



Evaluation of Accuracy and Reliability of a Mobile Screening Audiometer in Normal Hearing Adults

Angela Colsman¹, Gernot G. Supp^{2,3}, Joachim Neumann⁴ and Till R. Schneider^{1*}

¹ Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, ² Sonormed GmbH, Hamburg, Germany, ³ Independent Researcher, Hamburg, Germany, ⁴ Mindo Software SLU, Barcelona, Spain

OPEN ACCESS

Edited by:

Etienne De Villers-Sidani,
McGill University, Canada

Reviewed by:

Adrian Fuente,
Université de Montréal, Canada
Philippe Fournier,
Aix-Marseille Université, France

*Correspondence:

Till R. Schneider
t.schneider@uke.de

Specialty section:

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

Received: 18 August 2019

Accepted: 26 March 2020

Published: 29 April 2020

Citation:

Colsman A, Supp GG, Neumann J and Schneider TR (2020) Evaluation of Accuracy and Reliability of a Mobile Screening Audiometer in Normal Hearing Adults. *Front. Psychol.* 11:744. doi: 10.3389/fpsyg.2020.00744

Quantifying hearing thresholds via mobile self-assessment audiometric applications has been demonstrated repeatedly with heterogenous results regarding the accuracy. One important limitation of several of these applications has been the lack of appropriate calibration of their core technical components (sound generator and headphones). The current study aimed at evaluating accuracy and reliability of a calibrated application (app) for pure-tone screening audiometry by self-assessment on a tablet computer: Audimatch app installed on Apple iPad 4 in combination with Sennheiser HDA-280 headphones. In a repeated-measures design audiometric thresholds collected by the app were compared to those obtained by standardized automated audiometry and additionally test-retest reliability was evaluated. Sixty-eight participants aged 19–65 years with normal hearing were tested in a sound-attenuating booth. An equivalence test revealed highly similar hearing thresholds for the app compared with standardized automated audiometry. A test-retest reliability analysis within each method showed a high correlation coefficient for the app (Spearman rank correlation: $\rho = 0.829$) and for the automated audiometer ($\rho = 0.792$). The results imply that the self-assessment of audiometric thresholds via a calibrated mobile device represents a valid and reliable alternative for stationary assessment of hearing loss thresholds, supporting the potential usability within the area of occupational health care.

Keywords: audiometry, hearing test, air conduction, mobile application, automated audiometry, self-assessment, hearing threshold

INTRODUCTION

Hearing loss together with its negative personal and socio-economic consequences represents a serious health issue across cultures and around the globe. According to the World Health Organization (WHO, 2017) about 466 million people are already affected by a disabling hearing loss, while 1.1 billion young people (12–35 years) are at risk of hearing loss. The personal and socio-economic consequences of hearing loss are severe (Shield, 2006). In adults the consequences can range from mild impairments in the ability to communicate with others, pretending to hear, avoiding social situations (Thomas and Herbst, 1980; Kerr and Cowie, 1997) and increased stress

(Hasson et al., 2011), to lower productivity, absenteeism, or loss of employment (Shield, 2006; Jung and Bhattacharyya, 2012). Hence, hearing loss cannot be confined solely as an individual problem, but it is rather directly linked to serious social and economic consequences (Mohr et al., 2000). The WHO (2017) underlines the importance of preventing and treating hearing loss as early as possible. For this purpose, mobile self-assessment audiometry represents an essential entry point to prevention and early treatment.

Hearing tests are necessary for the detection of hearing loss and are usually obtained by audiologists in clinical settings where the patient is seated in a sound-attenuating room (booth) during the testing. This procedure is time consuming, personnel-intensive and hence cost-intensive for the health care system. Therefore, audiometry is not implemented as a general assessment across the adult population, but rather performed selectively in individuals who are either suspected to have hearing loss or might be at risk of hearing loss (Davis et al., 1992; Gianopoulos et al., 2002; Shield, 2006). A cost-effective solution is beneficial to provide access for large parts of the population in order to detect early hearing loss. Self-assessment mobile audiometers might represent a possible solution for screening larger populations, provided that valid hearing thresholds are delivered (Margolis and Morgan, 2008). The intended use of the investigated app is to serve as a screening audiometer that could be applied in occupational health or in public institutions under supervision of trained personnel.

Automated audiometry in general has successfully shown to deliver hearing thresholds and several studies have provided evidence on the validity of hearing thresholds by automated audiometry comparable to manual audiometry (Mahomed et al., 2013). To date, the technical feasibility of self-assessment of hearing loss via audiometric applications, implemented on mobile devices, has been demonstrated by several studies (Bright and Pallawela, 2016), but only for a few self-assessment devices (uHear, EarTrumpet, hearTest, ShoeBox audiometry) published validation results are available (Szudek et al., 2012; Foulad et al., 2013; Handzel et al., 2013). Additionally, as an important limitation, only a minority of these studies reported on mobile audiometers with appropriately calibrated components (Thompson et al., 2015; Saliba et al., 2017; van Tonder et al., 2017).

The validity of a freely available iOS-based audiometric application (uHear) for hearing loss has been investigated in several studies (Szudek et al., 2012; Handzel et al., 2013; Khoza-Shangase and Kassner, 2013; Peer and Fagan, 2015; Abu-Ghanem et al., 2016). This device is connected to consumer earbud headphones, which are not calibrated with appropriate reference threshold values. The average difference in hearing thresholds compared to standardized automated audiometry was reported to be quite high: >8 dB HL (decibel hearing level) in a sound booth and >14 dB HL in a waiting room (Szudek et al., 2012). In a sample of elderly participants (>65 years) (Abu-Ghanem et al., 2016) found that the application overestimated hearing thresholds by up to 17 HL. Yet the authors note that it could serve as a useful screening device, especially for people with limited accessibility to audiometric facilities.

A self-assessment, tablet-based hearing test (ShoeBox audiometry) was validated in adults using a calibrated audiometric headphone revealing hearing thresholds within an acceptable range compared to conventional clinical audiometry (Thompson et al., 2015). The authors reported a deviation of hearing thresholds within 10 dB from those obtained by conventional audiometry and concluded that the device serves as a valid screening instrument.

For hearing tests, it is crucial that measurements of hearing levels gathered by different devices are comparable. Audiometric devices differ in their hardware components, i.e., in the tone-generating module and in the type of headphones. Headphones differ in frequency characteristics depending on their specific details of construction. Thus, the use of different types of headphones (circumaural, supra-aural, earbud, or insert headphones) can lead to large variations in amplitude of the test signal depending on their frequency response (Shojaemend and Ayatollahi, 2018). Consequently, inaccuracies introduced by using non-standardized headphones, frequently applied in consumer products, can be resolved by applying calibrated, professional headphones designed for audiometry. Furthermore, adequate calibration requires the devices (sound generator and headphones) to be tested as a couple and calibration cannot be transferred to another device even when it is the same model.

Accurate interpretation of hearing thresholds requires calibration of the audiometric device, as comparability with established hearing tests cannot be ensured otherwise. To comply with this prerequisite a newly developed self-assessment tablet-based hearing test, calibrated with audiometric headphones, was validated and its reliability was tested in the current study. This app was developed for adults as a screening test. To evaluate the accuracy and reliability of the audiometric application we measured hearing-thresholds with the new app and compared these to standardized automated audiometry. For the comparison of thresholds, automated audiometry based on the Hughson-Westlake algorithm was applied, as this ensures a standardized procedure and high test-retest reliability. To answer the question whether differences in hearing thresholds between devices were sufficiently similar, an equivalence test (TOST-P) was computed for all frequencies (Rogers et al., 1993). Furthermore, we assessed the degree of reliability of the new app (and of the automated audiometer for comparisons) using within subject test-retest reliability analyses.

MATERIALS AND METHODS

Study Design

The newly developed audiometric application was compared to standardized automated audiometry using a repeated-measures within-subject design. A conventional audiometer with automatic recording of hearing thresholds standardized according to DIN EN ISO 8253-1 (2011) was chosen as the reference (further denoted as AM). Thresholds were collected at eleven test frequencies ranging from 125 to 8000 Hz (see **Table 2** for details). Test-retest reliability was computed on repeated hearing tests within subjects in both devices.

The sample size was comparable to similar studies testing calibrated mobile audiometric devices (Yeung et al., 2013, 2015; Thompson et al., 2015).

Participants

Sixty-eight participants were recruited either via local advertisements or from the database of the Department of Neurophysiology and Pathophysiology of the University Medical Center Hamburg-Eppendorf. Participants received remuneration for their participation (10 €/hour). Inclusion criteria for the participants were: normal hearing (according to self-report), no hearing aids, and an intact outer auditory canal. Participants were recruited so that the sample evenly covered the age range between 19 and 65 years ($M = 42.1$ years; $SD = 13.74$) and was roughly balanced according to gender (32 females, 36 males). Two participants had to be measured a second time on a separate day as their data was not stored correctly during the initial measurement.

Equipment and Materials

Hearing thresholds were recorded automatically and independently for the two ears using pure-tone air-conduction applying the method of Hughson-Westlake following DIN EN ISO 8253-1 (2011). Hearing thresholds were recorded using (1) a conventional audiometer (AM: MAICO MA25 eIIID, Maico Diagnostics, Eden Prairie, MN, United States) and (2) a self-assessment, tablet-based audiometric application (app: Sonormed, Hamburg, Germany), which was executed on an iPad 4 (Apple Inc., Cupertino, CA, United States), equipped by default with a 32-bit digital-to-analog converter. The use of a 32-bit digital-to-analog converter is necessary to produce an adequate range of sound pressure levels (see **Table 1** for ranges). The two devices are classified with medical class 1 and audiometer class 4. They were both equipped with Sennheiser HDA-280 headphones (Sennheiser, Wedemark, Germany), for which audiometric norms exist (Poulsen and Oakley, 2009). The supra-aural version of the HDA-280 was used in combination with the AM (Poulsen and Oakley, 2009) and the circumaural version of the same headphone (version HDA-280 CL, Parsa et al., 2017) was used with the iPad. The app and the specific combination of tone generator (iPad) and headphones were calibrated according to

the IEC 60645 norm (IEC (60645)-1, 2017) by Audio-Ton GmbH (Hamburg, Germany) (**Supplementary Material**). The AM was calibrated according to the same norm (IEC 60645). The app was developed by Sonormed GmbH (Hamburg, Germany) and its core parts are marketed under the name “Audimatch” as an audiometric device.

Both audiometric tests (app, AM) were conducted in a sound-attenuating room (Acoustair, Moerkapelle, Netherlands), which fulfilled the requirements of DIN EN ISO 8253-1 (2011) for air conduction audiometry. The manual for the setup of the app specifies the following prerequisites: a calm and distraction-free environment, comfortable room temperature, sufficient ventilation, and a rested state of the participant. The app automatically evaluates environmental noise and repeats measurement steps if the noise exceeds a defined level according to ANSI S3.1-1999 (ANSI S3 1-1999 (R2013), 2013). The investigator was outside the room during all tests but able to communicate with the participant via camera and intercom system. To compare test durations of the two devices, the duration of the hearing test with the AM was measured by the investigator using a stopwatch. The duration of the hearing test with the app was measured by the app itself. For both the app and the audiometer, the measurement of duration was started with the beginning of the familiarization phase (after the instruction) and ended with the response to the last tone.

Procedures

The test session started with an explanation of the entire test procedure and written consent by the participants. An inspection of the outer ear and two screening tests for hearing according to Weber and Rinne were performed, before the main testing started.

The audiometric procedure of AM and app was technically similar (differences are listed in **Table 2**). The Hughson-Westlake procedure, which is an established method for automated audiometric tests (Carhart and Jerger, 1959; Rhebergen et al., 2015), was implemented in both devices: The hearing threshold at each frequency was successfully defined and accepted at the lowest level at which the participant detected two out of three ascending trials (DIN EN ISO 8253-1, 2011). The sound level was decreased by 10 dB if a positive response was given and increased by 5 dB if the participant did not indicate to hear the test tone.

To reduce effects of learning, fatigue or attention the order of tests was counterbalanced either starting with the app or the AM. Both tests were conducted twice in alternate order (AM, app, AM, app).

Standard Automated Audiometer - AM

The investigator provided a standardized instruction according to DIN EN ISO 8253-1 (2011). Then the headphones were placed by the investigator. Before the main test started a familiarization procedure at 1 kHz was conducted on the right ear. Following the familiarization, the order of the eleven test frequencies was: 1 to 8 kHz, again 1 kHz and finally down to 125 Hz. Pure tones with a fixed duration of 2 s were used as test tones. Responses were registered via a hand-held response button. Participants were instructed to press the button only when they heard a tone.

TABLE 1 | Specification of the minimum and maximum hearing level, which can be produced by the tone generator of the iPad using the app.

Frequency (Hz)	Lower limit (dB HL)	Upper limit (dB HL)
125	-20	75
250	-20	100
500	-20	110
750	-20	110
1000	-20	110
1500	-20	110
2000	-20	110
3000	-20	110
4000	-20	110
6000	-20	100
8000	-20	90

TABLE 2 | Comparison of audiometric procedures for the app (Audimatch) and the audiometer (MA25 eIII D).

Feature	MA25 eIII D (AM)	Audimatch (app)
Type of tone	Pure tone	Pure tone
Duration of tone	2s	1s
Response method	Signal detection, only “yes” responses	Alternative forced choice, “yes” or “no”
Possibility of response straight after beginning of tone	yes	no
Duration of response window	2s	unlimited, but pop-up window stating < < Choose “no” if you are not sure > > after 3s
Pause after response	2 – 3,6 s (random interval)	0.7s
Familiarization (right ear)	1000 Hz, presentation until double confirmation of threshold	1000 Hz, presentation until double confirmation of threshold
Familiarization level	Start value: 40 dB Level decrease for correct response: 10 dB Level increase for wrong response: 5 dB	Start value: 40 dB Level decrease for correct response: 20 dB Level increase for wrong response: 10 dB
Criterion for hearing-threshold	2 confirmations out of 3 presentations of the same ascending threshold at each frequency (shortened ascending method according to DIN EN ISO 8253-1)	2 confirmations out of 3 presentations of the same ascending threshold at each frequency (shortened ascending method according to DIN EN ISO 8253-1)
Ascending and descending step size (according to Hughson-Westlake algorithm)	5 dB HL up, 10 dB HL down	5 dB HL up, 10 dB HL down
Default intensity when changing frequency	10 dB decrease	20 dB increase
Default intensity when changing ear	30 dB	Hearing threshold from 1000 Hz in the right ear
Frequency [Hz] order right and left ear	1000, 1500, 2000, 3000, 4000, 6000, 8000, 1000, 750, 500, 250, 125, (1000)	1000, 1500, 2000, 3000, 4000, 6000, 8000, 1000, 750, 500, 250, 125
Re-Test at 1000 Hz	Yes	no
Procedure if deviation in re-test > 10 dB HL	If difference at 1000 Hz \leq 5 dB HL, change ear. Otherwise retest all frequencies until difference \leq 5 dB HL.	not applicable
Class of medical device	1	1
Audiometer class	4	4

The tone was interrupted as soon as a response was given. If no response was given within 2 s after tone offset, the next tone was presented. The inter-stimulus interval varied from 2 to 3.6 s. When a threshold was determined for one frequency, the sound pressure level was increased by 10 dB for the next frequency. Before changing to the opposite ear, the hearing threshold at 1 kHz was tested again. If the deviation at 1 kHz was 5 dB or less, the more sensitive threshold was taken. If the threshold deviated more than 5 dB from the previously measured value, the test sequence started again until deviations were not larger than 5 dB. The default intensity of the start frequency at the opposite ear was 30 dB HL.

Mobile Screening Application - App

A standardized instruction was implemented in the app as a startup-wizard, which each participant had to execute. To guarantee voluntary participation, the user could terminate the test at any time. The startup-wizard included the following: a reminder to take a comfortable seating position and to avoid noise, e.g., through unnecessary movements; a reminder to draw full attention to the test. The headphones were placed by the users themselves. The correct left/right positioning of the headphones was tested by presenting test tones to each ear. Instructions on the screen requested the user to take off glasses, head covers, and jewelry that could interfere with the correct placement of the headphones.

A familiarization procedure at 1 kHz was conducted before the main test. At all times pure tones with a fixed duration of 1 s were used as test tones and it was not possible to respond before the

end of the tone. Following the tone, participants had to respond to the question “Did you hear a tone?” with “yes” or “no.” The choice screen initiated the response interval. If the response time lasted longer than 3 s, the following message was presented: “choose *no* if you are not sure.” The inter-stimulus interval was fixed to 0.7 s. When a threshold was determined for one frequency, the sound pressure level was increased by 20 dB for the subsequent frequency. Following the familiarization, the order of the eleven test frequencies was: 1 to 8 kHz, again 1 kHz and finally down to 125 Hz. For the opposite ear the threshold of the first ear at 1 kHz was taken as the starting level.

At the end of the procedure, app test results were presented to the participants in five grades according to the criteria of the WHO (2018): normal hearing (\leq 25 dB HL), mild (26–40 dB HL), moderate (41–60 dB HL), severe hearing loss ($>$ 60 dB HL), for each ear and for four frequencies separately: 500, 1000, 2000, and 4000 Hz. In case of hearing loss of more than 25 dB HL at least in one ear and in one frequency, or in case of any doubts on subjective hearing, the app recommends to consult a hearing health professional.

Total Duration of Audiometric Testings

The total duration of the four audiometric tests was approximately 60–80 min, including the instruction and breaks, which were scheduled after each audiometric test. During the breaks with self-determined duration, participants had the possibility for refreshment. The duration of the breaks did not exceed 5 min. All test sessions were conducted within the timespan from 9 a.m. to 6 p.m. Resulting hearing thresholds were

explained to participants after the complete test-procedure with reference to hearing thresholds appropriate for age and gender according to ISO 7029 (Eyken et al., 2007).

Data Analysis

The main goals of the analysis were (1) to evaluate the accuracy of the app, i.e., the agreement of hearing thresholds recorded by the app compared to the AM, and (2) to assess the reliability of the app, i.e., evaluating the degree of precision of measurements of the app in a test-retest analysis. The first goal was addressed by testing whether differences in hearing thresholds between devices fall within an acceptable range of deviation to be expected for audiometric measurements using an equivalence test and furthermore by calculating the correlation between the audiograms of app and AM thresholds. The second goal was evaluated by test-retest reliability as estimated by correlating hearing thresholds of the same participants and the same device (correlations within app and within AM thresholds).

Data analysis was performed using R and R Studio (R Version 3.3.3, R Studio Version 1.0.153). For the comparison of thresholds, ears of participants were treated as independent units of observation, as hearing loss can occur in one ear independently from the other one. A significance level of $p < 0.05$ was defined for all statistical tests.

Accuracy

We examined the similarity of hearing thresholds obtained with the app and the AM. The most frequently used method to evaluate the accuracy of measurement instruments is the report of average differences and absolute average differences including standard deviations (Mahomed et al., 2013). In addition to these descriptive statistics, an inferential statistical approach was used to test the accuracy of the app. A two one-sided paired t -test procedure (TOST-P, Rogers et al., 1993) was computed as an equivalence test, in order to test whether hearing thresholds obtained with the app and the AM are sufficiently equivalent. The acceptable range of deviation was defined by standards described in ISO 8253. Equivalence tests were developed in social sciences for research questions asking rather for the equivalence of impact of two treatments than for their difference. With the TOST-P procedure two H_0 hypotheses are proposed, assuming that the mean difference of hearing thresholds falls outside a certain *a priori* defined range (Mara and Cribbie, 2012):

$$H_{01}: \mu_1 - \mu_2 \geq \delta \text{ and } H_{02}: \mu_1 - \mu_2 \leq -\delta$$

With μ_1 and μ_2 being the expected mean value in the population, and δ and $-\delta$ being the upper and lower boundary of the accepted range of deviation between two devices, denoted as the interval of equivalence (Mara and Cribbie, 2012). The test will only be significant, if both null hypotheses (H_{01} , H_{02}) are rejected. In this case the H_1 , i.e., the mean value lies within the accepted range of deviation, will be accepted. The critical t -value for the H_{01} is calculated with the mean values of hearing thresholds in the AM

(M1) and the app (M2) according to Mara and Cribbie (2012):

$$t_2 = \frac{M_1 - M_2 - \delta}{\frac{s_{Diff}}{(n-1)}}$$

and t -value for H_{02} according to:

$$t_2 = \frac{M_1 - M_2 - (-\delta)}{\frac{s_{Diff}}{(n-1)}}$$

with s_{Diff} being the standard deviation of the differences and n the sample size. Hence H_{01} would be rejected if $t_1 \leq -t_{\alpha, (n-1)}$ and H_{02} would be rejected if $t_2 \geq t_{1-\alpha, (n-1)}$, respectively, with α being the significance level. As 11 tests (one for each frequency) were performed, testing the hypothesis that the devices did not differ in accuracy, a Bonferroni correction for multiple comparisons was applied by dividing the significance level by the number of tests: $p = 0.05/11 = 0.0045$. The Bonferroni correction reduces the probability of false positives. Of note, with the TOST-P this is the probability that the test erroneously assumes that the two devices do not differ in their thresholds, although they do differ.

The interval of equivalence was defined as the maximally tolerable deviation (δ and $-\delta$) due to the measurement uncertainty according to DIN EN ISO 8253-1 (2011). The calculation of the measurement uncertainty yielded a maximum tolerable deviation of 10 dB HL in all frequencies from 125 to 4 kHz, and 13 dB HL above 4 kHz. Hence a deviation of ± 10 dB HL (and ± 13 dB HL for 6 and 8 kHz) was chosen as the boundary values for the interval of equivalence in the TOST-P. The measurement uncertainty considers that audiometric data varies due to several aspects of the test setting, the participant, the nature of repetition, and expected values close to the normal hearing threshold.

As a second approach to quantify the agreement of the measured values, correlation coefficients between hearing thresholds obtained by the two devices (app and AM) were computed. A Spearman rank correlation was calculated, because hearing thresholds were not normally distributed.

Test-Retest Reliability

The correlation of hearing thresholds obtained with the same device and same participants at different points in time, the test-retest reliability coefficient (r_{tt}), can be considered as the method's degree of precision. It is assumed that the true values of participants do not differ between measurements (Moosbrugger and Kelava, 2012). Again, Spearman rank correlations were calculated, because hearing thresholds were not normally distributed.

Test Duration

To examine a potential difference in test duration between the two devices, the time intervals to complete the respective audiometric measurements were compared and statistical significance of the difference was tested with a two-sided paired t -test. To assess whether the test duration within the app differed

between the first and the second run, a two-sided Wilcoxon signed rank test was computed, as duration differences between t1 and t2 were not normally distributed.

RESULTS

According to standardized automated audiometry, 66 of 68 participants had normal hearing defined by the WHO criteria, i.e., pure-tone average (PTA) of 0.5, 1, 2, 4 kHz: Normal hearing (≤ 25 dB HL), mild (26–40 dB), moderate (41–60 dB), severe (61–80 dB), or profound hearing loss like deafness (> 81 dB) on the better hearing ear. Two participants had mild hearing loss with ~ 35 dB HL on the better hearing ear. This result was obtained identically by the app and the AM. This decision of using the WHO criteria for hearing loss bears the potential disadvantage that unilateral hearing loss, or hearing loss in single frequencies is missing. Further results are very detailed and contain all single frequencies and both ears. The following results are reported for 136 ears of 68 participants.

Accuracy

Descriptive analyses of the mean hearing levels (in dB HL) revealed that thresholds measured with the app were slightly higher than thresholds measured with the AM (Figure 1). This relationship was observed in all frequencies. Mean differences in hearing levels between the app and the AM ranged between -4.12 and -0.48 dB (AM -app) with standard deviations between 3.49 and 6.7 dB across the frequency spectrum, while mean absolute differences ranged between 3.27 and 4.96 dB with standard deviations between 2.93 and 4.7 dB (Table 3).

The distribution of differences between hearing thresholds of the app and the AM across subjects show that at t1 approximately 97% of all differences fall within the range of ± 10 dB HL and at t2: 96% (Figure 2C). These values indicate that most differences lie within the interval of equivalence.

For the statistical evaluation of similarity of the hearing thresholds obtained with the app and the AM, the data was analyzed between devices using the first run (t1) only. The equivalence test revealed that differences in hearing thresholds fall into the interval of equivalence for all frequencies (Figure 3). The smallest difference between the two devices was observed at 3000 Hz ($M_{\text{Diff}} = -0.48$; $SD = 4.63$, $t(135) = 23.984$, $p < 0.001$) and the largest at 1500 Hz ($M_{\text{Diff}} = -4.12$; $SD = 4.71$, $t(135) = 14.566$, $p < 0.001$). Results of the TOST-P for all frequencies can be found in Table 4.

Furthermore, the hearing thresholds of the app were correlated with those of the AM to test the agreement of the hearing thresholds collected by the app. The partial Spearman rank correlation between the hearing thresholds of all frequencies obtained at t1 was $\rho_{\text{APP,AM}} = 0.777$ ($p < 0.001$). A scatter plot of the hearing thresholds indicates a strong association between the two devices (Figure 4C).

Test-Retest Reliability

For an adequate interpretation of the results obtained by the analysis of the validity, the reliability of the app needs

to be quantified explicitly. Therefore test-retest reliability was computed within both devices. A partial Spearman rank correlation, corrected for frequency, resulted in a test-retest correlation coefficient of $\rho_{\text{tt,app}} = 0.829$ ($p < 0.001$) for measurements with the app and $\rho_{\text{tt,AM}} = 0.792$ ($p < 0.001$) for measurements with the AM (Figures 4A,B).

The distribution of differences in thresholds between the first and second measurement obtained with the app revealed that 93% of threshold values differed by 5 dB or less (Figure 2A). With the AM 88% of threshold values differed by 5 dB or less (Figure 2B). A difference of 0 dB was observed for 56% of threshold values when tested with the app, and 47% when tested with the AM.

Hearing thresholds within devices show a strong correlation between the first and the second session, also when considering single frequencies (Table 5). Correlation coefficients were highest in the high frequencies and lowest at 125 Hz within both devices. Test-retest reliability values within frequencies (Table 5) in the app vary from $\rho_{\text{(tt)125}} = 0.728$ ($p < 0.001$) at 125 Hz up to a very high correlation at 8000 Hz $\rho_{\text{(tt)8000}} = 0.919$ ($p < 0.001$). Test-retest reliability values within frequencies in the AM vary from $\rho_{\text{(tt)125}} = 0.614$ ($p < 0.001$) at 125 Hz up to a very high correlation at 8000 Hz $\rho_{\text{(tt)8000}} = 0.902$ ($p < 0.001$).

Test Duration

To assess whether durations to complete the test with the app and the AM are comparable, the duration of each audiometric test was measured. For both devices the time interval included the familiarization phase, but not the instruction. Due to transmission errors, durations of four subjects were lost in the data set of the app. The median of durations at t1 to complete the test with the app was $MD = 12.38$ min (IQR = 2.179) and with the AM it was $MD = 16.51$ min (IQR = 3.45). Minimum and maximum times to complete the test with the app were 10.05 min and 19.73 min and with the AM were 13 and 28.5 min. The time to complete the hearing test with the app was significantly shorter than with the AM at t1: The average difference was $M_{\text{Diff}} = -4.27$ min (95% CI = $[-4.866; -3.676]$), $t_{(63)} = -14.335$, $p < 0.001$.

To assess whether a learning effect had occurred in the use of the app, reflected in a shorter duration of the second testing, a two-sided Wilcoxon signed rank test was computed. The duration with the app at t2 was $MD = 11.98$ min (IQR = 2.566) and with the AM it was $MD = 16.23$ min (IQR = 3.49). Within audiometric tests no significant differences between t1 and t2 durations were observed ($MD = -0.4$ min; IQR = 1.758), $V = 0.713$, $p = 0.096$. For this comparison the duration values of seven participants were lost, due to the transmission error of the app data.

DISCUSSION

The results of the current validation study revealed similar audiometric thresholds for the mobile screening audiometer application compared to standardized automated audiometry. Differences between the audiograms of both devices ranged

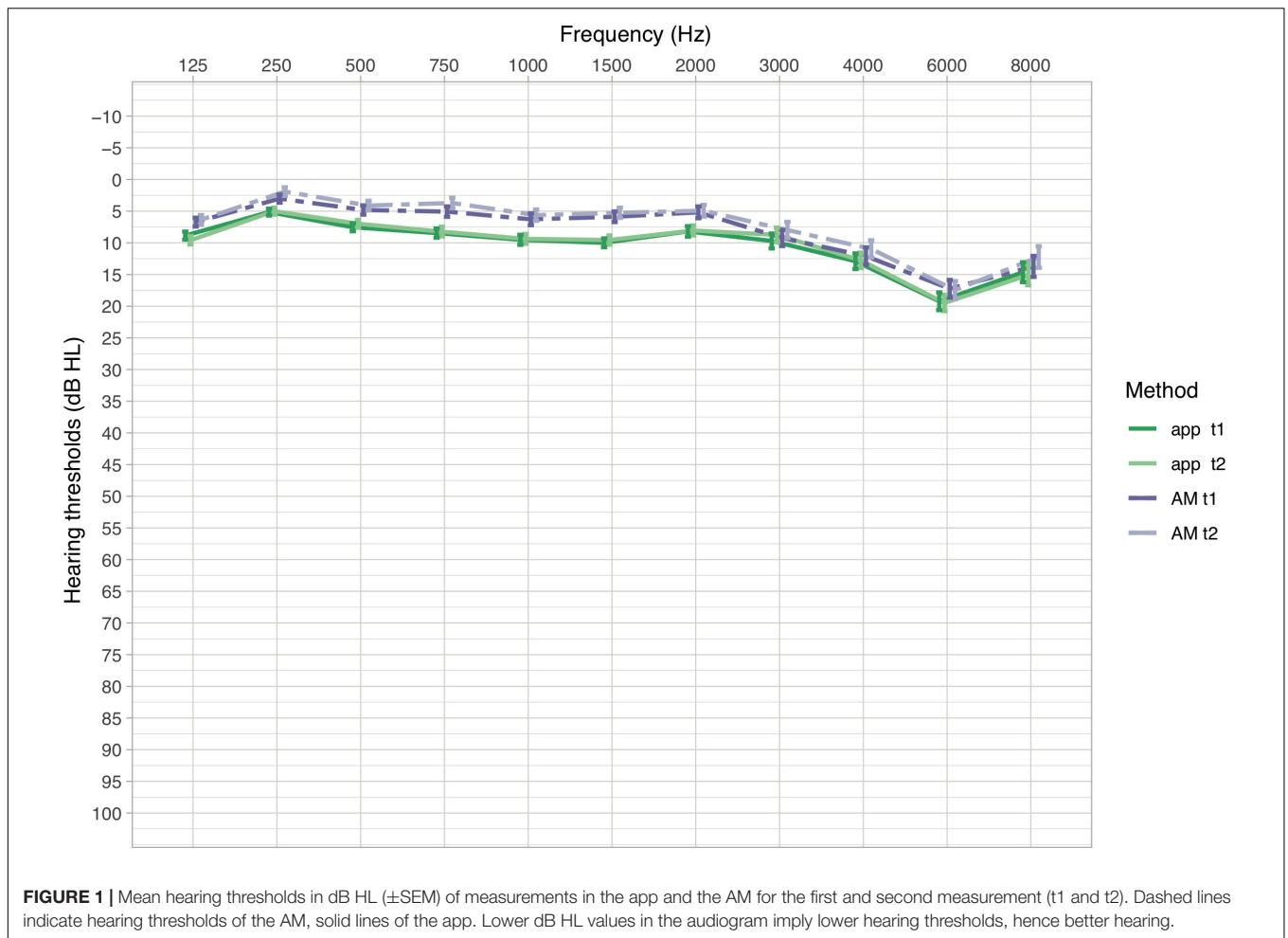


FIGURE 1 | Mean hearing thresholds in dB HL (\pm SEM) of measurements in the app and the AM for the first and second measurement (t1 and t2). Dashed lines indicate hearing thresholds of the AM, solid lines of the app. Lower dB HL values in the audiogram imply lower hearing thresholds, hence better hearing.

within a clinically acceptable range as indicated by an equivalence test. The convergence of both audiograms revealed by each device was demonstrated by the correlation of hearing thresholds. The degree of reliability of the mobile screening audiometer was quantified by a test-retest reliability analysis. The resulting correlation coefficient revealed a high degree of precision for the app and showed similar levels compared to standardized automated audiometry.

Evaluating Accuracy and Reliability

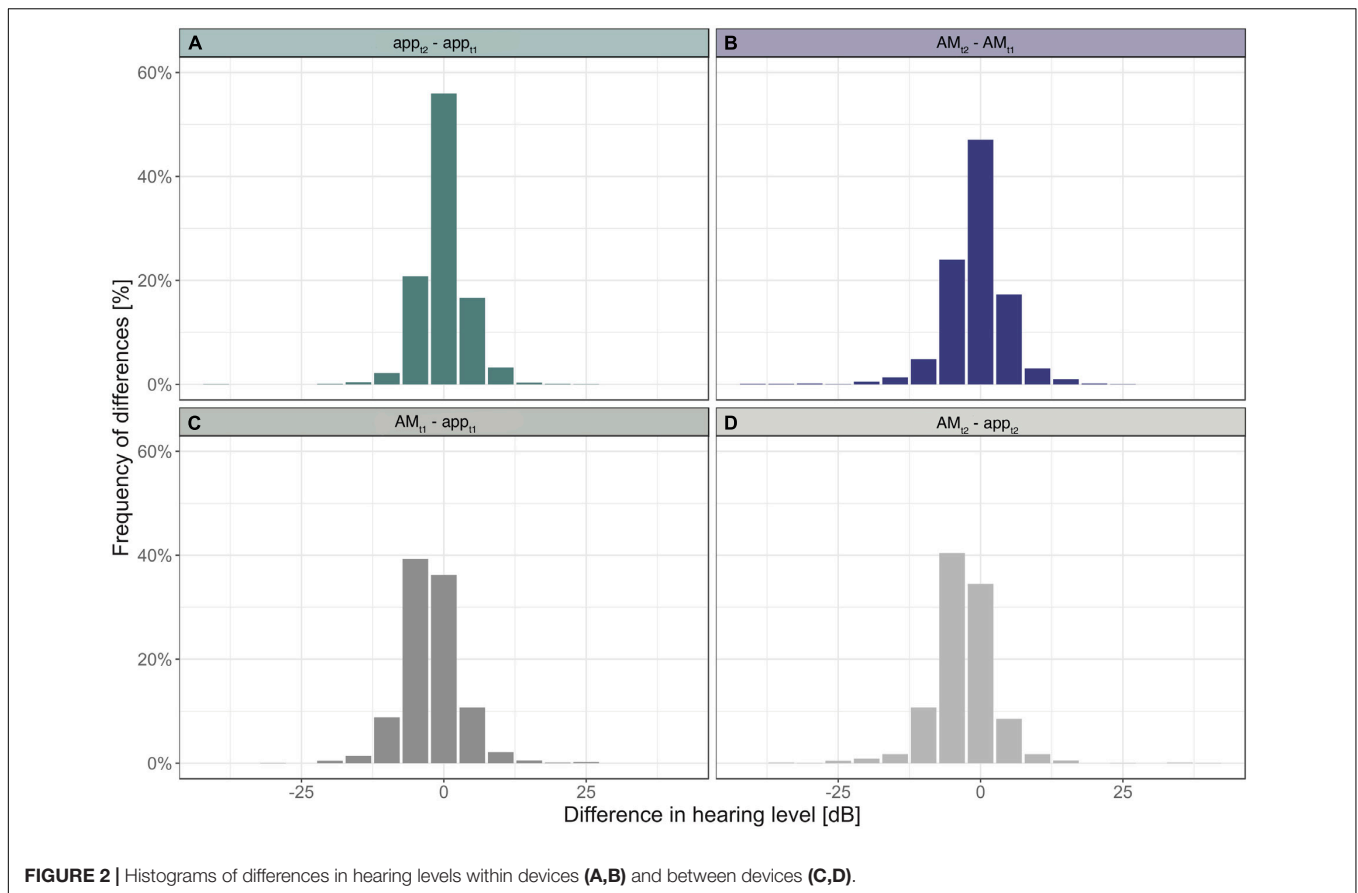
The first aim of the present study was to investigate the validity of the mobile screening audiometer. To that end, we compared the hearing thresholds of the app with hearing thresholds of a standard automated audiometer as used in clinical context. An equivalence test (TOST-P) indicated a good agreement between the auditory thresholds of both devices. The test revealed that mean differences between app and AM are significantly smaller than the limits of uncertainty that are common to audiometric procedures. Therefore, we concluded that the hearing thresholds gathered with the app seem to be within a clinically acceptable range of deviation compared to hearing thresholds obtained with a standard automated audiometer. Agreement of thresholds is also supported by the significant

correlation between hearing thresholds of the two devices. In the current study 97% of detected thresholds were within the clinically acceptable range of difference (10 dB HL) compared to standardized automated audiometry. Compared to other studies investigating the validity of the mobile audiometric applications,

TABLE 3 | Results of comparisons between App and AM: mean differences and mean absolute differences (dB HL) of hearing thresholds, in each frequency across all ears.

Frequency [Hz]	Mean diff. (AM-App) [dB HL]	SD	Mean abs. diff. [dB HL]	SD
125	-2.132	6.135	4.706	4.462
250	-2.132	4.200	3.382	3.271
500	-2.721	3.489	3.309	2.933
750	-3.382	3.746	3.824	3.291
1000	-3.235	4.851	4.338	3.888
1500	-4.118	4.709	4.853	3.941
2000	-3.015	5.453	4.485	4.315
3000	-0.478	4.630	3.272	3.299
4000	-0.809	5.027	3.456	3.728
6000	-1.985	6.033	4.265	4.696
8000	-0.919	6.700	4.963	4.574

SD = standard deviation.



the corresponding percentage of values within the acceptable range in the current study is comparable (Thompson et al., 2015; van Tonder et al., 2017) or even higher (Szudek et al., 2012; Foulad et al., 2013; Khoza-Shangase and Kassner, 2013).

Mobile audiometric devices without calibrated headphones show larger differences in hearing thresholds compared with conventional audiometry (Szudek et al., 2012; Foulad et al., 2013; Khoza-Shangase and Kassner, 2013). Several studies investigating the validity of non-calibrated mobile audiometric devices reveal average differences in hearing thresholds >14 dB compared to conventional audiometry (Khoza-Shangase and Kassner, 2013) and the percentage of thresholds ranging within 10 dB of the conventional audiogram at 67% when measured in a sound booth (Szudek et al., 2012). Foulad et al. (2013) investigated the validity of a self-assessment mobile application (EarTrumpet) with different iOS devices. However, the system used earbud headphones calibrated with a non-standardized procedure. Further studies, which investigated the validity of mobile hearing tests, either tested applications which were not designed for self-assessment (hearScreen and AudCal) (Swanepoel et al., 2014; Larrosa et al., 2015) or did not use calibrated headphones (Kam et al., 2012; Corry et al., 2017). Thus, they are not directly comparable to the audiometric application tested in the current study.

Applications which were tested with calibrated headphones resulted in much higher accuracies than the uncalibrated devices

(Yeung et al., 2013; Thompson et al., 2015; Saliba et al., 2017; van Tonder et al., 2017). In fact, the percentage of thresholds within 10 dB of the corresponding conventional audiogram ranged between 91 and 97%: the result of 97% was reported by Thompson et al. (2015), 94.4% by van Tonder et al. (2017), and 91.1% (EarTrumpet) and 95.5% (Shoebbox audiometry) by Saliba et al. (2017), which all used calibrated devices. These accuracy values are in the same range as the accuracy values of the app under investigation here.

Most mobile audiometric devices suffer from a limited range of sound intensities due to the technical limitations of headphones or the acoustic converter. At the lower boundary of sound intensities other studies report a technical limitation at 10 dB HL (van Tonder et al., 2017) or 15 dB HL (Thompson et al., 2015) leading to floor effects in normal hearing individuals. The app under investigation can produce sound intensities as low as -20 dB HL and up to 110 dB HL (see Table 1), which reduces the probability of floor or ceiling effects.

The second aim of the study was to evaluate the degree of reliability of measurements of the app by performing a test-retest analysis. This analysis revealed a high degree of precision for the app with slightly higher correlation coefficients compared to the audiometer. Only the evaluation of one other computerized and self-assessment hearing test (Kam et al., 2012) was reported to have higher retest-reliability values for the mobile application

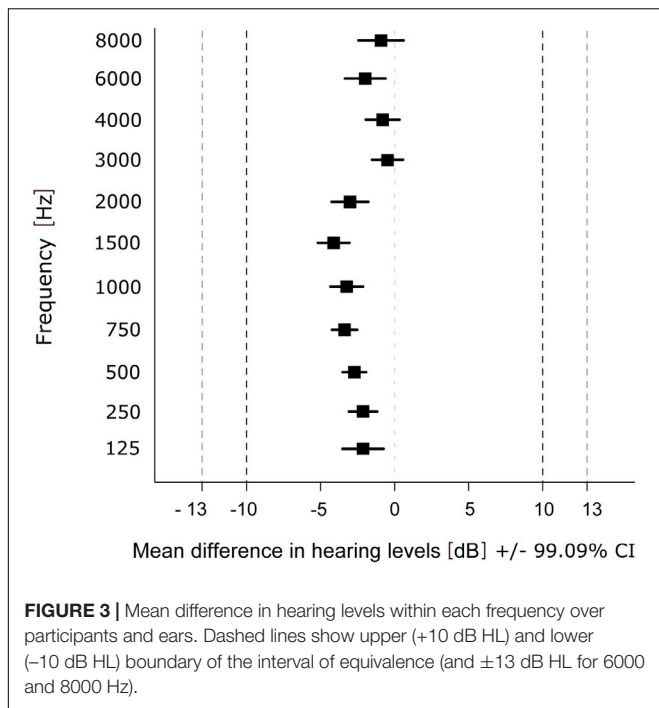


FIGURE 3 | Mean difference in hearing levels within each frequency over participants and ears. Dashed lines show upper (+10 dB HL) and lower (-10 dB HL) boundary of the interval of equivalence (and ±13 dB HL for 6000 and 8000 Hz).

TABLE 4 | Results of comparisons between devices: mean differences (dB HL) of hearing thresholds, *p*-values and 99.09% CI for the TOST-P in each frequency across all ears.

Frequency [Hz]	Mean diff [dB HL]	<i>t</i> (135)	<i>p</i>	99.09% CI
125	-2.132	14.955	<0.001	[-3.525, -0.74]
250	-2.132	21.845	<0.001	[-3.086, 1.179]
500	-2.721	24.331	<0.001	[-3.512, -1.929]
750	-3.382	20.603	<0.001	[-4.233, -2.532]
1000	-3.235	16.263	<0.001	[-4.336, -2.134]
1500	-4.118	14.566	<0.001	[-5.187, -3.049]
2000	-3.015	14.939	<0.001	[-4.252, -1.777]
3000	-0.478	23.984	<0.001	[-1.529, 0.573]
4000	-0.809	21.324	<0.001	[-1.95, -0.332]
6000	-1.985	21.291	<0.001	[-3.355, -0.616]
8000	-0.919	21.028	<0.001	[-2.44, 0.602]

(intra-class correlation = 0.95) than the ones found in the current study.

Taken together, the findings of the present study support the assumption that the app under investigation showed acceptable values of accuracy and high reliability which are in the range of automated clinical audiometers (Mahomed et al., 2013). The device could serve as a valid screening audiometer. These results extend findings of previous research on mobile self-assessment audiometry in terms of reliability and accuracy. Our results further support the necessity that for the self-assessment of hearing thresholds the use of appropriately calibrated devices according to ISO standards is beneficial and helps to obtain levels of accuracy and reliability. The potential areas of usage for mobile screening tests are versatile and include, but certainly are

not constrained to, occupational health care and public health institutions. Mobile audiometric devices could provide a low-threshold access to hearing tests within companies that offer this health-related service to their employees. In addition, such a screening device might be integrated in awareness campaigns addressing hearing health and prevention within occupational health care. Although screening results cannot substitute a detailed audiometric evaluation of a medical specialist, the result can motivate a person to visit a medical specialist in order to receive a diagnostic clarification and professional assessment.

Test Duration

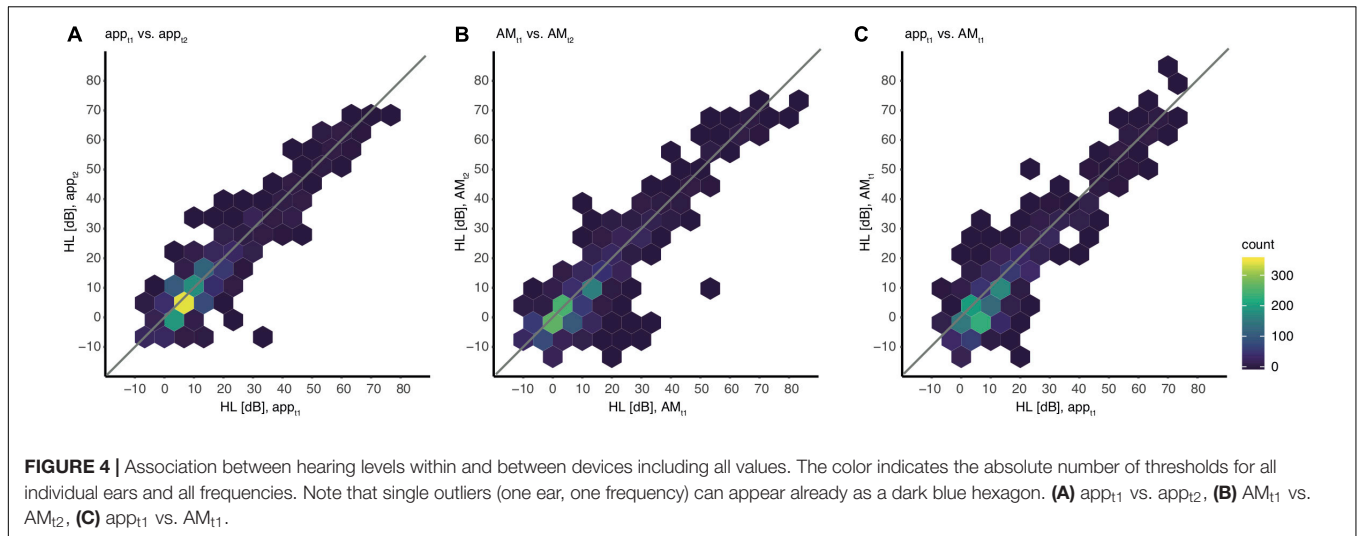
The duration to complete an audiometric test with the app was on average 4.27 min shorter than with the AM, pointing to a potential time-saving advantage for the investigator and reducing task-related fatigue for the participant. Three main features, which could affect the duration of the test, were different between the app and the audiometer: the duration of the test tones, the interval between response and the next tone, and the type of question the participants had to answer (press a button only when a tone is present vs. press different buttons for tone present/not present). These three differences could have led to faster detection of thresholds without strongly affecting the validity of the new audiometric device, as differences between both devices were within the limits (97% of thresholds were within 10 dB difference).

A learning effect for the execution of the app was not detected, as test durations in the first compared to the second run were similar. Similar durations for both runs also indicate that the instructions for the app were sufficiently well understood before the first run. Also, an effect of fatigue seems unlikely, given similar test durations at the beginning and the end of the whole testing session.

Limitations of the Study

Test sessions were conducted in a sound-insulated booth and it is therefore not evident whether results can be compared to the measurement of hearing thresholds in a noisy surrounding, like in a standard office. Thus, field studies with environmental noise could provide more insight on the accuracy and validity of the audiometric thresholds gathered by the app, for example in a waiting room of an otolaryngologist. Importantly, a validation with audiologicaly impaired patients would be necessary for the estimation of sensitivity and specificity of the app for clinically relevant hearing loss.

The audiometric application was designed for self-assessment. However, even though the whole audiometric procedure can be performed by the participant, the system is not intended for the use in private homes, outside the range of a trained person, as special headphones are needed and regular calibration of the iPad/headphone combination is a prerequisite. Apart from the calibration of the system, which has to be performed by a specialized company, this audiometric screening test can be operated by the user. Some supervision by trained personnel is helpful when starting the app, but it does not require the guidance of a health care professional. This stands also in contrast to the operation of the automated audiometer (which was used for



comparison) for which the placement of the headphones and the instruction of the participants had to be done by a trained person.

A further limitation of the study is concerning the sampling method. Gender and age range are well-known factors which influence hearing thresholds (Vaz et al., 2013). To avoid biases due to an overproportionate representation of these specific attributes, we used a sampling method, which allowed to collect data from a more representative sample than simple random sampling. This method accepts the consequence that the recruiting is not completely random. The proportion of males

and females in the sample was balanced, and age was uniformly sampled across the age range of the study.

Two authors of the current article were involved in the development of the app-based mobile hearing test which was evaluated in this study. This fact was disclosed before the start of the study so that a potential influence on the design of the study, data collection or the rational of data analysis could be contained beforehand.

CONCLUSION

Given the experimental settings of the present study, i.e., measuring audiometric thresholds in normal hearing participants in a sound-attenuated room, the results provide first evidence that the mobile self-assessment screening audiometer with calibrated headphones is measuring hearing thresholds, which are comparable to those obtained by a standard automated audiometer. Thus, the app could serve as a valid screening audiometer. Future studies could explore the validity of the app for patients with hearing loss or for children.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This research was conducted in conformity to the regulations and ethical standards of the Hamburg Medical Association (Ärztchamber Hamburg). The ethical review board of the Hamburg Medical Association waived the requirement for ethical approval for this study after screening the study protocol given that the experimental execution and data handling did not meet the criteria for an obligatory ethical review according

TABLE 5 | Results of test-retest reliability analyses within device and test frequency.

Frequency [Hz]	Device	$\rho_{01,12}$	p	95% CI
125	app	0.728	<0.001	[0.638; 0.799]
	AM	0.614	<0.001	[0.497; 0.709]
250	app	0.747	<0.001	[0.663; 0.814]
	AM	0.74	<0.001	[0.652; 0.807]
500	app	0.831	<0.001	[0.770; 0.877]
	AM	0.781	<0.001	[0.706; 0.839]
750	app	0.763	<0.001	[0.683; 0.826]
	AM	0.798	<0.001	[0.728; 0.852]
1000	app	0.75	<0.001	[0.665; 0.815]
	AM	0.719	<0.001	[0.627; 0.792]
1500	app	0.750	<0.001	[0.666; 0.815]
	AM	0.736	<0.001	[0.647; 0.804]
2000	app	0.84	<0.001	[0.782; 0.883]
	AM	0.754	<0.001	[0.670; 0.818]
3000	app	0.843	<0.001	[0.786; 0.886]
	AM	0.853	<0.001	[0.799; 0.893]
4000	app	0.899	<0.001	[0.861; 0.927]
	AM	0.868	<0.001	[0.82; 0.904]
6000	app	0.857	<0.001	[0.805; 0.896]
	AM	0.825	<0.001	[0.763; 0.872]
8000	app	0.919	<0.001	[0.888; 0.941]
	AM	0.902	<0.001	[0.865; 0.929]

to national guidelines and local legislation. Data storage and analysis were fully anonymized. Written informed consent was collected by all participants before participation according to the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

TS and GS conceived and designed the study. AC performed the testing. AC, TS, and GS analyzed the data. JN contributed materials or analysis tools. AC and TS wrote the manuscript. All authors discussed the results, edited, and approved the manuscript.

FUNDING

This work was supported by a grant from the Ministry of Health and Consumer Protection of the Free and Hanseatic City of

Hamburg (Behörde für Gesundheit und Verbraucherschutz der Freien und Hansestadt Hamburg, grant no.: Z|74779|2016|G500-02.10/10,010).

ACKNOWLEDGMENTS

The authors would like to thank Karin Deazle for assistance with participant recruitment, Maja Glahn for data acquisition and documentation, and Adrian Nötzel for providing the calibration procedure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00744/full#supplementary-material>

REFERENCES

- Abu-Ghanem, S., Handzel, O., Ness, L., Ben-Artzi-Blima, M., Fait-Ghelbendorf, K., and Himmelfarb, M. (2016). Smartphone-based audiometric test for screening hearing loss in the elderly. *Eur. Arch. Otorhinolaryngol.* 273, 333–339. doi: 10.1007/s00405-015-3533-9
- ANSI S3 1-1999 (R2013), (2013). *Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms*. New York, NY: American National Standards Institute.
- Bright, T., and Pallawala, D. (2016). Validated smartphone-based apps for ear and hearing assessments: a review. *JMIR Rehabil. Assist. Technol.* 3:e13. doi: 10.2196/rehab.6074
- Carhart, R., and Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *J. Speech Hear Disord.* 24, 330–345. doi: 10.1044/jshd.2404.330
- Corry, M., Sanders, M., and Searchfield, G. D. (2017). The accuracy and reliability of an app-based audiometer using consumer headphones: pure tone audiometry in a normal hearing group. *Int. J. Audiol.* 56, 706–710. doi: 10.1080/14992027.2017.1321791
- Davis, A., Stephens, D., Rayment, A., and Thomas, K. (1992). Hearing impairments in middle age: the acceptability, benefit and cost of detection (ABCD). *Br. J. Audiol.* 26, 1–14. doi: 10.3109/03005369209077866
- DIN EN ISO 8253-1, (2011). *DIN EN ISO 8253-1:2011-04 Acoustics – Audiometric test Methods – Part 1: Pure-Tone Air and Bone Conduction Audiometry (ISO 8253-1:2010); German version EN ISO 8253-1*. doi: 10.3109/03005369209077866
- Eyken, E. V., Camp, G. V., and Laer, L. V. (2007). The complexity of age-related hearing impairment: contributing environmental and genetic factors. *Audiol. Neurotol.* 12, 345–358. doi: 10.1159/000106478
- Foulad, A., Bui, P., and Djalilian, H. (2013). Automated audiometry using apple iOS-based application technology. *Otolaryngol.-Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.* 149, 700–706. doi: 10.1177/019459813501461
- Gianopoulos, I., Stephens, D., and Davis, A. (2002). Follow up of people fitted with hearing aids after adult hearing screening: the need for support after fitting. *BMJ* 325:471. doi: 10.1136/bmj.325.7362.471
- Handzel, O., Ben-Ari, O., Damian, D., Priel, M. M., Cohen, J., and Himmelfarb, M. (2013). Smartphone-based hearing test as an aid in the initial evaluation of unilateral sudden sensorineural hearing loss. *Audiol. Neurotol.* 18, 201–207. doi: 10.1159/000349913
- Hasson, D., Theorell, T., Wallén, M. B., Leineweber, C., and Canlon, B. (2011). Stress and prevalence of hearing problems in the Swedish working population. *BMC Public Health* 11:130. doi: 10.1186/1471-2458-11-130
- IEC 60645-1, (2017). *IEC 60645-1:2017 Electroacoustics – Audiometric Equipment – Part 1: Equipment for Pure-tone and Speech Audiometry*. Available at: <https://webstore.iec.ch/publication/32370> (accessed March 9, 2018).
- Jung, D., and Bhattacharyya, N. (2012). Association of hearing loss with decreased employment and income among adults in the United States. *Ann. Otol. Rhinol. Laryngol.* 121, 771–775. doi: 10.1177/000348941212101201
- Kam, A. C. S., Sung, J. K. K., Lee, T., Wong, T. K. C., and van Hasselt, A. (2012). Clinical evaluation of a computerized self-administered hearing test. *Int. J. Audiol.* 51, 606–610. doi: 10.3109/14992027.2012.688144
- Kerr, P. C., and Cowie, R. I. (1997). Acquired deafness: a multi-dimensional experience. *Br. J. Audiol.* 31, 177–188. doi: 10.3109/0300536400000020
- Khoza-Shangase, K., and Kassner, L. (2013). Automated screening audiometry in the digital age: exploring uhearTM and its use in a resource-stricken developing country. *Int. J. Technol. Assess. Health Care* 29, 42–47. doi: 10.1017/S02664623120000761
- Larrosa, F., Rama-Lopez, J., Benitez, J., Morales, J. M., Martinez, A., Alañon, M. A., et al. (2015). Development and evaluation of an audiology app for iPhone/iPad mobile devices. *Acta Otolaryngol. (Stockh)* 135, 1119–1127. doi: 10.3109/00016489.2015.1063786
- Mahomed, F., Swanepoel, D. W., Eikelboom, R. H., and Soer, M. (2013). Validity of automated threshold audiometry: a systematic review and meta-analysis. *Ear Hear* 34, 745–752. doi: 10.1097/01.aud.0000436255.53747.a4
- Mara, C. A., and Cribbie, R. A. (2012). Paired-samples tests of equivalence. *Commun. Stat. – Simul. Comput.* 41, 1928–1943. doi: 10.1080/03610918.2011.626545
- Margolis, R. H., and Morgan, D. E. (2008). Automated pure-tone audiometry: an analysis of capacity, need, and benefit. *Am. J. Audiol.* 17, 109–113. doi: 10.1044/1059-0889(2008/07-0047)
- Mohr, P. E., Feldman, J. J., Dunbar, J. L., McConkey-Robbins, A., Niparko, J. K., Rittenhouse, R. K., et al. (2000). The societal costs of severe to profound hearing loss in the United States. *Int. J. Technol. Assess. Health Care* 16, 1120–1135. doi: 10.1017/s0266462300103162
- Moosbrugger, H., and Kelava, A. (eds) (2012). *Testtheorie und Fragebogenkonstruktion*, 2nd ed. Berlin: Springer-Verlag.**
- Paras, V., Folkeard, P., and Hawkins, M. (2017). *Evaluation of the Sennheiser HDA 280-CL headphones*. London: National Centre for Audiology, University of Western Ontario.
- Peer, S., and Fagan, J. J. (2015). Hearing loss in the developing world: evaluating the iPhone mobile device as a screening tool. *S. Afr. Med. J. Suid.-Afr. Tydskrif Geneeskde* 105, 35–39.
- Poulsen, T., and Oakley, S. (2009). Equivalent threshold sound pressure levels (ETSPL) for Sennheiser HDA 280 supra-aural audiometric earphones in the

- frequency range 125 Hz to 8000 Hz. *Int. J. Audiol.* 48, 271–276. doi: 10.1080/14992020902788982
- Rhebergen, K. S., van Esch, T. E. M., and Dreschler, W. A. (2015). Measuring temporal resolution (release of masking) with a hughson-westlake up-down instead of a békésy-tracking procedure. *J. Am. Acad. Audiol.* 26, 563–571. doi: 10.3766/jaaa.14087
- Rogers, J. L., Howard, K. L., and Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychol. Bull.* 113, 553–565. doi: 10.1037/0033-2909.113.3.553
- Saliba, J., Al-Reefi, M., Carriere, J. S., Verma, N., Provencal, C., and Rappaport, J. M. (2017). Accuracy of mobile-based audiometry in the evaluation of hearing loss in quiet and noisy environments. *Otolaryngol.-Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.* 156, 706–711. doi: 10.1177/0194599816683663
- Shield, B. (2006). *Evaluation of the Social and Economic Costs of Hearing Impairment*. Available at: https://www.hear-it.org/sites/default/files/multimedia/documents/Hear_It_Report_October_2006.pdf (accessed May 2, 2018).
- Shojaeemend, H., and Ayatollahi, H. (2018). Automated audiometry: a review of the implementation and evaluation methods. *Healthc. Inform. Res.* 24, 263–275. doi: 10.4258/hir.2018.24.4.263
- Swanepoel, D. W., Myburgh, H. C., Howe, D. M., Mahomed, F., and Eikelboom, R. H. (2014). Smartphone hearing screening with integrated quality control and data management. *Int. J. Audiol.* 53, 841–849. doi: 10.3109/14992027.2014.920965
- Szudek, J., Ostevik, A., Dziegielewska, P., Robinson-Anagor, J., Gomaa, N., Hodgetts, B., et al. (2012). Can uhear me now? Validation of an iPad-based hearing loss screening test. *J. Otolaryngol. - Head. Neck Surg.* 41, S78–S84.
- Thomas, A., and Herbst, K. G. (1980). Social and psychological implications of acquired deafness for adults of employment age. *Br. J. Audiol.* 14, 76–85. doi: 10.3109/03005368009078906
- Thompson, G. P., Sladen, D. P., Borst, B. J. H., and Still, O. L. (2015). Accuracy of a tablet audiometer for measuring behavioral hearing thresholds in a clinical population. *Otolaryngol.-Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.* 153, 838–842. doi: 10.1177/0194599815593737
- van Tonder, J., Swanepoel, D. W., Mahomed-Asmail, F., Myburgh, H., and Eikelboom, R. H. (2017). Automated smartphone threshold audiometry: validity and time efficiency. *J. Am. Acad. Audiol.* 28, 200–208. doi: 10.3766/jaaa.16002
- Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., and Andreou, P. (2013). The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS ONE* 8:e73990. doi: 10.1371/journal.pone.0073990
- WHO (2017). *Deafness and Hearing Loss*. Available at: <http://www.who.int/mediacentre/factsheets/fs300/en/> (accessed November 28, 2017).
- WHO (2018). *Grades of Hearing Impairment*. Geneva: WHO.
- Yeung, J., Javidnia, H., Heley, S., Beaugard, Y., Champagne, S., and Bromwich, M. (2013). The new age of play audiometry: prospective validation testing of an iPad-based play audiometer. *J. Otolaryngol. - Head Neck Surg.* 42:21. doi: 10.1186/1916-0216-42-21
- Yeung, J. C., Heley, S., Beaugard, Y., Champagne, S., and Bromwich, M. A. (2015). Self-administered hearing loss screening using an interactive, tablet play audiometer with ear bud headphones. *Int. J. Pediatr. Otorhinolaryngol.* 79, 1248–1252. doi: 10.1016/j.ijporl.2015.05.021

Conflict of Interest: GS and JN received reimbursements or salary from Sonormed GmbH (Hamburg, Germany) to participate in the development of the app-based hearing test evaluated in the current study. JN received salary from Mindo Software SLU (Barcelona, Spain). The company was involved in the software development of the app.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Colsman, Supp, Neumann and Schneider. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.