Yağcıoğlu and Öztürk
The Egyptian Journal of Otolaryngology (2023) 39:148
https://doi.org/10.1186/s43163-023-00513-0

The Egyptian Journal
of Otolaryngology

ORIGINAL ARTICLE **Open Access**

Otoacoustic emission measurements: a test–retest reliability study

Ayşenur Aykul Yağcıoğlu[1*] and Burak Öztürk[2]

Abstract

Objectives Otoacoustic emissions (OAEs) are an important part of the audiological test battery and have many clinical uses. This study aims to determine the amplitude changes in the test–retest condition of distortion product otoacoustic emissions (DPOAEs) and transient-evoked otoacoustic emissions (TEOAEs), which are widely used in clinical settings.

Design DPOAE and TEOAE measurements were taken in 110 ears of 55 adults aged 18–35 years with normal hearing during three sessions. The repeatability of the measurements was evaluated by very short-term measurements taken 20 min after the first measurement and by short-term measurements taken 20 days after the first measurement.

Results There was no statistically significant difference between the three measurements in which DPOAE and TEOAE amplitudes were evaluated. The weakest reliability for TEOAEs was determined at frequencies of 1.0 kHz and 1.5 kHz, and the weakest reliability for DPOAEs was determined at 6728 Hz.

Conclusions The current findings indicate that DPOAE and TEOAE measurements are reliable for monitoring cochlear function over time. The data obtained from this study could help clinicians correctly interpret OAE changes and distinguish between physiological and pathological changes.

Keywords Otoacoustic emission, Cochlear function, Reliability, DPOAE, TEOAE

Background

Otoacoustic emissions (OAEs) are low-intensity sounds coming from the cochlea that can be measured in the outer ear canal. They can be recorded with or without auditory stimulation (evoked or spontaneous, respectively). Kemp (1978) was the first investigator to provide evidence that acoustic emissions can be detected in the human ear canal with clicks and short-duration tone bursts [1, 2]. The OAE test is an important part of the audiological test battery and has many clinical uses. Clinical applications of the OAE test include assessing cochlear integrity, neonatal hearing screening, pediatric evaluation, monitoring cochlear function, helping in ruling out cochlear—retrocochlear pathologies, and evaluating patients with functional or nonorganic hearing loss [2, 3]. The advantages of OAE measurement include that it does not require a behavioral response, it is an appropriate measurement for all ages, it provides a preneural auditory response, and it provides information specific to the ear and the frequency. The inability to make a hearing threshold estimation or evaluate the neural pathways is among the disadvantages [4]. OAE measurements can be measured using two different methods: spontaneous OAEs (SOAEs) and evoked OAEs (EOAEs). EOAEs occur in response to a stimulus presented to the ear. EOAEs occur in response to a pure-tone stimulus presented to the ear, and three different methods are used to reveal the response: stimulus frequency otoacoustic

---

This study was produced from the master thesis prepared by the first author under the supervision of the second author.

*Correspondence:
Ayşenur Aykul Yağcıoğlu
aysenurraykull@gmail.com
[1] Department of Audiology, KTO Karatay University, Akabe, Alaaddin Kap Cd. No:130, Karatay/Konya, Turkey
[2] Department of Audiology, Gazi Mustafa Kemal, Bakırcay University, Kaynaklar Cd, Menemen/İzmir, Turkey

emissions (SFOAEs), transient-evoked otoacoustic emissions (TEOAEs), and distortion product otoacoustic emissions (DPOAEs) [5, 6].

DPOAE measurements are divided into two functions: DP-gram and DPOAE input/output function [6]. The most commonly used measurement is the DP-gram obtained as a function of frequency by displaying the DPOAE amplitude at different frequencies for constant intensity levels. One of the most important measures of the value of audiological testing procedures or clinical equipment is the reliability and repeatability of the test results. To evaluate the repeatability of the procedure, it is necessary to know test–retest repeatability [7]. In the literature, the repeatability of DPOAEs and TEOAEs was generally evaluated with the Otodynamics ILO system OAE instrument [7–14] and the GSI-60 OAE instrument [15–17]. However, many other systems are used clinically today, and there is little data in the literature about their performance. In studies conducted in the literature, the test–retest reliability and repeatability of DPOAE and TEOAE measurements were evaluated at various stimulus intensities and frequencies for minutes, days, and weeks. Generally, high test–retest reliability and repeatability have been achieved in OAEs [7, 9, 10, 18]. However, the measurement repeatability of devices with different software has differed significantly.

In the present research study, the repeatability of DPOAE and TEOAE measurements, which are widely used clinically and provide objective information about the functional integrity of the cochlea, was evaluated by using Otometrics Madsen Capella^2 OAE instrument with OTOsuite software, which is different from the devices used in the literature. In this study, it is aimed to determine the test–retest repeatability and reliability of DPOAE and TEOAE with very short-term measurements made 20 min after the first measurement and short-term measurements made 20 days after the first measurement.

Methods

Subjects

The study was conducted in the Department of Audiology with the approval of the Ethics Committee of the University's Faculty of Medicine (Decision no: 2021/015). A total of 110 ears were included in the study out of 55 people aged 18–35 with normal hearing, 28 of whom were males (mean age 21,678 ± 3.277) and 27 were females (mean age 22,259 ± 4.128). Inclusion criteria of the study were bilateral normal hearing thresholds from 0.25 through 8.0 kHz ≤ 15 dB in pure-tone air conduction evaluation; speech discrimination score of 88% or better; normal otoscopic findings; a type A tympanogram, performed with an 85 dB SPL probe tone at 226 Hz (Titan Wideband Tympanometer, Interacoustic Inc.); middle-ear pressure varying between –100 and + 50 daPa; present ipsilateral acoustic stapedial reflex thresholds between 80 and 100 dB HL at 500 Hz, 1000 Hz, and 2000 Hz (Titan Wideband Tympanometer, Interacoustic Inc.); and correlation ≥ 80% in evaluation with otoacoustic emission test, which was necessary for inclusion in the study.

Individuals with a history of ototoxic drug use; a history of otological disease or surgery; and a history of additional disabilities, chemotherapy, radiotherapy history, neurological diagnosis, hearing loss, neuropsychiatric problem, head trauma, or noise exposure were excluded from the study. Participants were selected voluntarily, and informed consent forms were obtained from all individuals who agreed to participate in the study following the statements of the Declaration of Helsinki. After taking detailed anamnesis including demographic information of the participants, otoscopic examination, acoustic immitansmetric evaluation, pure tone audiometry, speech audiometry, and otoacoustic emission (TEOAE and DPOAE) measurements were performed.

Acoustic immitansmetric evaluation

Individuals with an abnormal external ear canal and/or tympanic membrane were determined by otoscopic examination and were not included in the study. Acoustic immitansmetric measurements were performed on individuals who made otoscopic examinations. Tympanograms in the range of + 200 and − 400 daPa were obtained at 85 dB SPL stimulus intensity by using a 226-Hz probe tone with an interacoustic brand TITAN broadband tympanometer device. With tympanometric evaluation, middle ear pressure (daPa), compliance (mmho), and equivalent external ear canal volume (cc) were evaluated. An acoustic reflex test was performed on those whose tympanogram was normal.

Pure-tone audiometry evaluation

Audiological evaluations were made in the soundproof cabin, which is in the Industrial Acoustics Company (IAC) standard for pure tone audiometry testing. The air-conduction pure tone hearing thresholds of the individuals included in the study were evaluated at frequencies of 250–8000 Hz using Interacoustics brand AC-40 model clinical audiometer and Telephonics brand TDH-39 supra-aural headphones. Bone conduction pure tone hearing thresholds were determined using Radioear brand B-71 bone vibrator at frequencies of 500–4000 Hz. Air and bone conduction pure tone hearing thresholds were determined using the Hughson-Westlake procedure. Individuals with pure tone hearing thresholds of ≤ 15 dB at frequencies of 250–8000 Hz were included in the study. Speech reception threshold and speech discrimination tests were administered to the individuals included in the study.

Otoacoustic emission evaluation

Before the appropriate probe was placed in the ear, the external ear canal was checked and the appropriate disposable probe was placed in the external ear canal. Otoacoustic emission assessments of all individuals included in the study were performed using the Otometrics Madsen Capella[2] portable otoacoustic emission device and OTOsuite software. Measurement was started after the device was in the appropriate measurement position with the appropriate configuration of the probe indicator and stimulus waveform in the device. Because OAE results were obtained independently of each other in both ears of the same individual, the right and left ears of all individuals were evaluated separately. Signal-to-noise ratio (SNR) values were determined for each frequency, and SNRs less than 6 dB were excluded from the study as criteria to verify sufficiently low noise levels. In the evaluation of the test, SNR and amplitude values in dB were determined. DPOAE measurement was made with DP-gram at 996 Hz, 1191 Hz, 1416 Hz, 1679 Hz, 2001 Hz, 2382 Hz, 2832 Hz, 3359 Hz, 4003 Hz, 4755 Hz, 5654 Hz, 6728 Hz, and 7998 Hz f2 frequencies. The difference between the DP-gram test protocol L1–L2 levels was determined to be 10 dB SPL (L1 = 65 dB SPL, L2 = 55 dB SPL), and the f2/f1 ratio was 1.22 in all measurements. DPOAEs were measured with the microphone in the external ear canal at the frequency 2f1-f2. TEOAE measurements were applied at the frequencies of 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, and 4 kHz in the range of 1–4 kHz with 50/s stimulus rate in nonlinear polarity with 85 dB SPL intensity. Registration of TEOAEs was terminated after 512 accepted sweeps with a noise rejection setting of 47.0 dB SPL.

Statistical analyses

All statistical analyses were performed using SPSS version 25.0 (SPSS Inc.). Descriptive statistics are presented as mean, standard deviation, and standard error values. The homogeneity of the variances, which is one of the prerequisites of the parametric tests, was checked using the “Levene” test. The assumption of normality was examined by the “Shapiro–Wilk” test.

A one- or two-way repeated measures analysis of variance (repeated measures ANOVA) was used to determine the changes in otoacoustic emission amplitudes between measurements according to the number of categorical variables. When significant differences were detected between the groups, these differences were determined by Bonferroni correction. For the gender effect repeated tests, the sphericity assumption was checked with Mauchly’s test, and when the sphericity assumption was encountered, the sphericity assumed test was applied. In cases where sphericity was not provided, the Huynh–Feldt test was evaluated for cases where it was greater than 0.75, and the Greenhouse Geisser test results were evaluated for cases where it was smaller by looking at the epsilon value. A p value less than 0.05 was considered statistically significant. The repeatability of OAE measurements was evaluated by the standard error of measurement (SEM), intraclass correlation coefficient (ICC), absolute differences between measurements, 95% confidence intervals, and repeatability standard deviations. The ICC was used to measure the degree to which each participant’s results are similar to those of the same participant as measured at other times [19]. SEM was used to assess whether two OAE measurements were statistically significant. The formula SEM = SS√(1 -ICC) was used to calculate the SEM [20]. In this formula, “SS” represents the standard deviation of all measurements. In this study, ICC values less than 0.5 were assumed to indicate poor reliability, between 0.5 and 0.75 moderate reliability, between 0.75 and 0.90 good reliability, and greater than 0.90 excellent reliability [21].

Results

The repeatability of TEOAEs and DPOAEs was measured in three sessions. The first session consists of the baseline measurement (measurement1 = M1). The second session was performed with probe-refitting (measure2 = M2) after 20 min. This session was followed by the third session (measure3 = M3) which was after 20 days. The mean values of the DPOAE amplitude values of the frequency variables taken from the test and retest conditions are shown in Fig. 1. When the findings are examined, it is seen that the maximum amplitude is at the frequency of 1416 Hz and the minimum amplitude is at the frequency of 7998 Hz. As a result of all measurements, the investigation of whether there was a significant difference between the mean of the differences of the obtained amplitude values in the test and retest situations was carried out with repeated measures ANOVA. There was no statistically significant difference between the three measurements in which DPOAE amplitudes were evaluated ($F = 1.712 = \rho = 0.019$). In 13 different f2 frequency variables, there was a statistically significant difference between the amplitude means obtained in the test–retest conditions, only between the second and third measurements at a frequency of 6728 Hz ($\rho = 0.022$). Figure 2 shows the DPOAE amplitude repeatability in the f2 frequency variable by gender. When the DPOAE amplitude values were examined by gender, higher amplitudes were observed in males at low frequencies of up to 1679 Hz, while higher amplitude values were observed in females

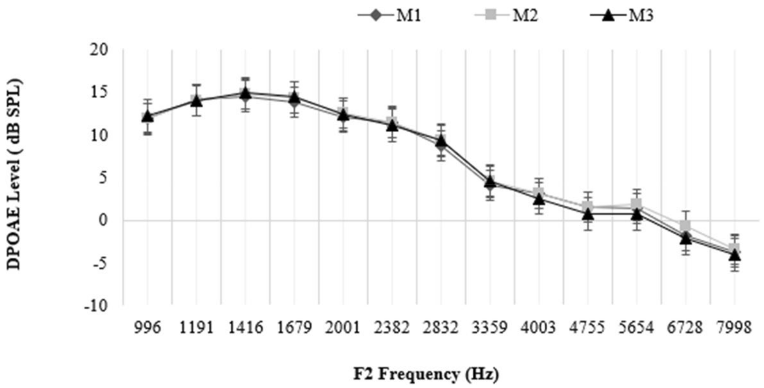

Fig. 1 The vertical lines show the distortion level in dB SPL and the horizontal lines the f2 frequency in Hz. DPOAE indicates distortion product otoacoustic emission.The first session consists of the baseline measurement (measurement1 = M1). The second session (measure2 = M2) after 20 min. Third session (measure3 = M3) after 20 days. The figure shows mean DPOAE amplitude of the frequency variables from the test and retest conditions

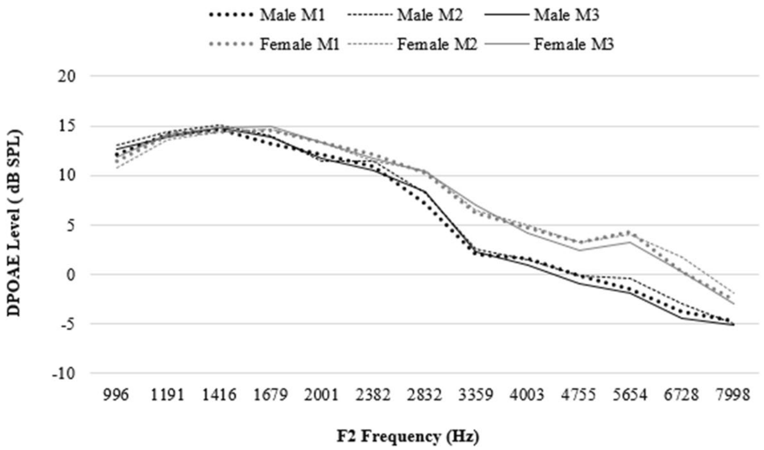

Fig. 2 The vertical lines show the distortion level in dB SPL and the horizontal lines the f2 frequency in Hz. DPOAE indicates distortion product otoacoustic emission. The figure shows mean DPOAE amplitudes repeatability in the f2 frequency variable by gender

at 1679 Hz and higher frequencies. The assumption of sphericity for repeated tests of gender effect in DPOAE amplitudes was checked with Mauchly's test, and no statistically significant difference was found according to gender at all the tested frequencies ($\rho > 0.05$). Additionally, the DPOAE three measurement evaluations of female and male were examined with repeated measures ANOVA and no significant differences were observed. A comparison of DPOAE amplitude, SNR values, ICC, and SEM values in dB in retest measurements according to the f2 frequency variable is presented in Table 1. When the obtained data were analyzed, the amplitude reliability of the DPOAE measurement was between 0.78 and 0.96 and the SNR reliability was between 0.71 and 0.88. The SEM for the amplitude value of the DPOAE measurement was in the 0.41–2.58 range, and the highest SEM was at 996 Hz, 6728 Hz, and 7998 Hz. The SEM of the DPOAE measurement for SNR was in the 1.19–2.99 range, with higher SEM values observed at low frequencies than at high frequencies.

The mean values of the TEOAE amplitude values of the frequency variables taken from the test and retest conditions are shown in Fig. 3. When the findings are examined, it is seen that the maximum amplitude is at the frequency of 1 kHz and the minimum amplitude is at the frequency of 4 kHz. There was no statistically significant difference between the three measurements in test–retest conditions in five different f2 frequency variables ($F = 0.550 = \rho = 0.058$). The p values of the TEOAE test–retest measurements were 0.281 at 1 kHz, 0.143 at 1.5 kHz, 0.831 at 2 kHz, 0.442 at 3 kHz, and 0.306 at 4 kHz. When the TEOAE test–retest measurement results were examined, no significant difference was found in any of the frequencies tested in the repeated TEOAE amplitude values ($\rho < 0.005$).

Table 1 Test–retest ICC and SEM of DPOAE measurements between test series

		F2 frequency (Hz)												
		996	**1191**	**1416**	**1679**	**2001**	**2382**	**2832**	**3359**	**4003**	**4755**	**5654**	**6728**	**7998**
AMPLITUDE	**ICC**													
	M1–M2	0.90	0.92	0.96	0.94	0.92	0.93	0.91	0.96	0.94	0.94	0.96	0.85	0.88
	M1–M3	0.86	0.92	0.93	0.91	0.88	0.87	0.84	0.93	0.90	0.91	0.91	0.78	0.88
	M2–M3	0.85	0.92	0.95	0.92	0.93	0.91	0.87	0.93	0.91	0.92	0.91	0.84	0.87
	SEM													
	M1–M2	1.14	0.77	0.42	0.59	0.74	0.62	0.96	0.41	0.56	0.72	0.57	1.76	1.46
	M1–M3	1.42	0.75	0.72	0.76	1.05	1.19	1.70	0.76	1.02	1.18	1.17	2.58	1.42
	M2–M3	1.46	0.84	0.54	0.66	0.61	0.80	1.36	0.77	0.92	1.09	1.18	1.77	1.53
SNR	**ICC**													
	M1–M2	0.82	0.81	0.87	0.73	0.82	0.83	0.85	0.88	0.80	0.88	0.87	0.83	0.82
	M1–M3	0.74	0.78	0.76	0.80	0.77	0.78	0.79	0.83	0.74	0.84	0.84	0.77	0,80
	M2–M3	0.77	0.77	0.86	0.76	0.79	0.84	0.71	0.82	0.80	0.88	0.84	.79	0.82
	SEM													
	M1–M2	1.78	2.03	1.31	2.52	1.64	1.53	1.49	1.18	1.70	1.33	1.43	1.37	1.19
	M1–M3	2.37	2.21	2.38	1.80	1.96	2.07	2.20	1.62	2.30	1.75	1.72	1.73	1.25
	M2–M3	2.16	2.28	1.47	2.03	1.86	1.53	2.99	1.73	1.75	1.33	1.72	1.58	1.19

SEM Standard measurement error, *ICC* İntraclass correlation coefficient, *SNR* Signal/noise ratio

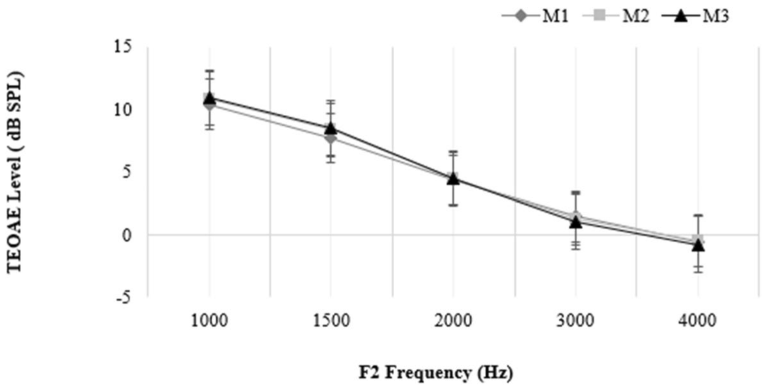

Fig. 3 The vertical lines show TEOAE level in dB SPL and the horizontal lines the f2 frequency in Hz. TEOAE indicates transient evoked otoacoustic emission.The first session consists of the baseline measurement (measurement1 = M1). The second session (measure2 = M2) after 20 min. Third session (measure3 = M3) after 20 days. The figure shows mean TEOAE amplitude of the frequency variables from the test and retest conditions

Figure 4 shows the TEOAE amplitude repeatability in the f2 frequency variable by gender. The assumption of sphericity for repeated tests of gender effect in TEOAE amplitudes was checked with Mauchly's test, and no statistically significant difference was found according to gender at all the tested frequencies ($\rho > 0.05$). Additionally, the TEOAE three measurement evaluations of female and male were examined with repeated measures ANOVA and no significant differences were observed. A comparison of TEOAE amplitude and SNR values, ICC, and SEM values in dB in retest measurements according to the f2 frequency variable is presented in Table 2. When the obtained data were analyzed, the amplitude reliability of the TEOAE measurement was between 0.85 and 0.99 and the SNR reliability was between 0.85 and 0.97. The SEM for the amplitude value of the TEOAE measurement was in the 0.12–0.92 range, and the highest SEM was at 1000 Hz and 1500 Hz. The SEM of the TEOAE measurement for SNR was in the 0.26–1.25 range, with the highest SEM value observed at 1000 Hz. The SEM for the amplitude value of the TEOAE measurement was in the 0.12–0.92 range, and the highest SEM was at 1000 Hz. There was no statistically significant difference in test–retest reliability in the measurements made at different times in the DPOAE and TEOAE tests. When the measurements made in general were examined, high ICC

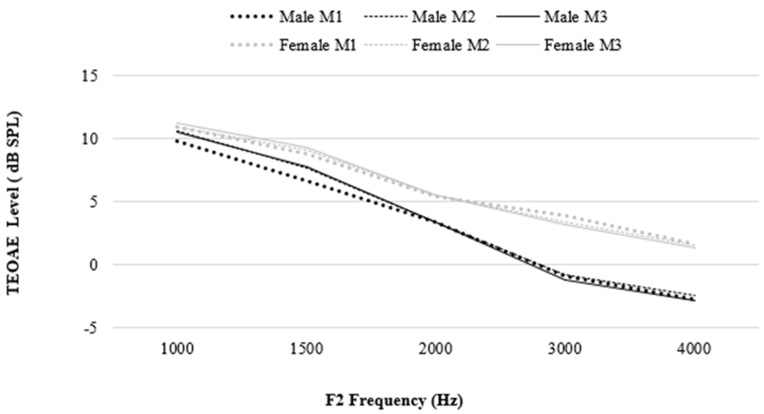

Fig. 4 The vertical lines show the TEOAE level in dB SPL and the horizontal lines the f2 frequency in Hz. TEOAE indicates transient evoked otoacoustic emission. The figure shows mean TEOAE amplitudes repeatability in the f2 frequency variable by gender

Table 2 Test–retest ICC and SEM of TEOAE measurements between test series

	F2 frequency (Hz)					
		1000	**1500**	**2000**	**3000**	**4000**
AMPLITUDE	**ICC**					
	M1–M2	0.94	0.90	0.98	0.99	0.98
	M1–M3	0.86	0.85	0.94	0.97	0.97
	M2–M3	0.90	0.92	0.95	0.97	0.98
	SEM					
	M1–M2	0.52	0.88	0.15	0.12	0.19
	M1–M3	0.92	0.92	0.46	0.32	0.23
	M2–M3	0.79	0.67	0.42	0.33	0.22
SNR	**ICC**					
	M1–M2	0.89	0.92	0.96	0.97	0.97
	M1–M3	0.88	0.91	0.94	0.95	0.95
	M2–M3	0.85	0.90	0.95	0.96	0.95
	SEM					
	M1–M2	1.03	0.72	0.33	0.34	0.26
	M1–M3	1.13	0.78	0.51	0.45	0.42
	M2–M3	1.25	0.83	0.46	0.40	0.41

SEM Standard measurement error, *ICC* İntraclass correlation coefficient, *SNR* Signal/noise ratio

values and low SEM values were determined. Additionally, reliability decreased with increasing time intervals between measurements.

Discussion

Every factor that affects EOAE amplitude and variability also affects EOAE repeatability. Measurement repeatability is affected by factors such as the test parameters, characteristics of test equipment, condition of the ear canal and/or middle ear conditions, probe placement with adequate isolation, ambient noise, participant-induced noise (breathing, swallowing, moving), and SNR value [11, 16, 22].

Franklin et al. (1992) examined the repeatability of DPOAE measurement over 4 weeks and found that the variability increased at 1 kHz and 8 kHz. In the measurements made at 65 dB SPL stimulus intensity, the reliability coefficient for amplitude responses at frequencies in the 2–8 kHz range, excluding 1 kHz, was generally around 0.90. Franklin et al. (1992) obtained the SEM value as the mean of 1.62 dB for measurements made between days and 2 dB for measurements made between weeks. Hallenbeck and Dancer [16] evaluated the DPOAE test–retest measurements in the 0.75–8 kHz frequency range and reported that the mean test–retest differences rarely exceeded 1 dB, which is clinically negligible. Ng and McPherson [7] evaluated the reliability of very short-term (after 20 min) and short-term (after a mean of 15 days) test–retest with DP-gram protocols using a 70-dB SPL pure sound stimulus in young adults with normal hearing and found no significant difference between the measurements. In the study, the SEM value for the DPOAE amplitude in the 1–6 kHz frequency range was determined to be in the 1.11–3.45 dB range in the test–retest measurements. In this study, the SEM value for the DPOAE amplitude was determined to be in the 0.41–2.58 dB range in the test–retest measurements. Wagner et al. [18] evaluated the repeatability of DPOAE with measurements made immediately without changing the probe with L1/L2 = 63/60 dB stimulus intensity and measurements made after days (5–10 days). The mean reliability was 0.89 in the measurements made after days, and the SEM value was approximately 1.44 dB in the 1–5 kHz frequency range and about 2.3 dB at 6 kHz. In this study, the test–retest differences were generally lower than 2 dB in the test sessions performed on the same day and days later. The mean reliability was 0.89 in the measurements made after days, and the SEM was approximately 1 dB in the 996–5654 Hz range and approximately 2.2 dB at 6728 Hz. The study findings were thought to be similar due to the inclusion of data with SNR ≥ 6 dB in both studies. Beattie et al. [15] evaluated the repeatability of SNR responses with very short-term and short-term measurements with the DPOAE measurements at L1 = L2 = 65 dB stimulus intensity. In the study, the SNR values were grouped into 3 dB, 6 dB, and 12 dB. The SEM values were examined, and it was reported that SNR did not have a significant effect on reliability. The SEM value at 550 Hz was reported to be about twice as high as the values at the other frequencies. In this study, the highest SEM value (2.37 dB) was obtained at 996 Hz, and the lowest frequency among the frequencies tested for SNR responses in short-term measurements.

Although the standard error in DPOAE amplitude responses was not highest at 996 Hz, high SEM values at low frequencies were observed in the SNR responses. This is thought to be because environmental noise and participant noise are higher at low frequencies and the SNR values are lower. In a study that examined the differences in mean DPOAE amplitude between repeated measurements in young adults with normal hearing, greater variability was observed at higher frequencies than at lower frequencies [23]. Another study involving young adults with normal hearing evaluated the DPOAE test–retest repeatability at L1/L2 = 65/55 dB SPL and 75/70 dB SPL stimulus intensity and found decreased reliability at 1 kHz, 4 kHz, and 8.0 kHz [14]. In this study, the reliability was lower at 996 Hz, 6728 Hz, and 7998 Hz than at the other frequencies. Wagner et al. [18] reported that in their study, the repeatability was significantly reduced at 6 kHz than at the 1–4 kHz frequency range. Hallenbeck and Dancer [16] found the best repeatability values at 2187 Hz and the worst at 5500 Hz in their study data. Ng and McPherson (2005) obtained higher variability above 6 kHz and found the best repeatability values at 1587 Hz and 2002 Hz. In the study, the SEM value of the DPOAE amplitude was found to be 2.59 dB in short-term retest measurements at 6.5 kHz [7]. In the current study, similar to the literature, the highest repeatability values were obtained at 1416 Hz, and the worst were obtained at 6728 Hz. The highest SEM value reached 2.58 dB at 6728 Hz. Reavis et al. [24] conducted a meta-analysis of studies that investigated DPOAE test–retest differences for adults using SEM statistics and found that the DPOAE variability increased as the time between the test repetitions increased. Although all the SEM values in the study increased as the time between the tests increased, the test–retest time interval was found to be statistically significant at 4 kHz and 6 kHz. In the current study, DPOAE variability increased as the time between tests increased. Although the SEM values increased as the test–retest times increased, the test–retest time interval was found to be statistically significant at 6728 Hz. Various studies investigating the repeatability of DPOAEs in normal hearing subjects have reported greater variability and lower reliability for short-term measurements compared to very short-term measurements [14,15, 25]. In this study, the probe effect on test repeatability was not examined. However, in this study, test–retest variability increased and reliability decreased in short-term measurements compared to very short-term measurements. It was thought that the increased time interval in short-term measurements created more changes in hearing (middle ear pressure changes, noise exposure, etc.), affecting test–retest variability and reliability.

In various studies in the literature examining the effect of gender on DPOAE measurement, DPOAE amplitudes were found to be significantly higher in women than in men [26, 27]. In a study examining the effect of gender on DPOAE amplitudes in individuals with normal hearing, higher DPOAE amplitude values were reported in women compared to men in the 2–6 kHz frequency range, and higher DPOAE amplitude values in men compared to women in frequencies below 2 kHz [10]. Bowman et al. (2000) found that DPOAE amplitudes at frequencies below 3.3 kHz were very similar between genders and reported that women had higher amplitudes at higher frequencies (3.3–13 kHz), but not at a statistically significant level [28]. McFadden et al. (2009) found that DPOAE amplitudes were higher in young women compared to men in the 1.5–4 kHz frequency range [29]. It has been suggested that the gender effect seen in

DPOAE amplitudes can be explained by the differences in head size and basilar membrane resulting from the male cochlea being longer than the female cochlea [30, 31]. In the present study, similar to the literature, DPOAE amplitudes were found to be higher in women than in men at medium and high frequencies. To the best of our knowledge, the effect of gender on OAE repeatability has not been investigated in the literature. In the current study, no statistically significant difference was found between very short-term and short-term measurements depending on gender in any of the f2 frequencies measured by DPOAE. At the same time, the lowest repeatability was observed at 996 Hz and 6728 Hz in females, while it was observed at 6728 Hz and 7998 Hz in males. This study is consistent with other studies where DPOAE amplitudes did not change in same-day measurements [7, 9, 27]. In the literature, it has been determined that the test–retest reliability decreases with the removal and repositioning of the test probe, increasing the test time from the same day to weeks, decreasing the L2 level, and increasing the f2 value above 4 kHz [7, 9, 11, 15, 16, 18, 23]. In studies evaluating DPOAE test–retest repeatability, SEM values were usually 2 dB and a reliability coefficient higher than 0.90 [9, 17, 18].

The general finding emerging from the literature can be summarized as decreased DPOAE repeatability at frequencies below 1 kHz and above 6 kHz in both children and adults [11, 15, 18, 23, 24, 32]. The high variability obtained from the low-frequency f2 values is thought to be due to low-frequency noise pollution, physiological respiration, vascular sounds, participant noise from small movements such as swallowing and coughing, and lower SNR values [18, 33]. It has been reported that standing waves affect stimulus and measurement responses at both low (3 to 7 kHz) and high (> 10 kHz) frequencies [34, 35]. The decreased repeatability at 6 kHz is thought to reflect the standing wave phenomenon occurring at higher frequencies due to the relationship between ear canal length and wavelength, and distort DPOAE amplitudes by increasing intrinsic random variability [22, 36]. Weaker reliability at high-frequency f2s was thought to be related to variability in probe placement due to limitations in current calibration techniques to correct for standing wave effects. It was concluded that the use of improved calibration methods such as depth-compensated SPL or forward pressure level would minimize frequency-specific errors and thus reduce test–retest variability [15]. The present differences in SEM values given in various studies are thought to be due to test parameters, application, equipment, and software differences [18, 37].

In various studies evaluating TEOAE test–retest reliability, no statistically significant difference was found in the measurements made on the same day or weeks later, and high test–retest reliability was obtained [8–10, 13, 37].

In the present study, evaluating short-term and very short-term TEOAE repeatability at 85 dB SPL stimulus intensity, high test–retest reliability was obtained at all tested frequencies and all time intervals. In a study evaluating TEOAE test repeatability with very short (20 min later) and short-term (mean after 15 days) measurements at 83 dBpeSPL stimulus intensity in young adults with normal hearing, the ICC values for SNR responses were found on the mean of 0.86 for very short-term measurements and 0.83 for short-term measurements [7]. In this study, the mean ICC values for SNR responses in TEOAE test–retest measurements were 0.94 for very short-term measurements and 0.92 for short-term measurements. In another study investigating the test–retest reliability of TEOAEs at 80 dBpeSPL stimulus intensity in adults with normal hearing, the mean ICC values were found to be 0.98 in very short-term (measurements made immediately after removing the probe and measurements taken approximately 1 h) measurements and 0.97 in short-term (measurement after 7 days) measurements [14]. When the variability between measurements was examined, higher variability was reported at 1 kHz and 1.4 kHz frequencies and the highest SEM was obtained at 1 kHz. In the study, the mean SEM was 0.62 dB in the very short-term measurement and 0.83 dB in the short-term measurement, and the SEM values increased with the increase in the time interval between the measurements [14]. Similarly, in the present study, the highest variability between measurements was detected at 1.5 kHz, and a decrease in inter-session reliability was observed with the increase in the time interval between measurements. In this study, the mean ICC values were 0.96 in very short-term measurements and 0.93 in short-term measurements, and the mean SEM for amplitude response was 0.37 dB in very short-term measurements and 0.53 dB in short-term measurements. It has been reported in the literature that the SEM value for TEOAEs evaluated with click stimulus is approximately 1 dB [8, 9]. In this study, SEM values of less than 1 dB were obtained for TEOAEs.

The higher TEOAE amplitudes in females compared to males in the study are consistent with previously reported studies [38, 39]. Therefore, the gender differences found in this study are thought to be due to the higher number of outer hair cells and the shorter cochlea in females [30]. To the best of our knowledge, gender-related repeatability has not been evaluated in the literature. In the current study, a not statistically significant difference was found between very short-term and short-term measurements depending on gender at f2 frequencies where TEOAE measurements were made. In addition, the lowest repeatability was observed at 1 kHz and 1.5 kHz frequencies in both genders. Compared with the studies using similar pure tone stimuli, TEOAE and DPOAE repeatability were higher in this study. The reason for this finding was thought to be due to the inclusion of data of high SNR values in the study. When the SEM values obtained from similar half-octave band frequencies were examined, it was reported that the reliability of TEOAEs was higher than that of DPOAEs [14, 40]. In the current study, the reliability of TEOAEs was obtained higher than DPOAEs and supports the literature.

Conclusions

This study demonstrated that the TEOAE and DPOAE measurements are reliable for monitoring cochlear function within minutes and weeks with high test–retest reliability. The minimum 6 dB SNR criterion, which is widely used in clinical OAE measurements, is considered to be an important criterion when the repeatability and reliability findings are examined. In DPOAE amplitude measurements, the 6 kHz and 8 kHz high-frequency regions are used to monitor the cochlear state, especially in cases such as exposure to ototoxic drugs or noise. In the study, 6728 Hz and 7998 Hz frequencies showed higher variability than other frequencies in DPOAE measurements. Since the 6 kHz and 8 kHz high-frequency regions in DPOAE amplitude measurements are used to monitor the cochlear state, especially in cases such as exposure to ototoxic drugs or noise, it is recommended that clinicians interpret the measurement results obtained at these frequencies carefully. During the TEOAE measurements, the highest acoustic background noise level was observed at frequencies of 1 kHz and 1.5 kHz. Therefore, it is recommended that clinicians take the initiative to reduce the background noise level, especially at these frequencies.TEOAEs and DPOAEs have high test–retest reliability, but reliability may vary depending on the recording device and test parameters. Further research is recommended to examine the test–retest reliability of DPOAEs and TEOAE measurements at different test parameters and in populations with varying degrees of hearing loss.

Acknowledgements
None

Authors' contributions

A. A. and B. Ö. designed and performed the study; A. A. collected the data; A. A. analyzed the data, and B. Ö. contributed to the interpretation of the analysis; A. A. and B. Ö. wrote and edited the article; all authors discussed the results and approved the submission of the article.

Funding
None.

Availability of data and materials

Data are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was conducted in the Department of Audiology with the approval of the Ethics Committee of the KTO Karatay University's Faculty of Medicine ((Decision no: 2022/011). The informed consent form and detailed anamnesis were obtained from all participants. The manuscript adheres to the ethical standards according to the Declaration of Helsinki.

Consent for publication

The written consent has been obtained from the that the information will be shared and utilized for educational purposes only.

Competing interests

The authors declare that they have no competıng ınterests.

Received: 9 August 2023 Accepted: 8 September 2023

References

1. -Kemp DT (2008) Active processes and otoacoustic emissions in hearing. Manley G. A. vd. (Ed.), Otoacoustic Emissions: Concepts and Origins. New York, Springer Science Business Media.
2. - Kramer S, Brown DK (2019) Audiology science to practice. B. Stach (Ed.), Evoked Physiologic Responses. San Diego, CA, Plural Publishing 244–251.
3. - Stach BA (2010) Clinical audiology: an ıntroduction. United States, Delmar Cencage Learning. 383-388
4. Dhar S, Hall JW (2012) Otoacoustic emissions: principles, procedures, and protocols. Plural Publishing, San Diego
5. - Welling DR, Ukstins CA (2013) Fundamentals of audiology for the speech-language pathologist. In Jones & Bartlett Publishers, 107-109
6. - Gelfand SA. (2016) Essentials of audiology. 4rd chapter. New York, NY, Thieme
7. Ng IH, McPherson B (2005) Test-retest reliability of DPOAEs in the 1 to 7 kHz range. Audiological Medicine 3:108–115
8. Harris FP, Probst R, Wenger R (1991) Repeatability of transiently evoked otoacoustic emissions in normally hearing humans. Audiology 30(3):135–141
9. Franklin DJ, McCoy MJ, Martin GK, Lonsbury-Martin BL (1992) Test/retest reliability of distortion-product and transiently evoked otoacoustic emissions. Ear Hear 13(6):417–429
10. - Dieler R, Shehata-Dieler WE, Klagges T, Moser LM (1999) Intra- and intersubject variability of acoustically evoked otoacoustic emissions. I. Transiently evoked otoacoustic emissions. Laryngo Rhino Otol 78(6):339–344.
11. Zhao F, Stephens D (1999) Test-retest variability of distortion-product otoacoustic emissions in human ears with normal hearing. Scand Audiol 28(3):171–178
12. Chan RH, McPherson B (2000) Test-retest reliability of tone-burst-evoked otoacoustic emissions. Acta Otolaryngol 120(7):825–834
13. Barboni M, Geralde AT, Goffi-Gomez VS, Schultz C, PecoraLiberman PH (2006) Test-retest variability of the transient otoacoustic emissions in normal hearing subjects. İnt Arch Otorhinolaryngol 10(2):371
14. Keppler H, Dhooge I, Maes L, D'haenens W, Bockstael A, Philips B, Swinnen F, Vinck B, (2010) Transient-evoked and distortion product otoacoustic emissions: a short-term test-retest reliability study. Int J Audiol 49(2):99–109
15. Beattie RC, Kenworthy OT, Luna CA (2003) Immediate and short-term reliability of distortion product otoacoustic emissions. Int J Audiol 42(6):348–354
16. Hallenbeck H, Dancer J (2003) Distortion-product otoacoustic emissions in ears with normal hearing sensitivity: test-retest variability. Perceptual Motor Skills 97(3:1):990–992
17. Stuart A, Passmore AL, Culbertson DS, Jones SM (2009) Test-retest reliability of low-level evoked distortion product otoacoustic emissions. J Speech Language Hear Res 52(3):671–681

18. Wagner W, Heppelmann G, Vonthein R, Zenner HP (2008) Test-retest repeatability of distortion product otoacoustic emissions. Ear Hear 29(3):378–391
19. Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19:231–40
20. Demorest ME, Walden BE (1984) Psychometric principles in the selection, interpretation, and evaluation of communication self assessment inventories. J Speech Hear Disord 49(3):226–240
21. Koo TK, Li M Y (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 15(2):155–163
22. Mills DM, Feeney MP, Drake EJ, Folsom RC, Sheppard L, Seixas NS (2007) Developing standards for distortion product otoacoustic emission measurements. J Acoustic Soc Am 122(4):2203–2214
23. Dreisbach LE, Long KM, Lees SE (2006) Repeatability of high-frequency distortion-product otoacoustic emissions in normal-hearing adults. Ear Hear 27:466–479
24. Reavis KM, McMillan GP, Dille MF, Konrad-Martin D (2015) Meta-analysis of distortion product otoacoustic emission retest variability for serial monitoring of cochlear function in adults. Ear Hear 36(5):251–260
25. Pilka E, Jedrzejczak WW, Kochanek K, Skarzynski H (2019) Variability of high-frequency distortion product otoacoustic emissions measured by the smartoae device: preliminary study. J Hear Sci 9(3):60–65
26. Gaskill SA, Brown AM (1990) The behavior of the acoustic distortion product, 2f1-f2, from the human ear and its relation to auditory sensitivity. J Acoustic Soc Am 88(2):821–839
27. Cacace AT, McClelland WA, Weiner J, McFarland DJ (1996) Individual differences and the reliability of 2F1-F2 distortion-product otoacoustic emissions: effects of time-of-day, stimulus variables, and gender. J Speech Hear Res 39(6):1138–1148
28. Bowman DM, Brown DK, Kimberley BP (2000) An examination of gender differences in DPOAE phase delay measurements in normal-hearing human adults. Hear Res 142(1–2):1–11
29. McFadden D, Martin GK, Stagner BB, Maloney MM (2009) Sex differences in distortion-product and transient-evoked otoacoustic emissions compared. J Acoustic Soc Am 125(1):239–246
30. Sato H, Sando I, Takahashi H (1991) Sexual dimorphism and development of the human cochlea. Computer 3-D measurement. Acta Otolaryngol. 111(6):1037–1040
31. Krizman J, Skoe E, Kraus N (2012) Sex differences in auditory subcortical function. Clin Neurophysiol 123(3):590–597
32. Sockalingam R, Lee Choi J, Choi D, Kei J (2007) Test-retest reliability of distortion-product otoacoustic emissions in children with normal hearing: a preliminary study. Int J Audiol 46(7):351–354
33. Thorson MJ, Kopun JG, Neely ST, Tan H, Gorga MP (2012) Reliability of distortion-product otoacoustic emissions and their relation to loudness. J Acoustic Soc Am 131(2):1282–1295
34. Siegel JH (1994) Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes. J Acoust Soc Am 95:2589–2597
35. Dreisbach LE, Siegel JH (2001) Distortion-product otoacoustic emissions measured at high frequencies in humans. J Acoust Soc Am 110:2456–2469
36. Siegel JH (2002) Calibrating otoacoustic emission probes. In: M. S. Robinette, & T. J. Glattke (Eds.), Otoacoustic emissions: clinical applications. New York: Thieme, 416-438
37. Kochanek KM, Śliwa LK, Puchacz K, Piłka A (2015) Repeatability of transient-evoked otoacoustic emissions in young adults. Med Sci Mon 21:36–43
38. McFadden D, Loehlin JC, Pasanen EG (1996) Additional findings on heritability and prenatal masculinization of cochlear mechanisms: click-evoked otoacoustic emissions. Hear Res 97:102–119
39. Stuart A, Kerls AN (2018) Does contralateral ınhibition of transient evoked otoacoustic emissions suggest sex or ear laterality effects? Am J Audiol 27(3):272–282
40. Lapsley Miller JA, Marshall L (2001) Monitoring the effects of noise with otoacoustic emissions. Semin Hear 22(4):393–403

Publisher's Note