregarded for the present analysis. The total MoCA score in returner participants ranged from 22 to 30. Five participants achieved a total score below the cutoff of 26 recommended for this brief cognitive screening tool (Nasreddine et al., 2005). As the Cochrane Review of Davis et al. (2015) provided evidence for a high false-positive rate (40%) when using 26 as the cut-off score, we decided to keep the data of these comparatively low-performing participants in the sample.

EMA Survey

EMA Hardware. For this field study, the EMA system olMEGA was used, which allows for managing digital surveys and continuous collection of objective acoustical feature data (Groenewold et al., 2018; Kowalk et al., 2017). Objective acoustical data were not used for the present analysis. olMEGA consists of an android smartphone (LG Nexus 5 in this study), a Bluetooth transmitter unit, and two cable-connected microphones that were attached to the frames of the participants' eyeglasses. olMEGA was developed at the Jade University of Applied Sciences, Oldenburg. Construction manuals, software implementation and Supplemental Material for olMEGA are available at <https://github.com/IHAB-RL> (Open Source License for all parts, including hardware design).

Survey Items. Surveys were collected using the digital questionnaire app that is part of the olMEGA system and was installed on the smartphone. The complete list

Table 1. Characteristics of Participants and EMA Data Collection.

Subject number	Returner participants																Nonreturner participants (only EMA1)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1-16	17	18	19	20	21	22	23	24	17-24						
Age (years) ^a	50	75	60	65	55	60	75	70	50	75	60	70	60	70	65	70	67	60	60	75	75	60	70	70	70	69						
ISCO skill level ^b	2	3	3	4	4	3	3	4	1	3	4	4	2	2	4	4	4	4	3	2	4	4	3	2	1							
HA																																
First-time user	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes							
Acquisition decided	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	Yes						
Adjustment (days)	86	22	132	77	27	40	99	113	95	162	88	110	136	160	56	139	97															
PTA ^c (dB HL)																																
Better ear	33	48	29	39	18	19	36	30	8	45	25	31	43	29	35	45	32	33	21	23	30	26	44	26	51	28						
Worse ear	40	53	36	46	23	24	46	38	33	49	33	31	50	34	38	53	38	34	26	26	36	46	53	34	55	35						
HHIE ^d (total score)	32	53 ^e	16	12	20	20 ^e	20	12	38	8	34	32	36	30	28	40	29	76	22	2	40	36	20	26	64	31						
EMA1																																
Usage time (hours)	27	31	39	33	47	47	29	30	36	23	47	49	43	48	17	53	37	45	18	18	35	17	30	46	41	32						
Queries (counts)	50	40	57	27	66	68	53	43	61	27	66	81	61	75	36	74	59	67	13	34	35	9	49	66	64	42						
Compliance _A (%)	83	44	62	30	64	55	62	48	74	11 ^f	47	60	48	44	74	48	51	38	28	65	47	12	49	45	55	46						
Compliance _{A+M} (%)	92	65	74	41	70	72	91	71	85	58 ^f	70	82	71	79	107	69	72	75	37	96	50	26	82	72	79	73						
EMA2																																
Usage time (hours)	24	32	33	30	33	30	27	38	45	48	36	33	44	32	38	62	33															
Queries (counts)	45	48	39	21	42	49	54	58	66	52	55	57	44	46	63	81	51															
Compliance _A (%)	87	59	48	27	58	65	49	49	67	51	56	63	36	59	73	24	57															
Compliance _{A+M} (%)	93	75	60	35	64	82	99	77	74	55	77	86	50	72	83	65	75															

Note. md = median; EMA = ecological momentary assessment; HA = hearing aid; dB HL = decibel hearing level; HHIE = Hearing Handicap Inventory for the Elderly; ISCO = International Standard Classification of Occupations; PTA = pure-tone average.

^aAge was rounded to full and half decades.

^bProfessional skill level based on the current or former job according to the International Standard Classification of Occupations (2012).

^cPure-tone average at 0.5, 1, 2, and 4 kHz.

^dHearing Handicap Inventory for the Elderly (Ventry & Weinstein, 1982).

^eMissing answer for either one or two items replaced by the rounded individual mean score derived in the respective subscale.

^fDifferentiation between prompted and self-initiated surveys probably biased due to connectivity problems of oMEGA.

Table 2. Assessment Dimensions Used in the IHAB App.

No.	Dimension	Extreme categories
1	<u>Speech understanding</u>	Nothing at all \longleftrightarrow perfect
2	<u>Listening effort</u>	No effort \longleftrightarrow extreme effort
3	<u>Localization of sounds</u>	Not at all \longleftrightarrow perfect
4	<u>Loudness perception</u>	Too soft \longleftrightarrow too loud
5	<u>Pleasantness of sounds</u>	Very unpleasant \longleftrightarrow very pleasant
6	Perception of <u>disability</u>	Not disabled at all \longleftrightarrow extremely disabled
7	<u>Involvement</u> in group communication	Definitively left out \longleftrightarrow definitively right in the middle (5 cat.)
8	<u>Difficulty by conversation partners</u>	Not difficult at all \longleftrightarrow extremely difficult
9	<u>Importance</u> of hearing well	Completely unimportant \longleftrightarrow very important
10	<u>Mood</u>	Very unhappy \longleftrightarrow very happy (5 smiley icons)

Note. The extreme categories refer to 7-point scales if not otherwise stated. Note that the direction of the scales has been reversed for some items in this article to improve the readability. The underlined words are hereinafter used as short forms for the respective items.

of items with the corresponding response options, in the order in which they were displayed in the olMEGA app, is provided in the supplement to this article (Supplementary Material). The questionnaire has an adaptive flow, that is, preceding answers determine the display of subsequent questions. The surveys contained a maximum of 13 items. Participants were initially asked to indicate how much time had elapsed as the situation occurred that they were going to assess (now, about 2–3, 5, 10, 15, 20, or 30 min ago) and to determine a general situation type (at home, in transit, in society, at work, other/manual entry option) and related activities. Subsequently, participants were asked to indicate sound sources and, if any sound source were reported, to specify the target source. Situations-specific selection lists were provided for choosing up to eight sound sources and one single target source.

Table 2 lists the assessment dimensions. Prior environmental and intentional descriptions determined eligible items, except the item on perceived disability, which was included at the end of each survey. If not otherwise stated, 7-point categorical scales were used.

Sampling Schedule. Every participant used the EMA system for at least 3 days. Depending of the day of the week and the individual agendas, the system was returned from the third to sixth day after handover. The data collection did not follow a fixed daily timetable. The participants were instructed to use the system all day long and to charge the system overnight. Data collection was deactivated during the charging process because the Bluetooth transmitter and the smartphone were disconnected. Otherwise, as soon as the smartphone and the Bluetooth transmitter were connected, data collection was automatically activated. The participants were instructed to take a survey when the acoustic situation changed markedly and when they were prompted by an alarm. An alarm was activated

approximately every 30 min. This time interval was randomly shortened or lengthened up to a maximum of 5 min. The alarm consisted of a 30 s vibration signal. In addition, the smartphone screen lit up and the text “EMA questionnaire” was displayed in large letters. No acoustical signal was emitted. As the data collection was closely timed, the prompt was kept unobtrusive and discreet, to avoid annoying the participants and bystanders. The screen remained in the request mode until a survey was conducted. If olMEGA was in the request mode, each survey was logged as a prompted response. Otherwise, the survey was logged as a self-initiated response. Thus, surveys taken less than about 30 min after the previous survey were interpreted as self-initiated responses. Due to this data logging, latency to a prompt was not calculable in this study.

The participants were carefully instructed in the use of the system before EMA1 and again before EMA2. The instruction at the first visit took about 1.5 h and included explanations of the survey items, handling, and repeated hands-on training. The participants received an illustrated manual in which they could find the most important information. In addition, the research assistants were available via mobile phone at all times.

Analysis

Categories and Descriptives. The auditory reality sampled in the EMA data is described using the CoSS intention and task categories displayed in Table 3.

The participants’ detailed descriptions given in the adaptive questionnaire, particularly the details on target sound sources, were used to assign the surveys to first level (listening intention) and second level (listening task) CoSS categories. Categorizing the data down to the third level (scenarios) led to undue data fragmentation and was therefore skipped for this analysis. The category “focused listening” conceptually contains

Table 3. Partial List of the Common Sound Scenarios Proposed by Wolters et al. (2016).

1st level	Intention	Speech communication		Focused listening		Nonspecific		
		2 people	> 2 people	Through device	Live sounds	Through media device	Monitoring surroundings	Passive listening
	Task	Two people having a shared conversation	Several people having a shared conversation	≥ 2 people having a shared conversation through a communication device	Focused listening to sound without being able to control the sound source	Focused listening to sound while being able to control the sound source	Conscious or unconscious screening of sound of relevance to current activity	Unconscious perception of environmental sounds, without relevance to current activity

Note. Scenario categories (third level) are not shown.

speech and nonspeech events. Therefore, estimates for speech understanding and listening effort in this category were limited to surveys that referred to speech-listening events (target signal “speech”). The listening task “monitoring surroundings” refers to events during which sound sources, but no target sources, were specified; for example, mobility activities such as driving a car or biking that require conscious or unconscious screening of sounds. The task “passive listening” includes events for which no sound source (quiet condition) or no target signal relevant to the current activity was specified; for example, when working at the PC at home. In addition, the basic olMEGA categories (home, in transit, society, and work) were used for reporting.

Statistical Analysis. Conventional procedures of descriptive statistics were used to show the distribution of CoSS categories. Spearman correlation coefficients were calculated for pure-tone average (PTA at 0.5, 1, 2, and 4 kHz), for HHIE total scores, the duration of the HA adjustment period, mean EMA scores, the distribution of CoSS intention categories in EMA1 and EMA2, and for evaluating the association of ratings of different dimensions.

The type and the distribution of the EMA data did not meet the assumptions of mean-based statistics. As no consensus has been reached on how to analyze time series of categorical data from study participants who provide greatly varying amounts of data, competing concerns were balanced when comparing ratings from EMA1 and EMA2. To provide conventionally interpretable as well as statistically reliable results, the analysis of EMA data was twofold. On the one hand, ordinal data were treated as continuous, that is, categories were linearly transformed to numerical values ranging either from 1 to 7 or from 1 to 5. On the other hand, the analyses for the main research questions were performed using methods that are considered appropriate for ordinal outcome data.

On the individual level, HA benefit, in terms of the assessment change from EMA1 to EMA2, is reported both as the difference of the mean scores yielded in EMA1 and EMA2 and as nonoverlap of all pairs (NAP). NAP is a nonparametric effect size (ES) measure suggested by Parker and Vannest (2009). They argued that the area under the curve (AUC) from the receiver operating characteristics analysis can be interpreted as “the percentage of data which improve across phases” (Parker et al., 2011). Thus, the assessment scores from individual EMA1 and EMA2 data were submitted to a receiver operating characteristics analysis with a binary outcome variable (0/1) encoding the phases EMA1 and EMA2. The null hypothesis is the chance level (AUC = 0.5), which can be interpreted as a complete overlap of the scores from EMA1 and EMA2, that is,

showing neither an improvement nor a deterioration of the ratings. In the opposite case, if the assessment scores from EMA1 and EMA2 did not overlap at all, that is, if even the worst rating in one EMA is still better than any rating in the other EMA, the AUC is either 0 or 1. If the 95% confidence interval (CI) of AUC included 0.5, NAP was considered to be nonsignificant.

At the group level, the assessment change is only reported for first-time HA users. As the data consist of repeated measurements for each individual participant, mixed models of the following general type were fitted: $EMAScore = \beta * EMAPhase + \gamma * ParticipantID + c$. The indicator variable for EMA1 versus EMA2, denoted $EMAPhase$, was included as a fixed effect. The indicator variable for individual participants, denoted $ParticipantID$, was included as a random effect. Each individual participant was fitted with a random intercept and random slope. More specifically, cumulative link mixed models (CLMM) were fitted. CLMM provide a mixed-modeling framework for categorical responses (Bauer & Sterba, 2011). A cumulative logit link function was used, which allows for a convenient interpretation in terms of proportional odds. If the beta coefficient estimate is positive, it indicates that a higher outcome category is more likely in EMA2 as compared with EMA1. If the beta estimate is negative, the reverse applies. All statistical analyses were performed using IBM SPSS 26.0.

Weighting Procedure. The number of surveys varied widely between the participants. To avoid biasing the descriptive statistics through participants with a high number of surveys, sample weights were calculated as follows: Let $N \in \mathbb{N}^+$ be the total number of subjects. Furthermore, let $i \in \{1, \dots, N\}$ denote an individual participant and $n_i \in \mathbb{N}^+$ be the number of completed surveys of participant i . The sample weight w_i for participant i is then defined by:

$$w_i = \frac{\frac{1}{N} \sum_{j=1}^N n_j}{n_i} \quad (1)$$

Sample weights were only used for reporting the distribution of CoSS categories in the ‘‘Auditory Reality’’ subsection and the related statistics. Figures 3 and 4 show how weighting impacted the results.

Compliance. In EMA studies, it is particularly important to determine how well the participants complied with the protocol. Due to the data-collection procedures in this study, it is difficult to calculate a percentage figure that allows a straightforward comparison to compliance estimates from other studies. As the surveys were not only prompted by closely timed alarms, but also self-initiated,

two compliance parameters were estimated for each study participant and EMA period as follows:

$$Compliance_A = \frac{\text{Count of surveys prompted by timed alarm}}{\text{Count of timed alarms}} \quad (2)$$

$$Compliance_{A+M} = \frac{\text{Total count of surveys (prompted \& self-initiated)}}{\text{Count of timed alarms}} \quad (3)$$

The count of alarms was estimated by dividing the usage time of olMEGA in each EMA period per participant by 30 min. The total usage time was calculated by summing up the usage time of olMEGA for each day in either EMA phase. The daily usage time was defined as the time elapsed from the first survey in the morning to the last survey before sleep.

Reactivity. Reactivity, defined as ‘‘the potential for behavior or experience to be affected by the act of assessing it’’ (Shiffman et al., 2008, p. 20), was examined using EMA1 data for speech understanding, listening effort, and perceived disability. As it was unclear whether such a behavior or experience change would be expected to be gradual or rather be spontaneous, the individual participant’s assessments were graphically evaluated for each CoSS listening intention in their chronological order. Moreover, two other analyses were performed. Check 1: On the ratings of speech understanding, listening effort, and disability separate CLMM of the following general type were fitted: $EMAScore = \beta * Sequence + \gamma * ParticipantID + c$. In each case, the chronological sequence of surveys, denoted $Sequence$, was included as a fixed effect and participants, denoted $ParticipantID$, as a random effect. Each individual participant was fitted with a random intercept and random slope. All CLMM used a logit link function. Check 2: The ratings of disability for the chronologically first three surveys (early) and the last three surveys (late) of EMA1 were separately averaged for each participant’s most frequently specified CoSS intention category. These mean values were regressed to two virtual measurement points (early vs. late). If the estimated beta coefficients were large, and even proved to be significant ($\alpha = .05$, two-sided), this could indicate that on a group level, the assessment criteria might have changed over time.

Results

Data Overview

In total, 2,049 surveys were retrieved from the EMA systems. Figure 2 shows the numbers of surveys collected at each stage of the study. Given that no fixed daily timetable was defined, the daily olMEGA usage time and the total time of data collection varied individually

according to life style. Table 1 lists the individual usage time and the number of surveys separately for each participant in EMA1 and EMA2. Each EMA phase (median) lasted 3 days and 9.2h per day. Overall, the participants took an EMA survey every 40 to 45 min. One survey took 73 s (median).

Compliance and Reactivity. In EMA1, compliance_A and compliance_{A+M} averaged 50% and 73%, respectively, and differed considerably between the participants.

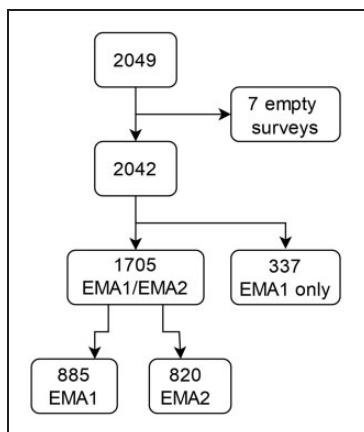


Figure 2. Overview of the Total Numbers of EMA Surveys Collected in IHAB-RL From 24 Study Participants in EMA1 and EMA2. Eight nonreturner participants conducted only EMA1. EMA = ecological momentary assessment.

Detailed results for each participant are given in Table 1. Individual compliance was not related to demographic variables or MoCA scores. Visual inspection and CLMM analyses were used to check for reactivity. No signs for reactivity were detected in individual time-series data on speech understanding, listening effort, and disability. Beta coefficients from CLMM for the change of assessments over time were all below $|0.01|$ (Check 1). Modeling the mean “early” and “late” disability ratings for the individually most-frequent listening intention yielded a comparatively increased beta estimate of 0.3 (95% CI $[-0.2, 0.8]$) but did not reach significance (Check 2). In summary, these results do not suggest that the participants fundamentally changed their decision criteria over the course of EMA1.

Momentary and Delayed Assessment. The core idea of the EMA method is to collect assessments at the time of the experience. For this reason, it is of interest to examine in which proportions and in which situations the assessments were made more or less *at the moment*. In total, 1,166 (57%) of the 2,042 surveys were, strictly speaking, momentary, referring to an assessment delay of 0 min, and 375 (18%) related to situations that occurred within the preceding 5 min. An assessment delay between 5 and 15 min was reported in 163 (8%) surveys and 185 (9%) of the surveys referred to situations more than 15 min, and less than 30 min, back in time. Information was missing for 153 (8%) surveys. To some extent, the assessment delay related to the location and activity. When the participants were “in transit” (e.g., bike, car, bus,

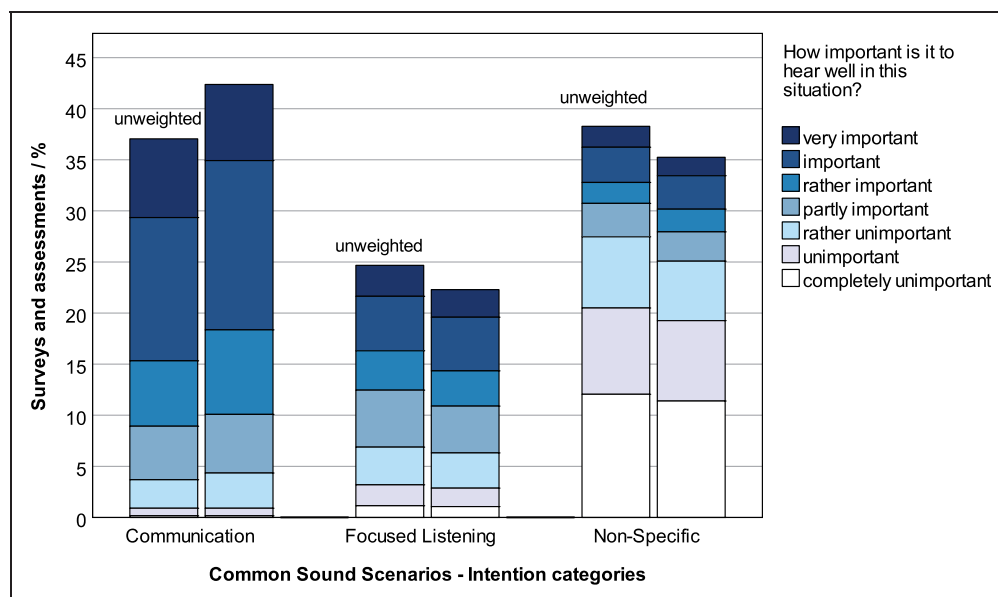


Figure 3. Distribution of CoSS Intention Categories and the Corresponding Assessments on the Dimension “Importance” in EMA1. The left bar of every cluster shows the results derived from the unweighted data, and the right bar shows weighted results (adjusted for the number of surveys given by each participant).

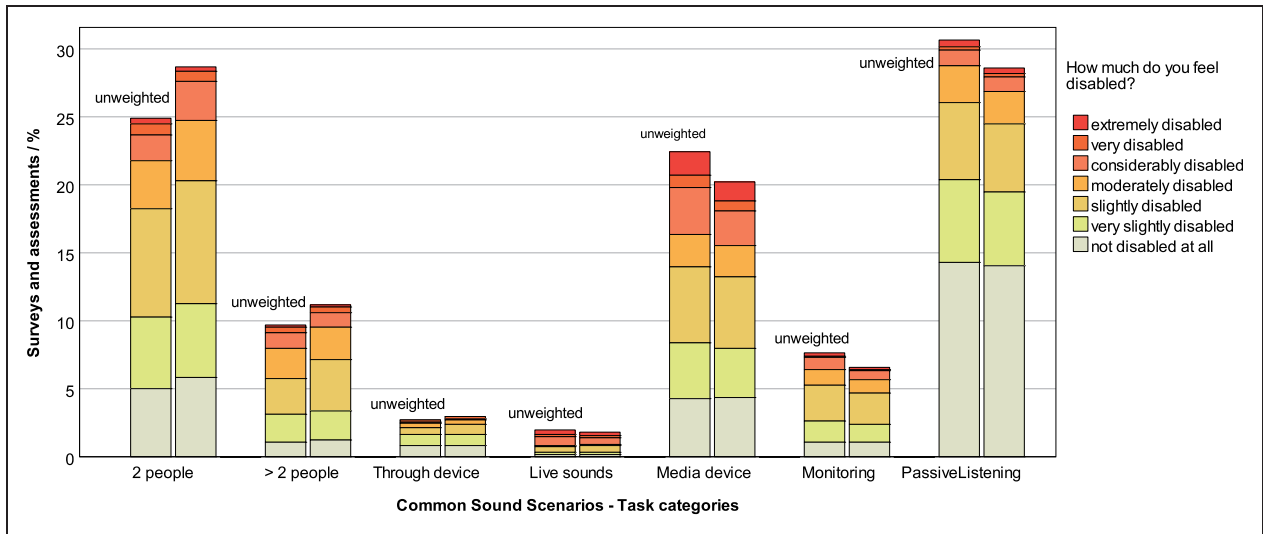


Figure 4. Distribution of CoSS Task Categories and the Corresponding Assessments in the Dimension “Disability” in EMA1. The left bar of each cluster shows the results derived from the unweighted data, and the right bar shows weighted results (adjusted for the number of surveys given by each participant).

walking), 64% of the assessments were delayed, followed by 52% during work time, 44% in societal situations (e.g., shopping, visits, lectures), and 39% when at home.

Auditory Reality

The descriptive statistics in this section are based on the EMA1 data of all 24 participants ($n = 1,222$). When not otherwise stated, sample weights were applied in this subsection.

The distribution of CoSS categories is shown in Figures 3 and 4, together with the corresponding ratings on the situational importance of hearing and disability. Results are shown for the unweighted and weighted data. On the intentional level, most assessments related to speech communication (42%), and the fewest to focused listening (23%). In events assigned to these two intention categories, good hearing was considered almost equally important. On the other hand, good hearing was considered much less important in situations assigned to the nonspecific intention category that accounted for 35% of the assessments. Broken down to the CoSS task level (Figure 4), a similarly high proportion of surveys were attributable to dialog communication situations (28%) and passive listening (29%). The perceived overall disability was generally higher in communication than in passive listening tasks. A minor disability was mostly stated for 50% of the passive listening tasks. A disability of varying degrees was reported for 78% of the conversation events with one partner and for 88% of the communication events with more than one partner. Overall, 21% of the listening tasks were related to media listening for

which similar levels of disability were reported. Monitoring surroundings, communication through a device, and focused listening to live sounds accounted for relatively small proportions of 7%, 3%, and 2% of the total data, respectively.

According to the original IHAB app categories, 59% of the surveys related to situations in private domestic environments, 16% to work environments, 13% to “in transit” situations, and 11% to societal situations and activities. In less than 2% of the surveys, the situation was labeled as “other,” without any further specification.

Overall, ratings on different dimensions were correlated. Moderate-to-strong correlations with $r \geq |.6|$ were observed between speech understanding, listening effort, and disability ($p < .001$). Localization was correlated with $r \geq |.6|$ to speech understanding and listening effort, but not to disability.

HA Benefit

The results presented in this section refer to 1,705 surveys collected by returner participants (Nrs. 1–16) in EMA1 and EMA2. Of these surveys, 933 related to speech-listening events.

To examine whether the participants’ auditory reality was comparable in EMA1 and EMA2, the shares of the CoSS intention categories in the surveys were calculated separately for each returner participant and each EMA phase. The agreement of individual EMA1 and EMA2 data with regard to the distribution of CoSS intention categories is displayed as a scatter plot in Figure 5. Spearman correlation on the percentage shares resulted in $r = .52$ ($p < .001$).

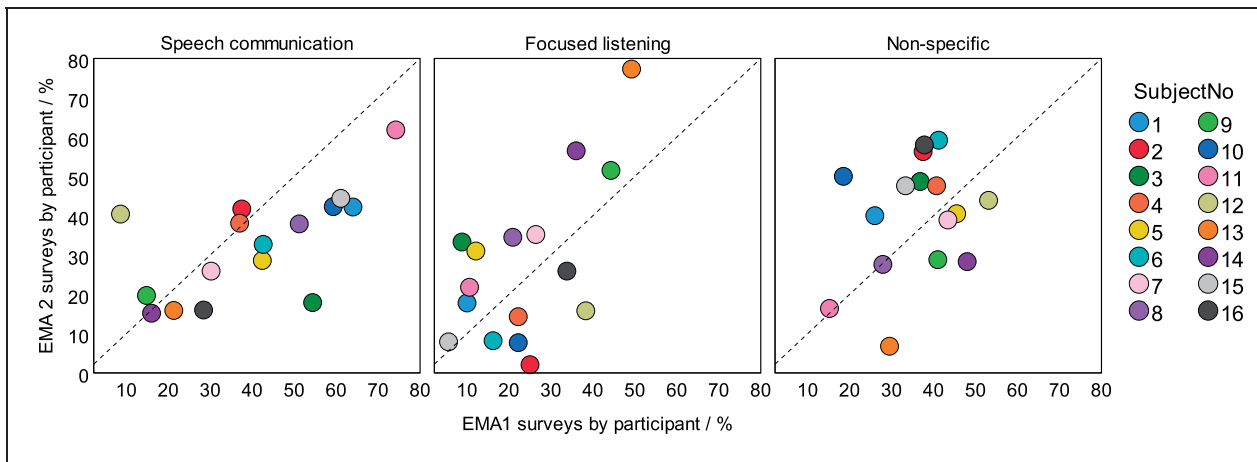


Figure 5. Proportions of the CoSS Intention Categories “Speech Communication” (Left Panel), “Focused Listening” (Middle Panel), and “Nonspecific” (Right Panel) in EMA1 and EMA2 Surveys for Each Returner Participant. Participants are numbered as denoted in Table 1 and coded by colors. EMA = ecological momentary assessment.

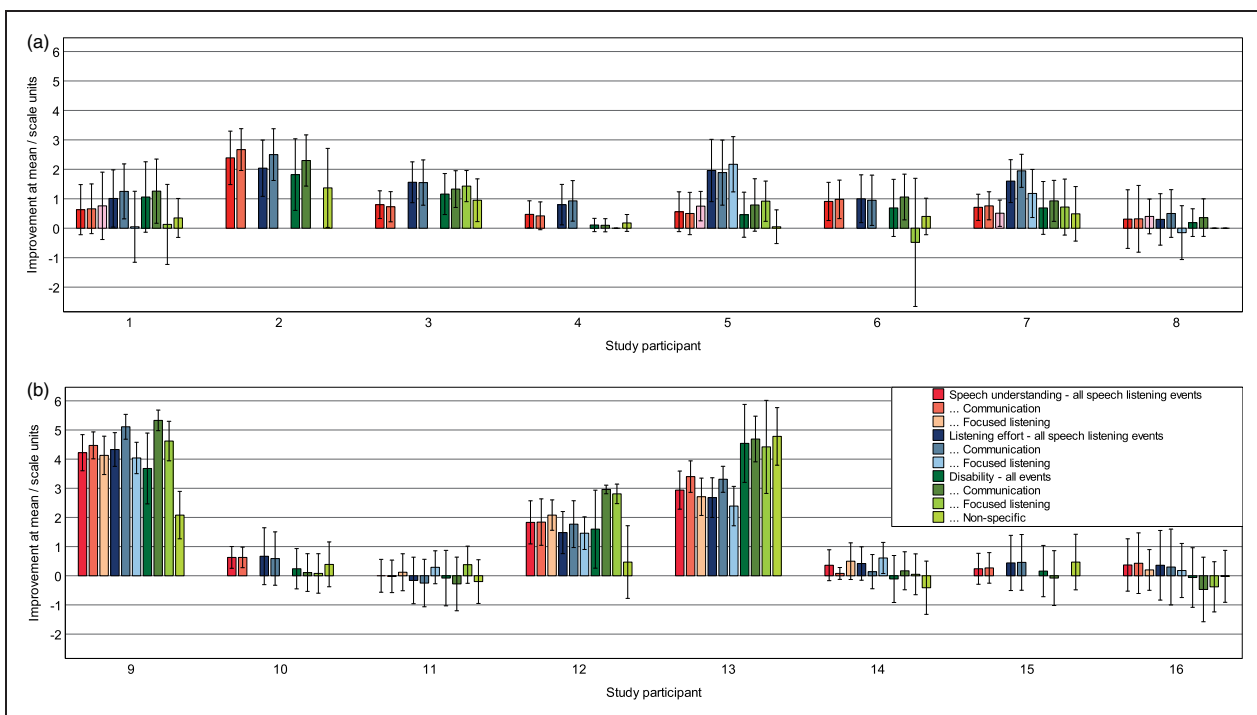


Figure 6. Individual Score Change for Three Hearing Dimensions in 13 First-Time HA Users (Nrs. 1–13) and 3 Follow-up HA Users (Nrs. 14–16). Score differences were calculated by subtracting the mean score from EMA1 and EMA2, if a minimum of three surveys was available in both phases. Otherwise, no score difference was calculated (no bars shown). Error bars show the pooled standard deviation which was calculated as the square root of the average of the variances for EMA1 and EMA2 scores. Data basis: $n = 933$ (speech understanding and listening effort), $n = 1,700$ (disability). The numbering of participants corresponds to Table 1.

Difference in the Mean. Figure 6 shows the difference in the average scores from EMA1 to EMA2 for each study participant for the dimensions speech understanding, listening effort, and disability. The scale was reversed for “listening effort” and “disability” to facilitate the comparison, that is, positive values show improvements. Results are shown separately for CoSS intentions

categories when appropriate. A mean score was only calculated if there were at least three assessments in each EMA phase.

With few exceptions, the intraindividual standard deviations were higher in EMA1, ranging from 0.3 to 1.7 (median 0.8), than in EMA2, with a range from 0 to 1.2 (median 0.6). From EMA1 to EMA2, all

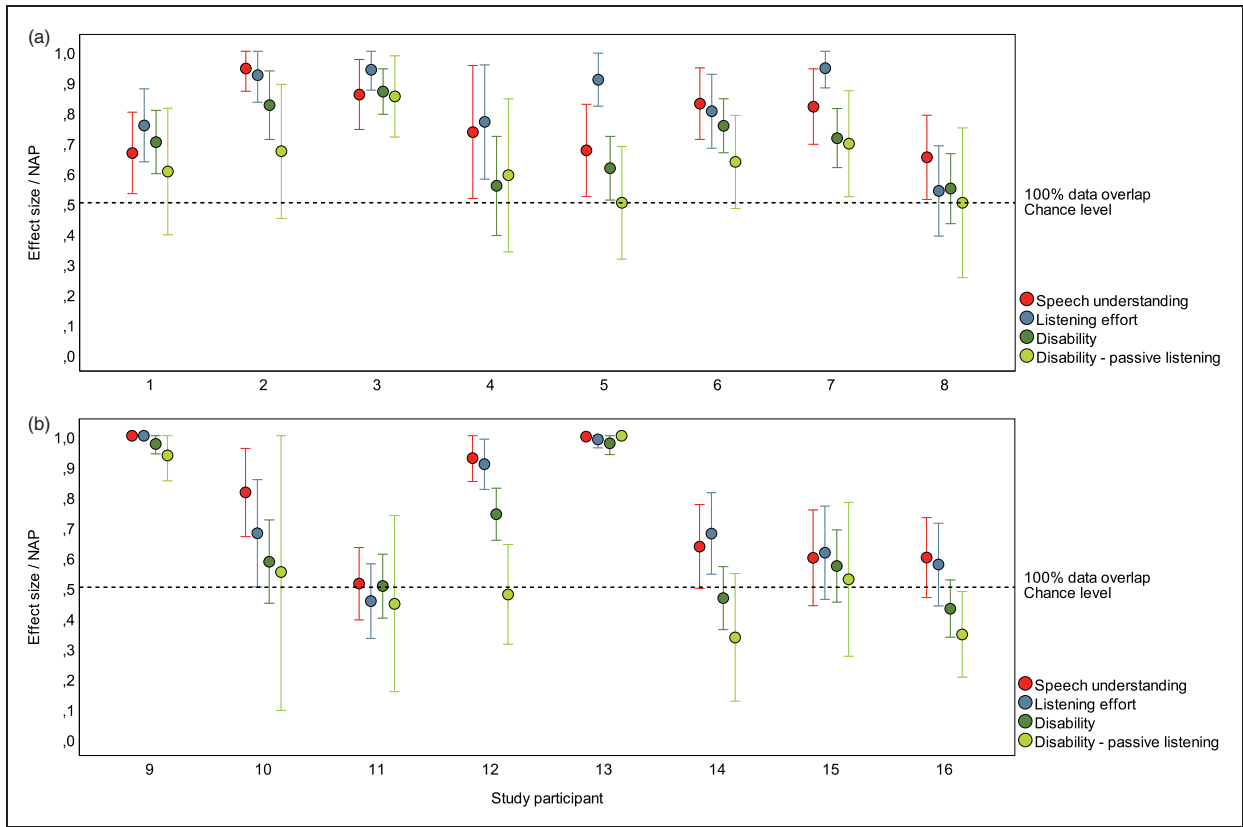


Figure 7. Individual ES and 95% CIs for Three Hearing Dimensions in 13 First-Time HA Users (Nrs. 1–13) and 3 Follow-up HA Users (Nrs. 14–16). Data basis: $n = 933$ (speech understanding and listening effort), $n = 1,700$ (disability), $n = 531$ (disability—passive listening). Numbering of participants corresponds to Table 1. NAP = nonoverlap of all pairs.

participants showed an improvement ranging between 0.5 and 2 scale units for most participants and dimensions. Deterioration from EMA1 to EMA2 was rare and numerically small. Experienced HA users reported less disability than first-time HA users in EMA1 and were among the participants whose average ratings hardly changed from EMA1 to EMA2. Score differences of more than three scale units were observed in only two participants (Nrs. 9, 13). The improvements were often comparable for the CoSS categories speech communication and focused listening across all dimensions. There are exceptions, such as Participants 5 and 7, who clearly showed stronger improvements in listening effort compared with speech understanding and disability. Participant 13, conversely, showed a larger score difference for the disability dimension than for speech understanding and listening effort. When experienced HA users were excluded, the changes in mean EMA scores were not correlated to either ear PTA or to duration of the HA adjustment period, but the HHIE total score did correlate significantly with the score change in speech understanding ($r = .64, p = .020$) and disability ($r = .59, p = .034$).

ES Estimate. ES based on NAP was calculated for all dimensions. Figure 7 shows ES and 95% CI for speech understanding, listening effort, and disability. ES for disability is shown separately for the surveys related to the CoSS task category passive listening and for all surveys. Based on all speech-listening events, ES median and interquartile range (IQR) were 0.773 (IQR 0.638–0.909) for speech understanding and 0.785 (IQR 0.630–0.935) for listening effort, with NAP reaching significance in 12 and 11 participants, respectively. ES for localization was smaller, at 0.676 (IQR 0.591–0.858). Considering the disability dimension, the median ES was 0.657 (IQR 0.549–0.805) when all events were included and reached significance in nine participants. ES reached a ceiling in Participants 9 and 13. Conversely, very small and even negative effects were observed for first-time HA users 8 and 11, and particularly in experienced HA users.

If applicable, NAP analyses were broken down into CoSS intentions and tasks categories. The report is limited here to dimensions for which the mixed-model results (see later) suggested analyzing individual EMA data. Significant ES were observed for mood in seven

participants (Nrs. 1, 2, 3, 4, 9, 13, and 16) and for pleasantness in six participants (Nrs. 1, 2, 6, 8, 9, and 13). More importantly, no significant effects in the reverse direction were found for either dimension, except for Participant No. 11, who assessed sound to be less pleasant in the aided than the unaided condition. ES for disability in passive listening tasks achieved significance in four participants (Nrs. 3, 7, 9, and 13). No significant ES were found for loudness, except for Participant 11, with sounds being perceived as louder in the aided than in the unaided condition. ES for loudness was uncorrelated to the duration of the HA adjustment period (all returner participants: $r = .11$, $p = .684$; first-time HA users: $r = .019$, $p = .950$).

HA Benefit in the Group of First-Time HA Users. The baseline condition was completely different for experienced HA users and first-time HA users. The results presented in this section therefore refer to 1,330 surveys collected by first-time HA users (Nrs. 1-13).

Figure 8 shows the results from the CLMM analyses for all survey dimensions. Analyses were performed on all valid data. Nevertheless, the number of surveys included differs because of content and the adaptive flow of the digital questionnaire (i.e., speech understanding cannot be rated if speech is not the target). Active filtering was carried out for the assessment of loudness, pleasantness, and localization abilities. For conceptual reasons, 53 surveys that related to passive listening were removed from the analyses. If neither a target source nor a sound source was specified, it is unclear to which percept the ratings on these items apply when given retrospectively. However, filtering had practically no effect on the estimates in this study. Note that the beta coefficients do not refer to categorical units but to log odds ratios. The change of assessments was considered to be significant if 95% CI did not include 0. Beta coefficients ranged between $|1.6|$ and $|4.1|$ for those

hearing-related dimensions in which improvement was directly or indirectly targeted. The absolute beta coefficients were larger for speech understanding and listening effort than for the items belonging to the psychosocial-communicative domain, such as difficulty by conversation partner and involvement.

Decision for or Against Permanent HA Use

Among the 20 participants who were seeking a HA fitting for the first time, 13 opted for HA acquisition (Nrs. 1-4, 6-13, 24 in Table 1) and 7 canceled the HA fitting process (no. 5, 18-23). HA acquisition or rejection was not related to gender, professional skill level, or employment status. PTA and mean HHIE scores tended to be higher for participants who opted for HA acquisition but did not differ significantly from those who decided against it (wide overlap of bootstrapped 95% CI). With regard to their EMA1 profile, the compliance_{A+M} was, on average, higher in future HA users than in future nonusers, but 95% CI still overlapped widely. In addition, both groups' auditory reality in terms of CoSS intentions and the situations assessed was well comparable. Mean EMA1 assessments of speech understanding, listening effort, and disability ratings of both groups were compared in all EMA1 data as well as separately for CoSS intention categories (except focused listening category, due to the insufficient number of cases). Comparative analyses established no difference with regard to EMA1 assessments between the participants who decided for permanent HA use and those who decided against it.

Discussion

This study described the auditory reality of adults with mild-to-moderate hearing loss who were seeking a HA fitting. They conducted an EMA before HA fitting and after a period of HA adjustment. The EMA sampling

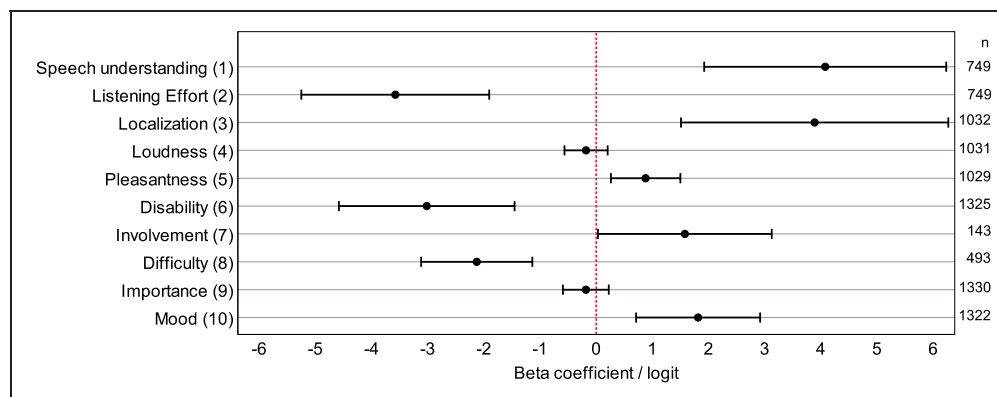


Figure 8. Beta Coefficients and 95% CI From CLMM for All Assessment Dimensions in 13 First-Time HA Users (Nrs. 1-13). The number of surveys included in each CLMM analysis is shown on the right side.

strategy was, to put it briefly, short and intense. Each EMA phase lasted 3 to 4 days, with surveys scheduled about every half hour. In total, 2,042 surveys collected by 24 participants were analyzed, focusing on the auditory reality of HA candidates and the HA benefit in natural environments.

Auditory Reality

The EMA data categorized according to the CoSS framework showed that most of the surveys related to speech communication (42%) and the least to the focused listening category (23%). Smeds et al. (2019) and Jensen et al. (2019) conducted EMA studies with experienced HA users of about the same age as in this study. They found lower shares for speech communication, 30% in Smeds et al. (2019) and a little over 30% in Jensen et al. (2019, depicted from plot), and comparable shares for the focused listening category. On the task level, the studies agree well with regard to the ranking of the shares. Speech communication referred mostly to conversations with one person, focused listening referred predominantly to media use, and nonspecific intentions were mostly related to passive listening tasks.

Most of the surveys in this study were taken at home and *momentary*, that is, taken while being in the situation. Surveys related to various out-of-home situations were less frequent and more often delayed. This result confirmed the findings of Schinkel-Bielefeld et al. (2020) who examined the situational context of delayed assessments. They concluded that social situations will likely be underrepresented in EMA. Anecdotal comments in this study also pointed in this direction. A few participants seemed to avoid a coincidence of family gatherings, journeys, or other special events, when arranging an appointment for EMA2.

HA Benefit

The mean score differences of the prefitting EMA1 and the postfitting EMA2, as well as ES, varied greatly from individual to individual but were uncorrelated with both hearing loss and the duration of HA adjustment period. For most participants, the ratings of real-life hearing abilities showed improvements from EMA1 to EMA2 that were mostly larger for first-time HA users than for experienced HA users.

Individual and Group Estimates. To some degree, the shares of CoSS intention categories in individual data differed from EMA1 to EMA2, indicating a distinct auditory reality. Nevertheless, ES and the individual mean score differences of EMA1 and EMA2 were often similar across the CoSS intention categories for the main hearing dimensions, particularly speech understanding and

listening effort. No, or hardly any, effect was observed in Participants 8 and 11, although they decided for HA acquisition. Participant 8 had an elevated mild hearing loss and a low hearing handicap, as indicated by a HHIE score of 12 points. Participant 11 had a very mild hearing loss but a high HHIE score of 34 points. Assuming that they not only bought HA but continued to use them, three aspects, among others, could possibly explain the overall small-to-negligible effect seen by EMA. First, events in which the devices were beneficial could have been avoided or simply missed in the short EMA phases. Second, the scales may not have been sensitive enough to map small, but noticeable, perceptual changes. Third and last, HA fitting could be driven by external factors with low intrinsic motivation.

Conversely, extreme improvements were observed in Participants 9 and 13. Participant 9 had a mild, asymmetric hearing loss, with almost perfect hearing in the better ear. Participant 13 had a bilateral, moderate hearing loss. Despite their completely different audiometric profiles, the measured benefit may still be linked to the audiometric type and degree of hearing loss. The PTA of Participant 13 was well above the cut-off value at which HA fitting is regarded as beneficial (Humes, 2019). Participant 9, in turn, may have experienced the asymmetry in hearing as a severe strain. The asymmetric profile had developed recently—in the last few years—and the normal-hearing ear might still have been perceived as an irritating reference. Explanations based on audiometric findings, however, are quite likely to fall short in the self-assessment of hearing abilities. As in established questionnaires, such as the Speech, Spatial, and Qualities of Hearing Scale (Gatehouse & Noble, 2004) and the APHAB (Cox & Alexander, 1995), EMA still captures the subjective perspective on activity limitations, and perfect agreement with audiometric findings is neither intended nor achieved (Banh et al., 2012; Dornhoffer et al., 2020; von Gablenz et al., 2018). Complex, temporally dynamic interactions of personality and context, which probably impact the self-assessment of abilities, are in general difficult to disentangle and certainly require complementary methods and a larger study sample when EMA data are concerned.

This study focused on individual analyses but also provided an effect overview for the subgroup of first-time HA users. As the data type and distribution did not meet the assumptions for conventional linear mixed modeling, CLMM analyses were performed. The sample size was too small to additionally consider participant-related covariates such as the degree and configuration of hearing loss. The distribution of CoSS categories, that is, differences in auditory reality in EMA1 compared with EMA2, was not effectively controlled for. The CLMM results should therefore be

appreciated qualitatively rather than according to their exact numerical coefficient estimate. Even controlling for CoSS, however, the authors would not have expected completely different results for the group analysis. The differences in HA benefit between the CoSS intention categories were rather small for most hearing dimensions and the individual distributional mismatch of CoSS categories can be considered random in this study.

Different scales, items and item wordings, and analysis methods hamper direct comparisons to the study of Timmer et al. (2018), which is the most similar to the present. However, after HA intervention, both studies found significant improvements for various hearing dimensions.

Assessment Dimensions. The largest improvements after the HA adjustment period were observed for speech understanding and listening effort. The ratings on these dimensions were correlated with each other and also with ratings on other dimensions, such as localization or difficulty by conversation partners. The latter was included to capture a psychosocial aspect of hearing-related communication constraints. Although the survey dimensions are conceptually different from the researchers' perspective, some of them are likely to coincide or be easier to assess from the experiences perspective (Moore & Picou, 2018). Such linking of assessments is known from retrospective questionnaires but also shows up in EMA in which retrospective considerations should be replaced by spontaneous decisions. For the dimensions mood and pleasantness of sounds, significant effects, in terms of being in a better mood and the sounds being perceived as more pleasant, were observed in more than one third of the participants. These results were confirmed on the group level for first-time HA users.

The dimension disability deserved particular attention, as the corresponding item was included in each survey (How much do you feel disabled?). As defined earlier, passive listening tasks refer to situations in which either no sound source, or no target signal relevant to the current activity, was present. One might argue that then, hearing should hardly play a role. The participants, however, often reported a disability in the EMA1 surveys even for these situations (Figure 4). On the one hand, participants might have felt insecure in certain environments or in the anticipation of an upcoming situation in which hearing plays a role. On the other hand, this finding may be partially explained by the stage at which the participants were recruited for the study. All participants were recruited at their first appointment with an acoustician, and started EMA within a few days, often the next day, so as not to further postpone HA fitting. By that time, the perception of hearing-related disability had possibly generalized and

peaked, followed by a flattening out in subsequent weeks and months. This could also explain the high hearing handicap (measured using HHIE) observed in this sample, corresponding to relatively less hearing loss, compared with the results reported by Ventry and Weinstein (1982). After the HA adjustment period, in turn, the participants' disability ratings could be influenced by the desire to meet expectations of a good outcome, both their own and others'. This could have led to an overestimate of effects in almost all dimensions. Thus, significant effects in the CoSS task category "passive listening" may serve as an indicator that effects may have been overestimated. Against this background, the results for Participants 3, 7, 9, and 13 (see NAP in Figure 7) should be interpreted with caution.

No change was observed for the ratings on the importance of hearing well and loudness. With regard to the latter, this finding might be unexpected, given that amplification is supposed to alter loudness perception. There can be two explanations for this, and perhaps they both apply. First, the participants conducted EMA2 when they felt comfortable with the HA settings and assumed the fitting process to be completed. Although the HA adjustment period was rather short for two participants, it took about 3 months or longer for two thirds of the participants. At this stage, they might have already adjusted to the acoustic changes and altered their loudness perception. Second, the scale used in this study is possibly insensitive to minor alterations. The scale categories are the same as used in the adaptive procedure for categorical loudness scaling, which was evaluated by Brand and Hohmann (2002). However, the half steps were removed in this study because the smartphone screens were too small for an 11-point scale.

Deciding Against HA Use

Seven of 20 potential first-time HA users canceled the HA fitting process after EMA1. Neither hearing loss, HHIE score, nor EMA1 offered any convincing explanation as to why some decided for and others against HA use. In general, the participants' motives and attitudes toward HA fitting do not show up in EMA data. It is important to note that the sample size was too small to support an EMA data analysis using advanced predictive statistical approaches. Therefore, current research cannot answer the question as to whether EMA in the unaided condition could predict who is more likely to dropout or to complete the HA fitting process.

Study Characteristics and Analysis

Reactivity, Compliance, and Data Quality. Shiffman et al. (2008) stated that researchers should be alert for reactivity in EMA studies. Henry et al. (2012) used the Tinnitus Handicap Inventory-Screening version (Newman et al.,

2008), whereas Timmer et al. (2018) and Galvez et al. (2012) applied the HHIE as an independent measure before and after an EMA phase. None of these studies found signs of reactivity. However, no evidence is provided that these questionnaires were sensitive enough to detect reactivity in EMA hearing studies. As a change in assessment criteria should be manifested in the EMA data itself, we examined the assessments in their original time sequence. Overall, no evidence of reactivity was found.

Compliance_A was 50% in this study, and thus lower than reported in other hearing studies, for example, Jensen et al. (2019) reported 79%, Galvez et al. (2012) 77%, and Henry et al. (2012) 90%. When self-initiated surveys were also considered, compliance_{A+M} increased to 73% on the group level. Individual compliance rates varied considerably, as was also observed in the studies cited earlier, and, more recently, by Timmer et al. (2017), Andersen et al. (2019), and Schinkel-Bielefeld et al. (2020). In general, comparing compliance rates across studies is doubtful if the numbers are not interpreted against the background of the particular study objective, the EMA sampling strategy, and the participant recruitment. In this study, the compliance rates were estimated based on the usage time of the EMA device, which is subject to uncertainty due to the technical configuration of olMEGA. The usage time might include time periods in which the participants did not wear the EMA system, for example, when doing sports or taking a nap. Moreover, the prompts were discreetly designed to not further increase the participants' burden and were thus probably often missed. Considering these uncertainties, together with the closely timed prompts and the participant recruitment from a research-inexperienced population, the overall compliance_{A+M} is quite good.

Importantly, the compliance figures do not substantiate whether the participants complied with the concept-related content of the instructions. We tried to increase this kind of higher order compliance by sharing the participant's results after each EMA phase. The general idea was to turn the data providers into beneficiaries of the EMA study, who then, in their own interest, stay authentic in their assessments. Nonetheless, EMA data sporadically contained assessment combinations that, at least from the analyst's perspective, called the consistency of the survey into question. We suspect that this phenomenon has been observed in other studies as well. Such possibly "murky" data samples, that are few in number and not attributable to specific participants, cannot be the subject of data cleansing, because the deeper situational context always remains unknown in EMA, and the participants are ultimately taken at their word. A run-in phase of several days, as suggested by Piasecki et al. (2007), and implemented, for example, by Wu et al. (2018), could have stabilized the assessments

but was not an option in this study to avoid unduly delaying HA fitting.

It is worth considering whether the feedback on EMA1 results given to the participants could have altered behavior in EMA2. From the authors' point of view, this seems less likely. The feedback graphics illustrated the participants' self-reporting. In this respect, it is not much different from using a questionnaire before and after HA intervention. Questionnaire respondents might also remember their earlier rating and somehow reference to it. Using the feedback graphic (or the memory of it) as a reference would have been even more difficult in EMA. The everyday listening situations were diverse and the prompts closely timed. On average, 3 months elapsed before returning for EMA2 and participants often passed the feedback printouts on to their HA acousticians. When they received the comparison between results for EMA1 and EMA2 at the final visit, it seemed that most participants did not remember the earlier outcome. Finally, if the feedback on EMA1 results hypothetically biased the ratings, then this assumption should apply both to first-time HA users and experienced HA users. The ratings of the latter, however, did not basically change from EMA1 to EMA2.

Effect Estimate for EMA Studies in AB Design. As noted by Jacobs and Kaye (2015), EMA is still in its "infancy" in hearing research. Further research is needed to extend and to refine the analytical approaches. In the current single-subject study, NAP was used to give a nonparametric ES estimate for the HA intervention. NAP belongs to the group of nonoverlap techniques (Manolov et al., 2016; Parker & Vannest, 2009; Parker et al., 2011) and combines various advantages. NAP provides a key figure that allows for comparison of data with any distribution and type. NAP is close to a visual data inspection, which is intuitively easy to understand, and therefore suitable for more practically oriented EMA applications. NAP is an interpretation of the *empirical* AUC and thus simple to compute. It comes with CIs that account for the uncertainty of the ES estimate, which is highly valuable both in research and clinical practice. Depending on the context of use, it may be reasonable to set lower levels of CI than 95%. Apart from these advantages, NAP shares the weaknesses of other ES measures. It is blind toward the distribution properties, trends, and autocorrelation (e.g., Archer et al., 2019; Kratochwill et al., 2010). Autocorrelation is present if error terms are serially correlated. More specifically, time-series data like EMA can share significant associations that are induced by the sampling strategy (Archer et al., 2019).

Strengths and Limitations

This study investigated the change of hearing abilities and related dimensions at two milestones in the aural rehabilitation process—on the verge of HA fitting and after an HA adjustment period. As its strengths, three aspects deserve special emphasis. First, the study participants were recruited from the general population of adults. Their primary interest was not participating in a research study but HA fitting. Thus, the recruiting presumably had a higher ecological fit than the recruiting from a pool of research-eager volunteers. Second, the data analysis was collaborative in nature. Since to date the sample size was small in EMA hearing studies, we refrained from employing the categories of oMEGA and used the CoSS framework. Third, an innovative approach to the analysis of EMA data was proposed. NAP has proven its value in other single-subject studies but has not yet been used for audiological EMA studies.

However, there are also several limitations that must be taken seriously. First, the AB design, in combination with the small sample size, did not permit causal inferences or allow the control of carry-over effects. This is basically the other side of the coin of the recruiting strategy in this study, as withdrawal of HA was ethically unacceptable. Second, the cooperation with several HA acousticians in the recruiting of participants impeded the calculation of a response ratio. It is unclear how many HA customers were invited to participate in the study and whether the acousticians applied unexplained and personalized criteria. Especially the last aspect can compromise one of the strengths of the study, the ecological fit of the participant recruitment. Third, information on the assessment delay is missing in 8 % of the surveys because the entries were kept nonmandatory for all items. Fourth, the EMA design of this study has two main disadvantages. The closely timed prompts probably increased both the autocorrelation in the EMA data and the participants' burden. Moreover, the EMA phases were very short. EMA1 could not be extended without unacceptably delaying the HA first fit. It is in fact questionable whether 3 to 4 days were sufficient to capture and to represent rare but important everyday situations. This leads to the fifth limitation; a differentiation of the surveys down to the CoSS task level was not possible for all participants, as a minimum of three assessments were deemed necessary. The analysis was largely focused on the CoSS intention category, within which rare and important events could be hidden. Differences in effects that may exist for diverse CoSS tasks or even scenarios could have been leveled out and remain undetected. EMA studies often face a classical trade-off problem, where insufficiently differentiated categorization and data fragmentation must be weighed against each other. However, this dilemma

might not be relevant in more practical applications such as HA counseling, where reporting of singular events can be instructive.

Conclusions

EMA was used to collect self-reports on hearing in participants who were seeking a first HA fitting or a renewal. Participants conducted EMA before HA fitting as well as after an individual HA adjustment period. Based on 2,042 surveys, this contribution studied three research questions: First, the auditory reality of hearing-impaired adults using the CoSS framework; second, the effects of HA fitting on self-rated hearing abilities and related dimensions; and third, the option to predict who would be likely to drop out or to complete the HA fitting process. The analysis showed that situations in which the participants conversed with natural speakers took the largest share in the preintervention EMA surveys. Hearing-related disabilities of varying degrees shaped the participants' auditory reality. Disabilities were most pronounced in communications tasks but also reported for situations for which no listening target had been specified. After the HA adjustment period, EMA showed in most participants better speech understanding as well as decreased listening effort and diminished disability. The effects were small in experienced HA users, but were in general greater in first-time HA users, with large inter-individual differences. The individual strengths of the effects were not associated with the duration of the HA adjustment period or hearing loss. Some participants canceled the HA fitting process after a few weeks or months. EMA data collected in the unaided condition did not predict who would be likely to drop out or to complete HA fitting. The nonparametric measure NAP was used to evaluate the strength of treatment effects. NAP was found to be an appropriate method that could guide the interpretation of individual EMA data in aural rehabilitation.

Acknowledgments

The authors thank Alejandro Garavito Arango for preparing the data export routines, Maximilian Hehl, Florian Schmitt, Kristin Sprenger, and Annäus Wiltfang for performing the laboratory measurements, the local HA acousticians for their kind cooperation, and two reviewers for helpful comments on earlier versions of the paper. English language services were provided by www.stels-ol.de.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Hearing Industry Research Consortium and the Research Fund of Jade University of Applied Sciences. Publication was supported by the Open Access Publication Funds of Jade University of Applied Sciences.

ORCID iD

Petra von Gablenz  <https://orcid.org/0000-0002-7346-7950>

Supplemental material

Supplemental material for this article is available online.

References

- Andersen, L. S., Edinger Anderesson, K., Wu, M., Pontoppidan, N., Bramslov, L., & Neher, T. (2019, August). Assessing daily-life benefit from hearing aid noise management: SSQ12 vs. ecological momentary assessment. *Proceedings of the International Symposium on Auditory and Audiological Research*, 7, 273–280. <https://proceedings.isaar.eu/index.php/isaarproc/article/view/2019-31>
- Archer, B., Azios, J. H., Müller, N., & Macatangay, L. (2019). Effect sizes in single case aphasia studies: A comparative, autocorrelation-oriented analysis. *Journal of Speech, Language, and Hearing Research*, 62(7), 2473–2482. https://doi.org/10.1044/2019_JSLHR-L-18-0186
- Banh, J., Singh, G., & Pichora-Fuller, M. K. (2012). Age affects responses on the Speech, Spatial, and Qualities of Hearing Scale (SSQ) by adults with minimal audiometric loss. *Journal of the American Academy of Audiology*, 23(2), 81–91. <https://doi.org/10.3766/jaaa.23.2.2>
- Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological Methods*, 16(4), 373–390. <https://doi.org/10.1037/a0025813>
- Brand, T., & Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *The Journal of the Acoustical Society of America*, 112(4), 1597–1604. <https://doi.org/10.1121/1.1502902>
- Burke, L. A., & Naylor, G. (2020). Daily-life fatigue in mild to moderate hearing impairment: An ecological momentary assessment study. *Ear and Hearing*. Advance online publication. <https://doi.org/10.1097/AUD.0000000000000888>
- Cox, R. M., & Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing*, 16(2), 176–186.
- Davis, D. H. J., Creavin, S. T., Yip, J. L. Y., Noel-Storr, A. H., Brayne, C., & Cullum, S. (2015). Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. *Cochrane Database of Systematic Reviews*, 10, 1–50. <https://doi.org/10.1002/14651858.CD010775.pub2>
- Dornhoffer, J. R., Meyer, T. A., Dubno, J. R., & McRacken, T. R. (2020). Assessment of hearing aid benefit using patient-reported outcomes and audiologic measures. *Audiology & Neurotology*. Advance online publication. <https://doi.org/10.1159/000506666>
- Galvez, G., Turbin, M. B., Thielman, E. J., Istvan, J. A., Andrews, J. A., & Henry, J. A. (2012). Feasibility of ecological momentary assessment of hearing difficulties encountered by hearing aid users. *Ear and Hearing*, 33(4), 497–507.
- Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, 43(2), 85–99. <https://doi.org/10.1080/14992020400050014>
- Groenewold, H., Franz, S., Holube, I., & Bitzer, J. (2018, May 23–26). Wearable mobile Bluetooth device for stereo audio transmission to a modified Android smartphone [Convention]. 144th Convention of the Audio Engineering Society, Milano, Italy.
- Hasan, S. S., Chipara, O., Wu, Y.-H., & Aksan, N. (2014). Evaluating auditory contexts and their impacts on hearing aid outcomes with mobile phones. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (pp. 126–133). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). <https://doi.org/10.4108/icst.pervasivehealth.2014.254952>
- Henry, J. A., Galvez, G., Turbin, M. B., Thielman, E. J., McMillan, G. P., & Istvan, J. A. (2012). Pilot study to evaluate ecological momentary assessment of tinnitus. *Ear and Hearing*, 32(2), 279–290.
- Holube I., von Gablenz, P., & Bitzer, J. (2020). Ecological momentary assessment (EMA) in audiology—Current state, challenges, and future directions. *Ear and Hearing*, 41(S1), 79S–90S. <https://doi.org/10.1097/AUD.0000000000000934>
- Humes, L. E. (2019). The World Health Organization's hearing-impairment grading system: An evaluation for unaided communication in age-related hearing loss. *International Journal of Audiology*, 58(1), 12–20. <https://doi.org/10.1080/14992027.2018.1518598>
- International Standard Classification of Occupations. (2012). *Structure, group definitions and correspondence tables—ISCO-08 (Vol. 1)*. International Labour Office.
- Jacobs, P. G., & Kaye, J. A. (2015). Ubiquitous real-world sensing and audiology-based health informatics. *Journal of the American Academy of Audiology*, 26(9), 777–783. <https://doi.org/10.3766/jaaa.15010>
- Jensen, N. S., Hau, O., Lelic, D., Herrlin, P., Wolters, F., & Smeds, K. (2019, September 9–13). Evaluation of auditory reality and hearing aids using an ecological momentary assessment (EMA) approach. In M. Ochmann, M. Vorländer, & J. Fels (Eds.), *Proceedings of the 23rd International Congress on Acoustics* (pp. 6545–6552).
- Keidser, G., Dillon, H., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, 1(1), 88–90. <https://doi.org/10.4081/audiores.2011.e24>
- Kowalk, U., Kissner, S., von Gablenz, P., Holube, I., Bitzer, J. (2017, August). An improved privacy-aware system for objective and subjective ecological momentary assessment. In S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjerg, T.

- Andersen, & T. Poulsen (Eds.), *Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing* (pp. 25–30B). The Danavox Jubilee Foundation.
- Kratochwill, T., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf
- Manolov, R., Losada, J. L., Chacón-Moscoso, S., & Sanduvete-Chaves, S. (2016). Analyzing two-phase single-case data with non-overlap and mean difference indices: Illustration, software tools, and alternatives. *Frontiers in Psychology, 7*, 32. <https://doi.org/10.3389/fpsyg.2016.00032>
- Moore, T. M., & Picou, E. M. (2018). A potential bias in subjective ratings of mental effort. *Journal of Speech, Language, and Hearing Research, 61*, 2405–2421. https://doi.org/10.1044/2018_JSLHR-H-17-0451
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699.
- Newman, C. W., Sandridge, S. A., & Bolek, L. (2008). Development and psychometric adequacy of the screening version of the tinnitus handicap inventory. *Otology & Neurotology, 29*(3), 276–281. <https://doi.org/10.1097/MAO.0b013e31816569c4>
- Noble, W. (2008). Auditory reality and self-assessment of hearing. *Trends in Amplification, 12*(2), 113–120. <https://doi.org/10.1177/1084713808316172>
- Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*(4), 357–367. <https://doi.org/10.1016/j.beth.2008.10.006>
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35*(4), 303–322. <https://doi.org/10.1177/0145445511399147>
- Piasecki, T. M., Hufford, M. R., Solhan, M., & Trull, T. J. (2007). Assessing clients in their natural environments with electronic diaries: Rationale, benefits, limitations, and barriers. *Psychological Assessment, 19*(1), 25–43. <https://doi.org/10.1037/1040-3590.19.1.25>
- Probst, T., Pryss, R. C., Langguth, B., Rauschecker, J. P., Schobel, J., Reichert, M., Spiliopouou, M., Schlee, W., & Zimmermann, J. (2017). Does tinnitus depend on time-of-day? An ecological momentary assessment study with the “TrackYourTinnitus” application. *Frontiers in aging neuroscience, 9*, article 253. <https://doi.org/10.3389/fnagi.2017.00253>
- Ricketts, T. A., Bentler, R., & Mueller, H. G. (2019). *Essentials of modern hearing aids: Selection, fitting, and verification*. Plural Publishing.
- Schinkel-Bielefeld N., Kunz, P., Zutz, A., & Buder, B. (2020). Evaluation of hearing aids in everyday life using ecological momentary assessment—What situations are we missing? *American Journal of Audiology, 29*(3S), 591–609. https://doi.org/10.1044/2020_AJA-19-00075
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Smeds, K., Dahlquist, M., Larsson, J., Herrlin, P., & Wolters, F. (2019, September 9–13). LEAP, a new laboratory test for evaluating auditory preference. In M. Ochmann, M. Vorländer, & J. Fels (Eds.), *Proceedings of the 23rd International Congress on Acoustics* (pp. 7608–7615).
- Stone, A., Shiffman, S., Atienza, A., & Nebeling, L. (2007). Historical roots and rationale of ecological momentary assessment (EMA). In A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research*. Oxford University Press.
- Timmer, B. H., Hickson, L., & Launer, S. (2017). Ecological momentary assessment: Feasibility, construct validity, and future applications. *American Journal of Audiology, 26*(3S), 436–442. https://doi.org/10.1044/2017_AJA-16-0126
- Timmer, B. H., Hickson, L., & Launer, S. (2018). Do hearing aids address real-world hearing difficulties for adults with mild hearing impairment? Results from a pilot study using ecological momentary assessment. *Trends in Hearing, 22*, 1–15. <https://doi.org/10.1177/2331216518783608>
- Ventry, I. M., & Weinstein, B. E. (1982). The Hearing Handicap Inventory for the Elderly: A new tool. *Ear and Hearing, 3*(3), 128–134. <https://doi.org/10.1097/00003446-198205000-00006>
- von Gablenz, P., Otto-Sobotka, F., & Holube, I. (2018). Adjusting expectations: Hearing abilities in a population-based sample using an SSQ short form. *Trends in Hearing, 22*, 1–21. <https://doi.org/10.1177/2331216518784837>
- 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. *Journal of the American Academy of Audiology, 27*(7), 527–540.
- Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J. (2018). Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear and Hearing, 39*(2), 293–304. <https://doi.org/10.1097/AUD.0000000000000486>