



Article

Next-Generation Auditory Steady-State Responses in Normal-Hearing Adults: A Pilot Test–Retest Reliability Study

Hanan Hamad ¹, Nilesh J. Washnik ^{2,*}  and Chandan H. Suresh ^{1,*}

¹ Department of Communication Disorders, California State University, Los Angeles, CA 90032, USA; hhamad2@calstatela.edu

² Department of Hearing, Speech, and Language Sciences, Ohio University, Athens, OH 45701, USA

* Correspondence: washnik@ohio.edu (N.J.W.); csuresh@calstatela.edu (C.H.S.)

Abstract: The Auditory Steady-State Response (ASSR) provides objective and ear-specific information essential for early and appropriate intervention. Test–retest reliability is essential for audiological monitoring. The test–retest reliability of the ASSR has received limited attention. Only a handful of studies found in the literature investigated the test–retest reliability of old-generation ASSR using amplitude or mixed modulated stimuli. However, to our knowledge, no published reports have specifically examined the test–retest reliability of the next-generation ASSR using Chirp family stimuli as implemented in the Interacoustics Eclipse system. This pilot study investigated (a) the test–retest reliability of air conduction (AC) ASSR thresholds across two test sessions and (b) the relationship between differences in ASSR thresholds across two sessions to the residual noise levels in normal-hearing adults. **Methods:** Fifteen normal-hearing adults (12 females) (30 ears) with an average age of 28 years were recruited for the study. The ASSRs were recorded using a two-channel recording montage. The automatic default stimuli and recording protocol using 90 Hz ASSR, and the accuracy method ($p < 0.01$) as implemented in the Eclipse system is used to measure ASSR. **Results:** The study demonstrated strong test–retest reliability for ASSR across frequencies (500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz). Notably, the highest reliability was observed at 500 Hz. The mean test–retest reliability of ASSR was found to be comparable to pure-tone thresholds, but the intra-subject variability is higher for ASSR compared to pure-tone thresholds. Additionally, no significant correlation was found between the difference in ASSR residual noise levels at the threshold and the difference in ASSR thresholds at all tested frequencies. **Conclusion:** The next-generation system demonstrated strong test–retest reliability across the frequencies examined in this pilot study. Particularly, an improvement in reliability was observed at 500 Hz compared to the old-generation ASSR. This enhancement can be attributed to the utilization of narrow-band CE-chirp stimuli, which generate large amplitude responses, and the implementation of an improved detection paradigm involving multiple harmonics spectral and phase analysis. This pilot study only enrolled adults with normal hearing, and future investigations should include a larger sample size comprising both normal-hearing and hearing-impaired individuals, as well as the pediatric population.



Citation: Hamad, H.; Washnik, N.J.; Suresh, C.H. Next-Generation Auditory Steady-State Responses in Normal-Hearing Adults: A Pilot Test–Retest Reliability Study. *J. Otorhinolaryngol. Hear. Balance Med.* **2023**, *4*, 6. <https://doi.org/10.3390/ohbm4020006>

Academic Editor: Jeroen C. Jansen

Received: 6 June 2023

Revised: 5 July 2023

Accepted: 7 July 2023

Published: 10 July 2023

Keywords: Auditory Steady-State Response (ASSR); next-generation ASSR; NB CE-chirps; test–retest reliability



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Auditory Steady-State Response (ASSR) is a sustained electrophysiological response that is phase-locked to the steady component of the stimulus (either the amplitude modulation frequency or the repetition rate of the stimulus), which maintains amplitude and phase throughout the presentation of the stimuli [1,2]. It facilitates frequency-specific estimation of hearing thresholds. The ASSR has been introduced as a possible alternative to the Auditory Brainstem Response (ABR) to reduce test time and improve auditory diagnostic evaluations [3]. The ASSR provides objective and ear-specific information essential

for early and appropriate intervention, such as fitting hearing aids or cochlear implants. Tone-burst-evoked ABRs are the gold standard for estimating hearing thresholds in infants, yet the ASSR is one of the emerging technologies that has improved test efficiency [3,4]. The advantages of the ASSR over the tone burst ABR includes completely objective detection of responses and shorter overall test time due to the multiple stimulus presentations [3–7]. The first-generation ASSRs using amplitude or mixed modulation stimuli with various carrier frequencies (500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz) were implemented to reliably record ASSR responses [1,2,5,7,8]. But, the early ASSR systems used a stimulus and detection paradigm that lacked sufficient sensitivity in finding low-amplitude responses, leading to large discrepancies between ASSR threshold and behavioral thresholds for individuals with normal hearing and lower degrees of hearing loss [8–11]. These large discrepancies led to reluctance in the routine use of ASSRs among clinicians.

However, the next-generation ASSR with a narrow-band CE-chirp stimulus presented at different repetition rates has offered promise with improved detection rates and a reduced test time [3,4,12]. The commercial Eclipse system incorporates a multi-harmonic detector that analyzes the amplitude and phase of the fundamental frequency (repetition rate of the individual stimuli) and up to 12 higher harmonics of the ASSR. Through statistical analysis, it compares the response with residual noise levels to objectively determine the presence or absence of the ASSR. Further details about the algorithm for automatic detection can be found in the comprehensive review by Sininger et al. [3].

In a retrospective study involving 667 patients, Ehrmann-Müller et al. [13] reported highly promising results for the application of narrow-band chirp ASSR. They found a significant correlation between Puretone audiometry results and ASSR measurements across the patient population. Additionally, a recent study conducted by Vanaja and Kunjir [14] focused on children aged 3 to 5 years. Their findings indicated that aided chirp ASSR can accurately predict aided behavioral thresholds. This provides further evidence for the use of ASSR in the hearing aid fitting and verification process. The use of next-generation ASSRs with narrow-band chirps shows great potential for assessing auditory function in both adult and pediatric populations, serving as a valuable diagnostic tool and aiding in hearing aid validation.

Test–retest reliability holds paramount importance in audiological monitoring as it ensures consistency and accuracy of measurements over time. This metric is commonly assessed by obtaining repeated measurements from the same individuals, which facilitates an evaluation of the extent to which a test produces consistent results under similar conditions. The ability to reliably measure and track changes in auditory conditions is crucial for clinicians to confidently interpret test results, monitor progressive hearing loss, and make informed decisions regarding diagnosis, treatment, and interventions.

To illustrate the significance of test–retest reliability in audiological monitoring, several examples can be considered. One such example is the monitoring of hearing changes associated with ototoxicity, which has been extensively reviewed by Campbell and Leprell [15]. Patients undergoing treatment with life-saving drugs such as aminoglycoside antibiotics or platinum-based chemotherapies require regular evaluations, typically weekly to biweekly, to detect any potential hearing changes resulting from the administration of these medications. In such cases, the test administered must provide reliable results to differentiate between the hearing changes caused by the drugs and the inherent variability of the test itself. For a significant change in hearing threshold to be identified, the threshold shift must exceed the test–retest variability of the test. As Campbell and Leprell [15] have noted, a shift greater than 10 dB is considered clinically significant.

Another example pertains to monitoring hearing changes in patients with progressive hearing loss. As highlighted by Barreira-Nielsen et al. [16], it is crucial to track and measure changes in hearing thresholds over time to determine the progression of the condition accurately and tailor appropriate interventions accordingly.

A comprehensive literature review identified five studies examining the test–retest reliability of the 80 Hz old-generation ASSR using amplitude or mixed modulation stim-

uli with various carrier frequencies [5,17–20]. Luts and Wouters [5] and Kaf et al. [18] explored the test–retest reliability of ASSR threshold estimation in normal-hearing and hearing-impaired adults, with Kaf et al. [18] also considering participants with simulated sensorineural hearing loss. Additionally, Luts and Wouters [5] compared two different systems, AUDERA and MASTER ASSR. Kim and Lee [19] conducted an investigation into the test–retest reliability of bone conduction ASSR in both normal-hearing and hearing-impaired individuals. In a similar vein, Wilding et al. [20] evaluated the repeatability of the ASSR amplitude through test–retest analysis in individuals with normal hearing and those with hearing impairment. Overall, these studies reported moderately strong to strong test–retest reliability for ASSR at carrier frequencies of 1000, 2000, and 4000 Hz, but moderate to weak reliability at 500 Hz. D’Haenens et al. [17] and Wilding et al. [20] highlighted high variability in the amplitude of the ASSR between test sessions, potentially affecting ASSR threshold consistency across sessions.

McFadden et al. [21] measured the intersession reliability of the 40 Hz ASSR for click and white noise stimuli using both amplitude and phase analysis. Their findings revealed a significant correlation between sessions 1 and 2. The study highlighted that phase coherence analysis exhibited higher inter-session reliability compared to amplitude analysis. The authors reported that click-train stimuli demonstrated higher reliability than white noise stimuli. This can be attributed to the higher signal-to-noise ratio associated with click-train stimuli. It is important to note that the study by McFadden et al. [21] focused specifically on click and white noise stimuli and did not consider tone burst or narrowband chirp stimuli in their assessment of inter-session reliability.

To date, no published reports have investigated the next-generation ASSR’s test–retest reliability, specifically as implemented in the Eclipse system. Therefore, the primary objective of this pilot study is to evaluate the test–retest reliability of the next-generation ASSR system, utilizing the multiple ASSR technique in the Eclipse system for air conduction stimulation among normal-hearing adults. Building upon findings from previous studies using the old-generation ASSR, we hypothesize that there will be strong test–retest reliability for ASSR thresholds across two sessions, with higher frequencies exhibiting greater reliability than lower frequencies. Additionally, we anticipate no significant relationship between differences in ASSR thresholds and residual noise levels.

2. Materials and Methods

2.1. Participants

Fifteen participants (12 females, 3 males) with a mean age of 28 years (range: 23–31 years) were recruited from both the Department of Communication Disorders at California State University, Los Angeles, and the local population. A total of 30 ears (right and left ears of 15 participants) were included in the study. The research protocol comprised two sessions: a test session and a retest session. All 15 participants successfully completed both sessions. To gather the required data, two separate one-hour test sessions were conducted for each participant, scheduled with a gap of 7 to 14 days between them. Informed consent was obtained from all participants in adherence to the protocol approved by the Institutional Review Board of California State University, Los Angeles.

2.2. Standard Inclusion and Clinical Procedures

The routine audiological test battery was conducted to determine the eligibility of participants for the study. The study employed the following criteria for inclusion: (1) the absence of recent hearing impairment or aural symptoms, (2) normal findings upon otoscopic examination, (3) normal behavioral audiometry, as defined by pure-tone thresholds no worse than 20 dB HL at each octave frequency ranging from 500 to 4000 Hz, utilizing the modified Hughson–Westlake procedure [22], (4) normal middle-ear function, as confirmed by the presence of a normal 226 Hz tympanogram, and (5) the absence of a significant air–bone gap.

2.3. ASSR Stimuli

Multiple simultaneous dichotic narrow-band (NB) frequency-specific chirp stimuli presentations were utilized with center frequencies of 500, 1000, 2000, and 4000 Hz as implemented in the commercial Interacoustics Eclipse system (software version: 2.1.0). Claus Elberling (CE)-Chirp was designed to compensate for the delay induced by the basilar membrane's traveling wave by presenting low-frequency energy before higher-frequency energy [23]. This approach results in synchronized activation of auditory nerve fibers, leading to larger amplitude responses and lower thresholds [24,25]. The broadband CE-chirp stimulus (level independent) decomposed to a narrow-band (NB) stimulus centered at 500, 1000, 2000, 4000, and 8000 Hz (referred to as NB CE-chirp stimulus of that particular frequency). A newer level-specific (LS) CE-chirp stimulus was developed by modifying the temporal characteristics of the chirp stimulus based on observed level effects related to changes in the cochlear traveling wave delay and the upward spread of the excitation [26]. Broadband CE-chirp LS decomposed to NB stimulus centered at 500, 1000, 2000, 4000, and 8000 Hz (referred to as a NB LS CE-chirp stimulus of that particular frequency). For this study, a dichotic stimulus presentation was chosen, involving the simultaneous delivery of eight stimuli (500 Hz CE-chirp LS, 1000 Hz CE-chirp LS, 2000 Hz CE-chirp LS, and 4000 Hz CE-chirp LS to each ear) through insert earphones (IP 30, RadioEar, Middelfart, Denmark). Each NB CE-chirp LS stimulus was presented at slightly different repetition rates around 90 Hz, ranging from 88 to 98 Hz, as utilized in the commercial Interacoustics Eclipse platform for ASSR recordings.

2.4. ASSR Recording

The ASSRs were collected and analyzed using the automated analysis feature of the Interacoustics Eclipse system. A two-channel recording setup was employed, with four surface electrodes used to record the responses. The non-inverting electrode was positioned on the high forehead at the hairline, while two inverting electrodes were placed on each ear lobe (A1 and A2). The ground electrode was placed on the forehead beneath the non-inverting electrode. Electrode impedance was maintained below 3 kOhms, with an inter-electrode impedance difference kept under 2 kOhms. The rejection level was set at 40 microvolts. Throughout the testing process, continuous monitoring of the EEG was carried out. In cases of excessive EEG noise due to patient movement or artifacts, the testing procedure was temporarily paused until the artifact was resolved.

The ASSRs were recorded using the accuracy-test method and the adult sleeping protocol implemented in the Eclipse software. The accuracy-test method priority considers the significance level of p -value ≤ 0.01 , indicating a 99% or higher confidence in the response. The algorithm employed for the automatic detection of ASSR in the Eclipse system is described in detail by Sininger et al. [3], providing a comprehensive analysis of its functioning and effectiveness. In summary, the multi-harmonic detector analyzes the amplitude and phase of the fundamental frequency (repetition rate) and higher harmonics (up to 12 harmonics in the commercial Eclipse system) of the ASSR to determine the presence or absence of a response. Statistical analysis was conducted to compare the response with the residual noise levels and determine the presence or absence of the ASSR response. Within the accuracy-test method protocol, an ASSR response was considered present if the statistical comparison between the signal and noise yielded a p -value of ≤ 0.01 (indicating a 99% confidence or higher). Conversely, an ASSR response was considered absent if the p -value was >0.01 (less than 99% confidence). The adult sleeping protocol establishes the repetition rates of the stimuli used, employing fast repetition rates of around 90 Hz. During the testing sessions, participants were comfortably positioned on a recliner and instructed to relax, minimize movements, and be encouraged to sleep if possible. Two one-hour test sessions were conducted, scheduled 7 to 14 days apart for all participants, to collect the necessary data.

2.5. ASSR Threshold Estimation

The modified Michael and Jørgensen [12] procedure, as depicted in Figure 1, was employed for determining the ASSR threshold. The measurement procedure started at 40 dBnHL and used a descending technique with 10 dB steps until no ASSR response was detected. A level 5 dB above the no ASSR response was recorded to determine the ASSR threshold. The ASSR threshold was determined as the lowest level at which ASSR was present with 99% confidence ($p = 0.01$), and a non-significant response ($p > 0.01$) was observed 5 dB below this level. This 5 dB accuracy in ASSR threshold measurement allowed us to perform a comparison with pure-tone thresholds, which are clinically measured using a 5 dB step size. The initial recording duration for ASSR was set at 6 min. It was repeated once if the response curve reached 99% confidence within one minute. Detection curves that did not reach 99% confidence within the 6 min recording had stopping criteria: (1) residual noise levels ≤ 20 nV, or (2) response confidence curve $< 50\%$. If the stopping criteria were not met at 6 min, the ASSR recording was extended by 3 more minutes. At a total of 9 min of recording, the stopping criteria were: (1) residual noise levels ≤ 20 nV, or (2) response confidence curve $< 75\%$. If the stopping criteria were not met at the 9 min mark, the ASSR recording was further extended by a final 3 min. If the 99% response confidence was not achieved even at the 12 min mark, the ASSR response was considered absent. Thus, the maximum recording time for ASSR at any specific level was 12 min. If no response was measured at the starting level of 40 dBnHL, the presentation level was increased in 10 dB steps until a response was obtained. The same testing protocols were followed during both test sessions (test and retest).

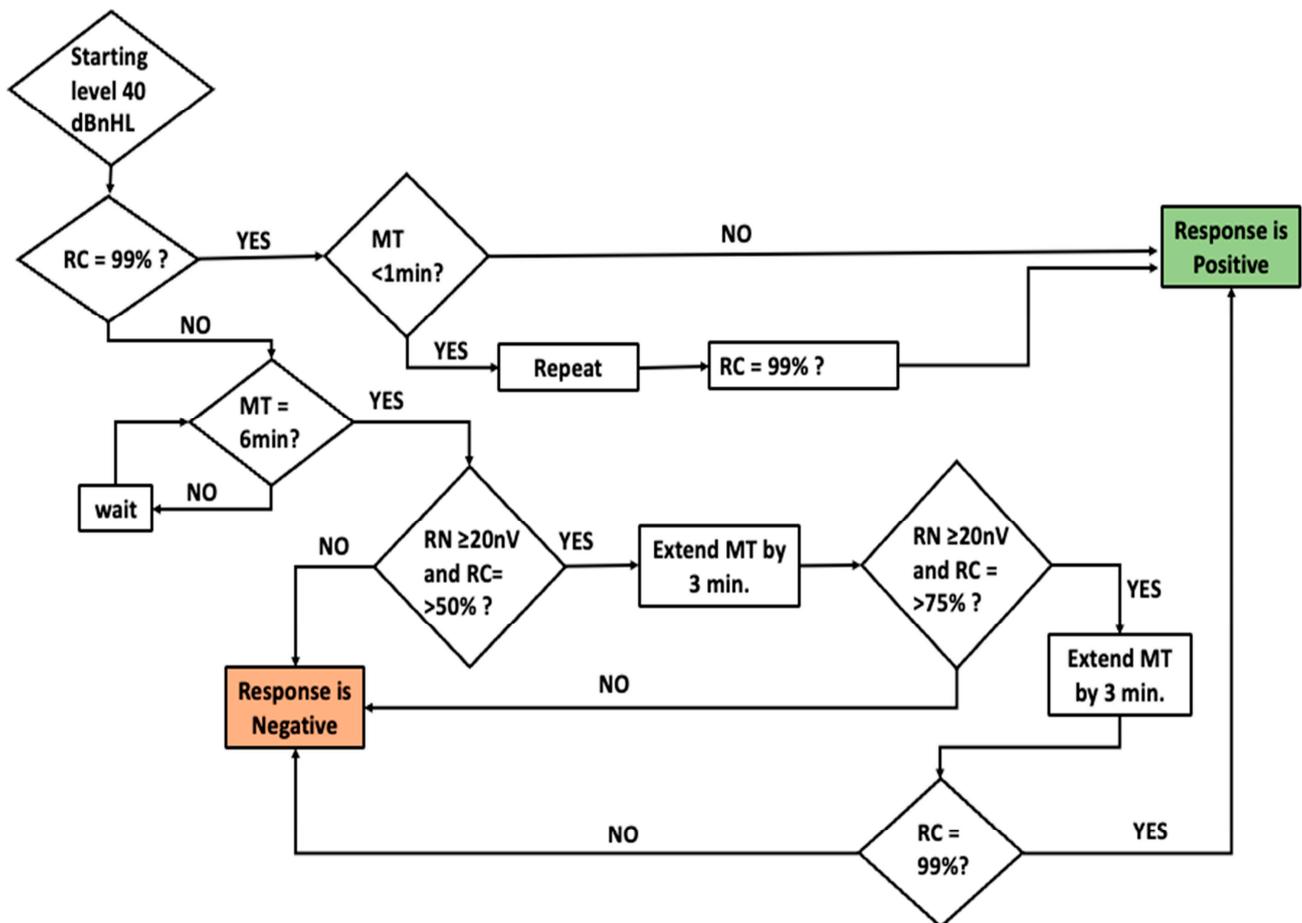


Figure 1. Flowchart illustrating the procedure for determining ASSR thresholds, incorporating response confidence (RC), measurement time (MT), and residual noise (RN).

The Eclipse system presented narrow-band (NB) CE-chirp LS stimuli with different center frequencies and fast repetition rates around 90 Hz. For a 6 min recording at a 90 Hz repetition rate, a total of 32,400 stimuli presentations were administered (90 stimuli per second, resulting in 32,400 stimuli over 6 min). For a 9 min recording, it corresponded to 48,600 sweeps, and for the maximum duration of 12 min, the recording consisted of 64,800 stimuli. As the stimuli were presented in the order of tens of thousands, the determination was made to adhere to the accuracy setting to a p -value of ≤ 0.01 . Also, an initial pilot study comparing the accuracy method ($p \leq 0.01$) and the speed method ($p \leq 0.05$) did not reveal any significant discrepancies, except for thresholds 5 dB lower with the accuracy method at a few individual frequencies.

2.6. Data Analysis

Statistical analyses were performed using Microsoft Excel and SPSS (version 27.0; Armonk, NY: IBM Corp). Pearson product-moment correlation was employed to assess the relationship between ASSR thresholds obtained during sessions 1 and 2, as well as the relationship between the difference in ASSR thresholds between sessions and the difference in residual noise levels at the threshold. To investigate the association between test–retest threshold measurements and frequency, three separate two-way (frequency \times session) ANOVAs were performed on the ASSR threshold, behavioral pure-tone threshold, and the difference between the ASSR and behavioral thresholds. Additionally, a separate two-way ANOVA analysis was conducted to compare the differences in thresholds between sessions 1 and 2 for different testing methods (ASSR and behavioral pure-tone threshold) and frequencies. Before conducting statistical inference, the assumptions of normality and homogeneity of variance were assessed using Levene’s test and visual inspection of normality plots. The data were found to meet the assumptions of normality and homogeneity of variance, thereby fulfilling the requirements for statistical inference. For all post hoc multiple comparisons, a Bonferroni significance level was employed with α set at 0.05.

To determine the likelihood of a repeated measurement falling within a specific range, we use a measure called the standard error of measurement (SEM). The SEM provides an estimate of the standard deviation of test–retest measurements conducted on an individual [27]. It is important to note that the SEM is distinct from the standard error of the mean (SE). Previous studies on the reliability of ASSR have also reported SEM values [17]. To calculate the SEM, we divide the standard deviation of the differences observed between the test and retest sessions by the square root of 2. Typically, a range of ± 1 SEM covers about 68% of the variability. Additionally, a range of ± 1.96 SEM (referred to as “2 SEM” for simplicity) represents a 95% variability range. For instance, if the 2 SEM value is 5 dB, it means that there is a 95% chance that the test and retest values would fall within a 5 dB range. This information is highly valuable in clinical audiological monitoring.

3. Results

3.1. Incidence of ASSR Responses

The occurrence of ASSR at different intensities and carrier frequencies ranging from 500 to 4000 Hz in the test (session 1) and retest session (session 2) is depicted in Figure 2. Measurements were performed on both the right and left ears of the 15 participants, resulting in a total of 30 responses for each individual session. The X-axis represents intensity levels from 0 to 45 dBnHL, while the Y-axis indicates the percentage of response occurrence at each specific intensity. The solid line represents the incidence values for session 1, and the dotted blue line represents the values for session 2. For suprathreshold levels (30 to 45 dB HL), the percentages of responses were relatively similar across carrier frequencies of 1000, 2000, and 4000 Hz. At lower intensity levels (20, 10, and 0 dB nHL), the incidence values were lowest for 500 Hz and highest for 4000 Hz. A similar trend was observed for the percentage of response occurrence across frequencies in both the test and retest sessions.

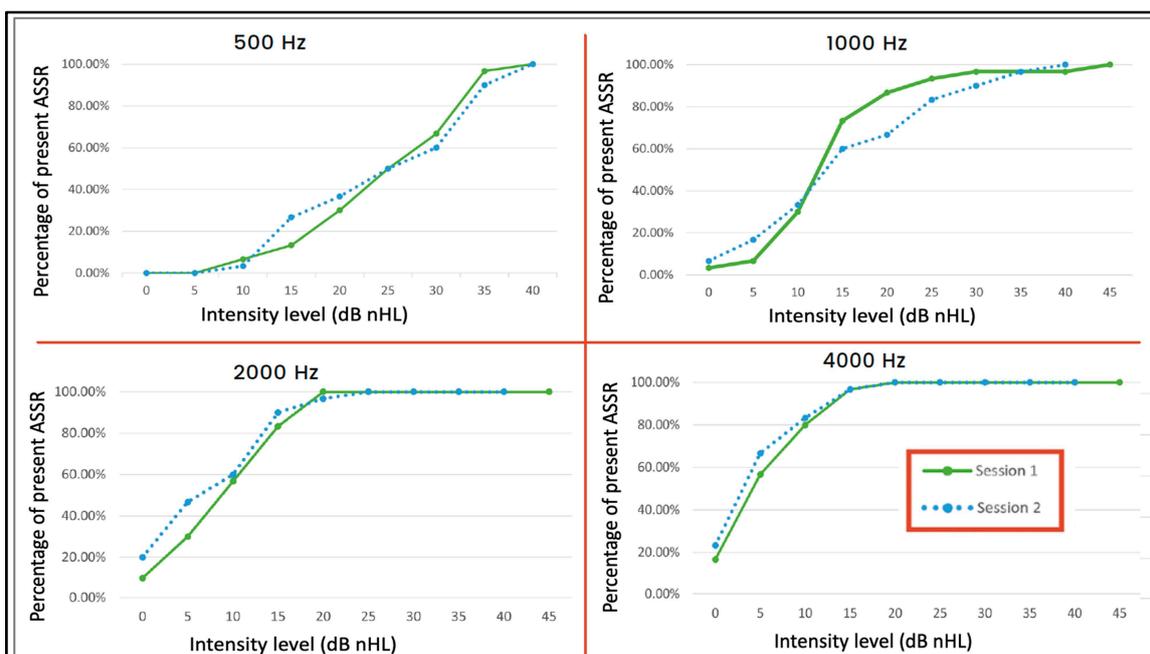


Figure 2. Response occurrence at different intensities and carrier frequencies (ranging from 500 to 4000 Hz) for sessions 1 and 2. Each plot for the frequencies 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz. The solid line represents the values from session 1, while the dotted line represents the values from session 2.

3.2. Test–Retest Reliability

Table 1 presents the ASSR thresholds, behavioral pure-tone thresholds, and difference thresholds (ASSR threshold minus behavioral pure-tone threshold) at each frequency. Consistent with previous studies, ASSR thresholds were higher at 500 Hz and lower at 4000 Hz. The results of the two-way ANOVA (frequency and session as main factors) revealed a significant main effect of frequency ($p < 0.001$). However, there was no significant main effect of sessions (test versus retest) ($p = 0.54$), suggesting that the sessions did not have a significant influence on the observed results. Additionally, there was no significant interaction effect between frequency and session ($p = 0.9$), implying that the relationship between frequency and session was not statistically significant. The frequency main effect showed that the 500 Hz ASSR was found to be higher compared to the thresholds of other frequencies. The threshold for the 1000 Hz ASSR was higher compared to the thresholds at 2000 Hz and 4000 Hz. The results of the two-way ANOVA between behavioral pure-tone thresholds and stimulus frequency yielded similar findings. There was only a significant main effect of frequency ($p < 0.001$), indicating that different frequencies had a significant influence on the behavioral pure-tone thresholds. However, there were no significant main effects of sessions or interaction effects between frequency and session. The frequency main effect showed that the pure-tone thresholds at 4000 Hz were lower than the other frequencies. Furthermore, the results of the two-way ANOVA (difference between ASSR and behavioral pure-tone thresholds across frequency and sessions) showed only a significant main effect of frequency ($p < 0.001$). The difference thresholds at 500 Hz were higher compared to the other frequencies, and the threshold difference at 1000 Hz was greater than at 2000 Hz ($p < 0.05$).

Plotted on the graphs in Figure 3 are the ASSR thresholds for each subject across session 1 (X-axis) and session 2 (Y-axis). Darker dots indicate thresholds that overlap for two or more subjects. Trend lines are included on the graphs for frequencies 500, 1000, 2000, and 4000 Hz. These trend lines serve to illustrate the correlation between the thresholds obtained in session 1 and session 2. The presence of a strong correlation, indicated by

the closeness of the trend lines with data points, suggests a high degree of similarity and consistency between the ASSR thresholds measured in both sessions.

Table 1. The average of ASSR, behavioral pure-tone thresholds, and the difference threshold between ASSR and behavioral pure-tone thresholds. The standard deviations are listed in parentheses.

Frequency (Hz)	ASSR Threshold (dB nHL)		Behavioral Pure-Tone Threshold (dB HL)		Difference Thresholds (dB)	
	Test	Retest	Test	Retest	Test	Retest
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
500 Hz	26.8 (8.2)	26.7 (9.4)	5.7 (6.1)	6.5 (5.1)	20.9 (10.4)	19.5 (9.5)
1000 Hz	17 (9.6)	17.3 (10.4)	6.7 (4.1)	6.1 (4.5)	10.4 (10.5)	10.9 (10.4)
2000 Hz	11 (6.2)	9.3 (7.0)	6.6 (5.5)	6.2 (5.8)	4.7 (7.8)	2.1 (7.0)
4000 Hz	7.5 (5.4)	6.5 (5.4)	0.7 (4.9)	1.3 (5.1)	6.5 (5.0)	5 (4.8)

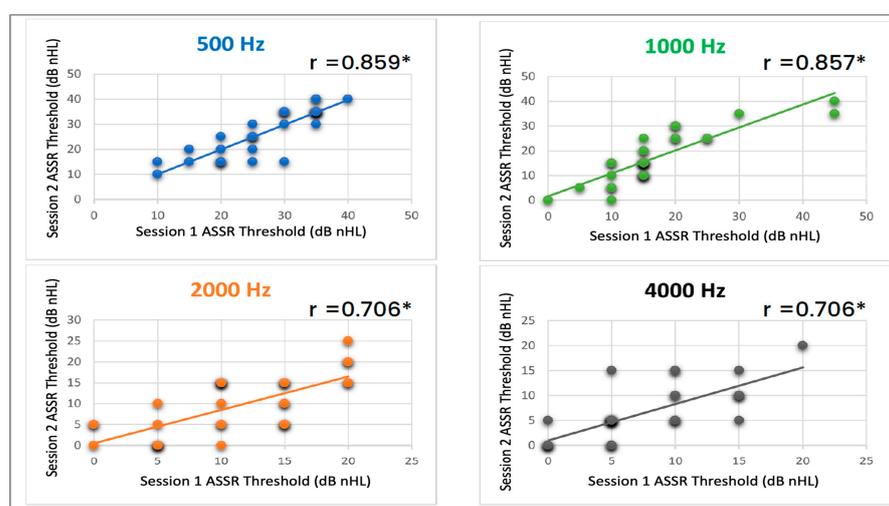


Figure 3. Correlation of ASSR thresholds for each subject between session one and session two at different carrier frequencies. The top left graph (blue) represents the correlation for 500 Hz, the top right graph (green) shows the correlation for 1000 Hz, the bottom left graph (orange) represents the correlation for 2000 Hz, and the bottom right graph (black) displays the correlation for 4000 Hz. The asterisks indicate that session one and session two were statistically significant ($p < 0.01$).

The ASSR thresholds between session one and session two were compared using the Pearson product-moment correlation coefficient to assess the level of correlation. The analysis revealed a highly significant correlation ($p < 0.01$) between the two sessions for all carrier frequencies examined: 500, 1000, 2000, and 4000 Hz. The individual panels in Figure 3 also display the corresponding R^2 values, which indicate the strength of the correlation between session one and session two for each carrier frequency. The R^2 values demonstrate a robust and consistent association between the ASSR thresholds obtained in sessions one and two across the carrier frequencies tested. The high correlation observed between the two sessions reinforces the consistency and stability of the ASSR thresholds, further highlighting the reliability of ASSR for evaluating auditory responses at different carrier frequencies.

3.3. Test–Retest Threshold Differences across Frequencies for the ASSR

Figure 4 and Table A1 illustrate the average differences in ASSR thresholds between session one and session two across various carrier frequencies. Among the four tested frequencies, the highest mean difference was observed at 2000 Hz, while the lowest mean difference was found at 500 Hz, with values of -1.6667 and -0.167 , respectively. To examine the impact of carrier frequency on the difference in ASSR thresholds between sessions, a single-factor analysis of variance (ANOVA) was conducted. The results indicated

that the carrier frequency did not have a significant effect on the differences in ASSR thresholds between session two and session one ($p = 0.4$). This implies that the differences in ASSR threshold between sessions are similar across frequencies.

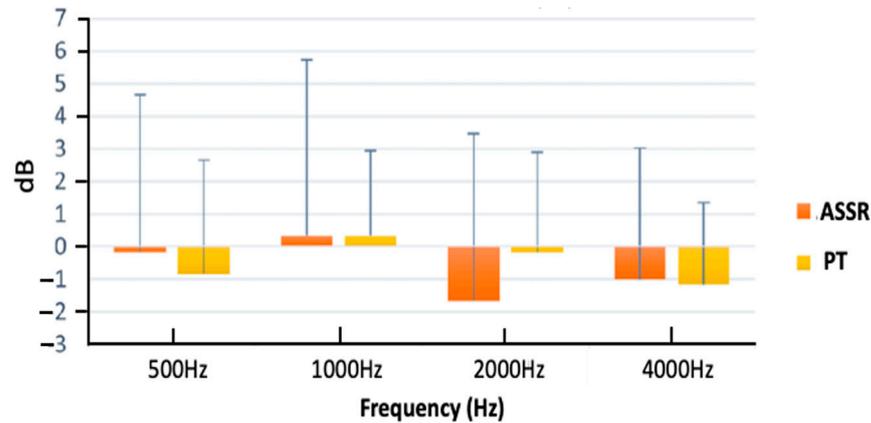


Figure 4. Mean differences in thresholds between sessions for ASSR and pure-tone thresholds across frequencies ranging from 500 to 4000 Hz. Error bars represent 1 standard deviation.

3.4. Comparison of ASSR Thresholds Differences between Sessions versus Pure-Tone Thresholds Differences between Sessions

Figure 4 and Table A1 display the mean threshold differences between sessions for ASSR and pure-tone thresholds at each carrier frequency. Notably, there was greater variability in ASSR thresholds between sessions compared to pure-tone thresholds. However, the mean values showed remarkable similarities in the variability between the two testing methods. In a two-way ANOVA analysis, there was no significant main effect of frequency, threshold method (ASSR and pure-tone thresholds), or the interaction between frequency and threshold method. These findings indicate that the test–retest variability is comparable between ASSR and pure-tone thresholds, highlighting the reliability and consistency of the ASSR threshold.

The SEM was calculated separately for each frequency, encompassing both behavioral pure-tone thresholds and ASSR thresholds (Table 2). For both types of measurements, the ± 1 SEM results indicate that the thresholds fall within a range of ± 5 dB. Additionally, the ± 2 SEM results were provided as they offer a 95% probability that the threshold would be repeated within this range. The 2 SEM for behavioral pure-tone thresholds is within ± 5 dB HL, while for ASSR recordings, it is within ± 10 dB nHL. It is worth noting that the overall intrasubject variability is higher for ASSR compared to pure-tone thresholds. Combining the findings from Figure 4, Tables 2 and A1, it can be inferred that the mean difference between the test and retest sessions is similar for both ASSR and pure-tone thresholds. However, the intra-subject variability is higher for ASSR than pure-tone thresholds.

Table 2. ± 1 SEM and ± 2 SEM values for ASSR and behavioral pure-tone thresholds.

Frequency (Hz)	ASSR (dB nHL)		Behavioral Pure-Tone Threshold (dB HL)	
	± 1 SEM	± 2 SEM	± 1 SEM	± 2 SEM
500 Hz	3.4	6.7	2.4	4.7
1000 Hz	3.8	7.5	1.9	3.7
2000 Hz	3.6	7.1	2.1	4.2
4000 Hz	2.8	5.6	1.8	3.5

3.5. Comparison of Residual Noise Levels at the Threshold to the Difference in the ASSR Thresholds across Sessions

The difference in residual noise levels at the threshold between sessions was determined by subtracting the residual noise values from session two from those of session

one for each participant. This difference in residual noise levels at the threshold was then compared to the difference in ASSR thresholds to investigate whether variations in residual noise levels contributed to the test–retest variability of ASSR measurements. The correlation analysis between the difference in residual noise levels at the threshold and the difference in ASSR thresholds between sessions did not demonstrate statistical significance ($p = 0.3$) (see Figure 5). These findings suggest that differences in residual noise levels at the threshold between test–retest sessions did not significantly contribute to the observed differences in ASSR thresholds between sessions.

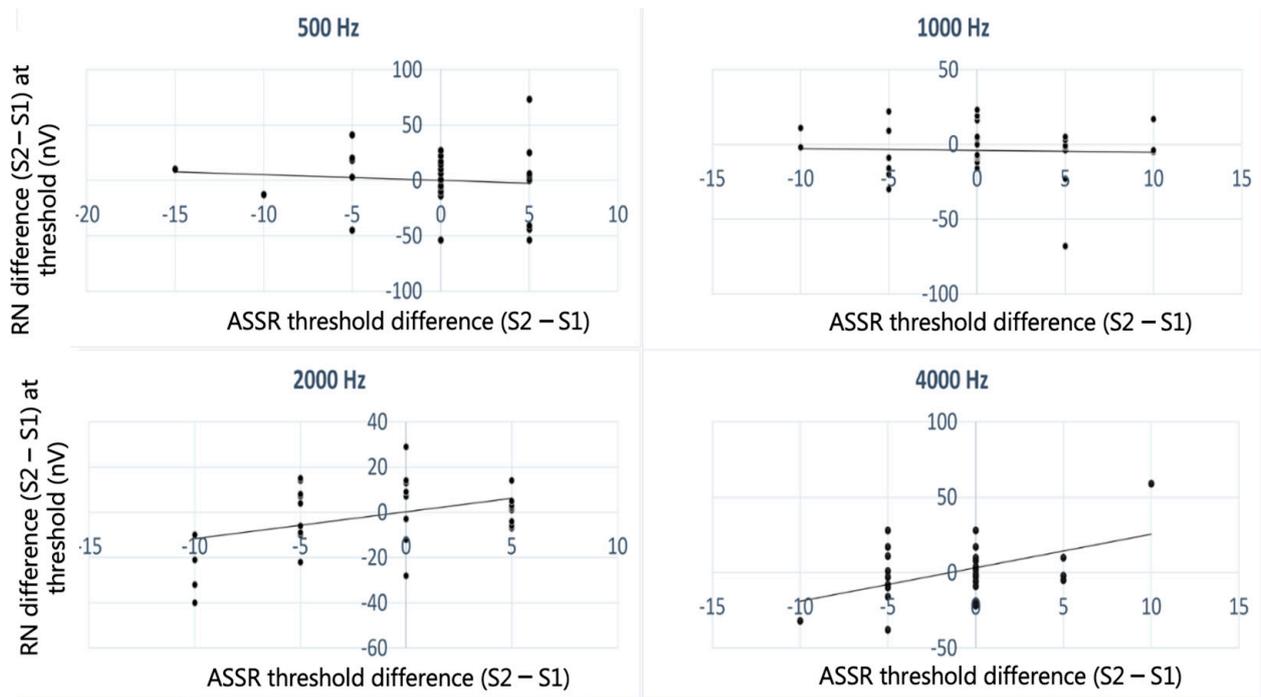


Figure 5. Scatter plot showing the differences in noise levels at threshold (Y-axis) plotted against the differences in air conduction ASSR thresholds across sessions (X-axis) for frequencies ranging from 500 Hz to 4000 Hz.

4. Discussion

The primary objective of this pilot study was to evaluate the test–retest reliability of the next-generation ASSR system using the multiple ASSR technique in the Eclipse system among normal-hearing adults. The results indicated strong test–retest reliability for ASSR across all tested frequencies, including 500, 1000, 2000, and 4000. Notably, the highest reliability was observed at 500 Hz, which differs from previous studies that studied the test–retest reliability using the old-generation ASSR [5,17,18]. Additionally, no significant correlation was found between the difference in ASSR residual noise levels at the threshold and the difference in ASSR thresholds between sessions across all frequencies. Although the mean test–retest variability of ASSR was similar to behavioral pure-tone thresholds, the ± 2 SEM (95% probability that the threshold would be repeated within this range) values were within a range of 5 dB for behavioral pure-tone thresholds and 10 dB for ASSR thresholds, suggesting that the intra-subject variability is higher for ASSR than for pure-tone thresholds.

4.1. ASSR Test–Retest Reliability

In the present study, the Pearson correlation coefficients for the air conduction ASSR thresholds ranged from 0.706 to 0.859 across the four center frequencies (500–4000 Hz), with 2000 Hz showing the weakest and 500 Hz showing the strongest correlation. The correlation between the test–retest ASSR thresholds was statistically significant across all

frequencies. These findings indicate a strong relationship between the thresholds obtained in sessions one and two. Luts and Wouters [5], Kaf et al. [18], and D'haenens et al. [17] reported moderately strong test–retest reliability for ASSR at 1000, 2000, and 4000 Hz and weaker test–retest reliability at 500 Hz. Overall, the findings of the current study are similar to previous studies using old-generation ASSR for 1000 Hz, 2000 Hz, and 4000 Hz, whereas improvement in test–retest reliability from weak to strong reliability is observed for 500 Hz.

The improved test–retest reliability observed with the new-generation ASSR can be attributed to several factors. Firstly, the utilization of a narrow-band CE-chirp stimulus in this new system generates larger amplitude responses even at lower intensity levels [28,29]. However, the Eclipse clinical system does not provide the amplitude of ASSR, preventing a direct comparison between the amplitudes of the old- and new-generation ASSR. Therefore, an interesting future investigation would involve comparing the amplitudes of ASSR using the old-generation amplitude or mixed modulated stimuli with the chirp family stimuli of the next-generation ASSR across different intensity levels. This data would provide insights into the contributing factors that lead to improved thresholds and test–retest reliability observed with the next-generation ASSR. A second contributing factor to the enhanced test–retest reliability lies in the differences in the nature of ASSR analysis between the old- and new-generation systems. The old-generation ASSR relied on amplitude or phase analysis solely at the modulation frequency [1,2], whereas the Eclipse system incorporates a multi-harmonic detector analysis that examines both the amplitude and phase of the fundamental frequency (repetition rate) as well as higher harmonics (up to 12 harmonics) [3,4]. This multi-harmonic statistical analysis likely contributes to a more reliable threshold determination between sessions.

The third factor pertains to the threshold determination criteria. Previous studies assessing ASSR test–retest reliability typically employed a criterion of $p \leq 0.05$ to determine the presence of ASSR [5,17,18]. However, this study utilized a more stringent criterion of $p \leq 0.01$. This strict criterion may have led to reduced residual noise levels and a consistent ASSR threshold between sessions, thereby contributing to improved test–retest reliability.

4.2. Relationship between Residual Noise Levels at the Threshold and ASSR Threshold Difference between Sessions

The residual noise level in a recording is influenced by factors such as baseline EEG activity, muscle noise, and the test duration. Through the use of narrow-band filtering and spectral averaging techniques, background noise with components at all frequencies can be filtered out, leading to an improvement in the signal-to-noise ratio. To investigate whether residual noise levels recorded during the testing could be responsible for the observed test–retest differences, the differences in residual noise levels at the threshold between session two and session one was compared to the differences in thresholds between session two and session one for each carrier frequency. The data analysis revealed no significant correlation between the variability in residual noise levels and the variability in thresholds across sessions for frequencies ranging from 500 Hz to 4000 Hz. There are no directly comparable published data regarding the relationship between differences in residual noise levels and threshold differences. However, a previous study conducted by Wilding et al. [20] compared the amplitude of the ASSR measured in dB levels referenced to adjacent noise bins. They reported high variability in the ASSR amplitude between test sessions, which affected test–retest consistency. Wilding et al. [20] further discussed that the variability in the amplitude of the ASSR is related to the variability in noise levels, as the amplitude of the ASSR is computed with reference to the noise levels in adjacent bins. Based on the findings of this study, it can be concluded that the test–retest variability observed in ASSR thresholds is not influenced by variations in residual noise levels.

4.3. Test–Retest Reliability Comparison between Pure-Tone Thresholds and ASSR Thresholds

The findings of the current study suggest that the mean test–retest variability for ASSR and behavioral pure-tone thresholds is comparable. However, there is greater variability

observed in ASSR thresholds compared to behavioral pure-tone thresholds. For behavioral pure-tone thresholds, the ± 2 SEM range (representing a 95% probability that the threshold will fall within this range) is within ± 5 dB HL, whereas for ASSR recordings, it is within ± 10 dB nHL. These higher ± 2 SEM values for ASSR compared to pure-tone threshold results align with the findings reported by D'haenens et al. [17], although the comparison is not explicitly discussed in this study. The data from D'haenens et al. [17] suggest similar mean test–retest variability between ASSR and behavioral pure-tone thresholds. However, there are differences in the ± 2 SEM values between D'haenens et al. [17] and the current study, with D'haenens et al. [17] showing nearly double the values for both ASSR and behavioral pure-tone thresholds compared to the present research. The reasons for these higher ± 2 SEM values of ± 10 dB HL for behavioral pure-tone thresholds reported by D'haenens et al. [17] are uncertain. It is possible that the higher ± 2 SEM values of ± 20 dB HL for ASSR, reported by D'haenens et al. [17], could be attributed to the use of old-generation ASSR, which produces lower amplitude responses, and the utilization of amplitude at modulation frequency for analysis. Wilding et al. [20] and Israelsson et al. [10] have also reported such high variabilities in ASSR thresholds due to fluctuations in amplitude and noise levels of the ASSR responses.

It is widely acknowledged that pure-tone audiometry is considered the gold standard for assessing changes in hearing over time, and numerous research studies have attempted to study the test–retest reliability of regular and extended high-frequency pure-tone audiometry [29–32]. The SEM reported in this study is comparable to previous research [29–32]. The findings of the current study also indicate that ASSR can be a reliable alternative for populations that cannot provide reliable behavioral pure-tone thresholds. The results of this study have significant implications for clinicians and researchers who work with individuals with hearing impairments. The findings suggest that ASSR can be a valuable tool for monitoring hearing changes in individuals who cannot provide reliable behavioral responses, such as infants, young children, or patients with severe cognitive or motor disabilities. With the use of ASSR, clinicians, and researchers can potentially gain greater insight into the progression of hearing loss and develop appropriate intervention strategies to mitigate its effects. Overall, the study's findings offer promising support for the use of ASSR in monitoring hearing changes over time. However, further research is needed to validate these results in a larger population of normal-hearing and hearing-impaired individuals.

4.4. Limitations and Future Research

The current pilot study had certain limitations that warrant consideration for future research. Firstly, the sample size of the study comprised only 15 adults (30 ears) with normal hearing. This falls below the recommended threshold of 40 participants for a test–retest study, as suggested by McMillan and Hanson [33]. To gain a comprehensive understanding of the reliability of ASSR, future studies should include a larger and more diverse population, encompassing both individuals with normal hearing and those with hearing impairments. This would provide the opportunity for a broader assessment of the ASSR system's performance across different hearing profiles.

Additionally, it is essential to investigate the differences between pediatric and adult populations to better understand the reliability of ASSR in each group. Children possess distinct physiological characteristics and developmental profiles, which may influence the reliability of ASSR measurements. One aspect to consider is the variance in baseline EEG levels between children and adults. Children's baseline EEG levels may differ from those of adults. These variations could potentially impact the reliability of ASSR measurements, particularly when using the fast repetition rate ASSR implemented in the Eclipse system. Therefore, it is imperative to conduct future research specifically aimed at comparing the test–retest reliability of ASSR between pediatric and adult populations. By examining age-related differences, researchers can determine the influence of age-related factors on the reliability of ASSR measurements. Moreover, the adaptive properties of the auditory system may differ between children and adults, especially in response to the fast repetition

rate utilized in the Eclipse system. Children's auditory systems exhibit ongoing maturation and plasticity, which could affect their response to auditory stimuli and, subsequently, the reliability of ASSR measurements. Therefore, a focused investigation comparing the test–retest reliability of ASSR between pediatric and adult populations is warranted. This research would provide valuable insights into any age-related differences and enable a more accurate evaluation of the reliability of ASSR in each group.

Significant variations can be observed in the outcomes obtained from different commercially available ASSR acquisition systems, as pointed out by Hatzopoulos et al. [34]. Consequently, it is crucial to conduct a comparative analysis between the results obtained from the next-generation ASSR implemented in Eclipse and the older-generation ASSR systems, as well as other commercially available systems. Another limitation of the study is the lack of availability of amplitude and phase coherence information for ASSR measurements in the Eclipse clinical system. This limitation prevented direct comparisons between the amplitudes and phase coherence information of the older and newer generation ASSR systems. To address this limitation, an intriguing avenue for future investigation would involve comparing the amplitudes of ASSR using the older-generation amplitude or mixed modulated stimuli with the chirp family stimuli of the next-generation ASSR across various intensity levels. Such research would provide valuable insights into the specific factors contributing to the improved thresholds and test–retest reliability observed with the next-generation ASSR.

In summary, future studies should aim to overcome the limitations of the current research by including larger and more diverse participant samples, exploring pediatric-adult differences, and conducting comparative investigations on the amplitude of ASSR using different stimuli. By addressing these areas, researchers can further enhance our understanding of the reliability and potential applications of the next-generation ASSR system.

5. Conclusions

The objective of this pilot study was to examine the test–retest reliability of the next-generation ASSR system implemented in the Eclipse system. The research evaluated the consistency of ASSR thresholds across multiple sessions and found strong test–retest reliability across all carrier frequencies. Notably, the carrier frequency of 500 Hz demonstrated the highest test–retest reliability, which was unexpected considering previous studies on the reliability of older-generation ASSR systems. Furthermore, the study compared the reliability of ASSR thresholds with behavioral pure-tone thresholds and found that they exhibited similar mean test–retest reliability. The ± 2 SEM values, representing a 95% probability of the threshold falling within this range, were within 5 dB for behavioral pure-tone thresholds and 10 dB for ASSR thresholds. These results suggest that the intra-subject variability is higher for ASSR compared to pure-tone thresholds. While pure-tone audiometry is widely regarded as the gold standard for monitoring changes in hearing over time, the study's findings indicate that ASSR can be a reliable alternative, particularly for individuals who are unable to provide consistent behavioral pure-tone thresholds. In conclusion, the study provides promising evidence supporting the use of ASSR in monitoring changes in hearing over time. The strong test–retest reliability observed in the next-generation ASSR system, particularly at 500 Hz, highlights its potential as a valuable tool in clinical settings for populations with limitations in providing reliable behavioral pure-tone thresholds.

Author Contributions: Conceptualization, H.H. and C.H.S.; methodology, H.H. and C.H.S.; formal analysis, H.H. and C.H.S.; investigation, H.H. and C.H.S.; data curation, H.H.; writing—original draft preparation, H.H., N.J.W. and C.H.S.; writing—review and editing, H.H., N.J.W. and C.H.S.; visualization, H.H., N.J.W. and C.H.S.; supervision, C.H.S.; project administration, C.H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of California State University- Los Angeles (protocol code 1836963 and date of approval January 2022).

Informed Consent Statement: Informed written consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: Special thanks to Margaret Winter, Miwako Hisagi, and Peter Ivory for providing feedback on all aspects of this project.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The average difference in the threshold between sessions for ASSR and pure-tone thresholds.

Frequency (Hz)	Difference in the Threshold between Sessions (dB)	
	ASSR Mean (SD)	Pure-Tone Thresholds Mean (SD)
500 Hz	−0.2 (4.8)	−0.8 (3.5)
1000 Hz	0.3 (5.4)	0.3 (2.6)
2000 Hz	−1.7 (5.1)	−0.2 (3.1)
4000 Hz	−1 (4.2)	−1.2 (2.5)

References

- Korczak, P.; Smart, J.; Delgado, R.; Strobel, T.M.; Bradford, C. Auditory Steady-State Responses. *J. Am. Acad. Audiol.* **2012**, *23*, 146–170. [[CrossRef](#)] [[PubMed](#)]
- Picton, T.W.; Dimitrijevic, A.; John, M.S. Multiple Auditory Steady-State Responses. *Ann. Otol. Rhinol. Laryngol.* **2002**, *111*, 16–21. [[CrossRef](#)] [[PubMed](#)]
- Sininger, Y.S.; Hunter, L.L.; Roush, P.A.; Windmill, S.; Hayes, D.; Uhler, K.M. Protocol for Rapid, Accurate, Electrophysiologic, Auditory Assessment of Infants and Toddlers. *J. Am. Acad. Audiol.* **2020**, *31*, 455–468. [[CrossRef](#)] [[PubMed](#)]
- Sininger, Y.S.; Hunter, L.L.; Hayes, D.; Roush, P.A.; Uhler, K.M. Evaluation of Speed and Accuracy of Next-Generation Auditory Steady State Response and Auditory Brainstem Response Audiometry in Children With Normal Hearing and Hearing Loss. *Ear Hear.* **2018**, *39*, 1207–1223. [[CrossRef](#)]
- Luts, H.; Wouters, J. Comparison of MASTER and AUDERA for Measurement of Auditory Steady-State Responses. *Int. J. Audiol.* **2005**, *44*, 244–253. [[CrossRef](#)]
- Herdman, A.T.; Stapells, D.K. Auditory Steady-State Response Thresholds of Adults with Sensorineural Hearing Impairments: Umbrales de las respuestas auditivas de estado estable en adultos con hipoacusia sensorineural. *Int. J. Audiol.* **2003**, *42*, 237–248. [[CrossRef](#)]
- Van Maanen, A.; Stapells, D.R. Multiple-ASSR Thresholds in Infants and Young Children with Hearing Loss. *J. Am. Acad. Audiol.* **2010**, *21*, 535–545. [[CrossRef](#)]
- Rance, G.; Roper, R.; Symons, L.; Moody, L.-J.; Poullis, C.; Dourlay, M.; Kelly, T. Hearing Threshold Estimation in Infants Using Auditory Steady-State Responses. *J. Am. Acad. Audiol.* **2005**, *16*, 291–300. [[CrossRef](#)]
- Attias, J.; Karawani, H.; Shemesh, R.; Nageris, B. Predicting Hearing Thresholds in Occupational Noise-Induced Hearing Loss by Auditory Steady State Responses. *Ear Hear.* **2014**, *35*, 330–338. [[CrossRef](#)]
- Israelsson, K.-E.; Bogo, R.; Berninger, E. Reliability in Hearing Threshold Prediction in Normal-Hearing and Hearing-Impaired Participants Using Mixed Multiple ASSR. *J. Am. Acad. Audiol.* **2015**, *26*, 299–310. [[CrossRef](#)]
- Rodrigues, G.R.I.; Lewis, D.R. Threshold Prediction in Children with Sensorineural Hearing Loss Using the Auditory Steady-State Responses and Tone-Evoked Auditory Brain Stem Response. *Int. J. Pediatr. Otorhinolaryngol.* **2010**, *74*, 540–546. [[CrossRef](#)]
- Michel, F.; Jørgensen, K.F. Comparison of Threshold Estimation in Infants with Hearing Loss or Normal Hearing Using Auditory Steady-State Response Evoked by Narrow Band CE-Chirps and Auditory Brainstem Response Evoked by Tone Pips. *Int. J. Audiol.* **2017**, *56*, 99–105. [[CrossRef](#)]
- Ehrmann-Müller, D.; Shehata-Dieler, W.; Alzoubi, A.; Hagen, R.; Cebulla, M. Using ASSR with narrow-band chirps to evaluate hearing in children and adults. *Eur. Arch. Otorhinolaryngol.* **2021**, *278*, 49–56. [[CrossRef](#)]
- Vanaja, C.S.; Kunjir, A. Auditory Steady-State Responses to Narrow-Band Chirps in Predicting Aided Behavioral Thresholds. *J. All India Inst. Speech Hear.* **2021**, *40*, 31–38. [[CrossRef](#)]

15. Campbell, K.C.M.; Le Prell, C.G. Drug-Induced Ototoxicity: Diagnosis and Monitoring. *Drug Saf.* **2018**, *41*, 451–464. [[CrossRef](#)]
16. Barreira-Nielsen, C.; Fitzpatrick, E.; Hashem, S.; Whittingham, J.; Barrowman, N.; Aglipay, M. Progressive Hearing Loss in Early Childhood. *Ear Hear.* **2016**, *37*, e311–e321. [[CrossRef](#)]
17. D’Haenens, W.; Vinck, B.M.; De Vel, E.; Maes, L.; Bockstael, A.; Keppler, H.; Philips, B.; Swinnen, F.; Dhooge, I. Auditory Steady-State Responses in Normal Hearing Adults: A Test–Retest Reliability Study. *Int. J. Audiol.* **2008**, *47*, 489–498. [[CrossRef](#)]
18. Kaf, W.A.; Sabo, D.L.; Durrant, J.D.; Rubinstein, E. Reliability of Electric Response Audiometry Using 80 Hz Auditory Steady-State Responses. *Int. J. Audiol.* **2006**, *45*, 477–486. [[CrossRef](#)]
19. Kim, S.-R.; Lee, J. Test–Retest Reliability of Bone-Conducted Auditory Steady-State Response. *Audiology* **2010**, *6*, 50–54. [[CrossRef](#)]
20. Wilding, T.S.; McKay, C.M.; Baker, R.J.; Kluk, K. Auditory Steady State Responses in Normal-Hearing and Hearing-Impaired Adults: An Analysis of between-Session Amplitude and Latency Repeatability, Test Time, and F Ratio Detection Paradigms. *Ear Hear.* **2012**, *33*, 267–278. [[CrossRef](#)]
21. McFadden, K.L.; Steinmetz, S.E.; Carroll, A.M.; Simon, S.T.; Wallace, A.; Rojas, D.C. Test–Retest Reliability of the 40 Hz EEG Auditory Steady-State Response. *PLoS ONE* **2014**, *9*, e85748. [[CrossRef](#)] [[PubMed](#)]
22. Carhart, R.; Jerger, J.F. Preferred Method For Clinical Determination Of Pure-Tone Thresholds. *J. Speech Hear. Disord.* **1959**, *24*, 330–345. [[CrossRef](#)]
23. Elberling, C.; Don, M. A Direct Approach for the Design of Chirp Stimuli Used for the Recording of Auditory Brainstem Responses. *J. Acoust. Soc. Am.* **2010**, *128*, 2955–2964. [[CrossRef](#)] [[PubMed](#)]
24. Ferm, I.; Lightfoot, G. Further Comparisons of ABR Response Amplitudes, Test Time, and Estimation of Hearing Threshold Using Frequency-Specific Chirp and Tone Pip Stimuli in Newborns: Findings at 0.5 and 2 KHz. *Int. J. Audiol.* **2015**, *54*, 745–750. [[CrossRef](#)]
25. Rodrigues, G.R.I.; Ramos, N.; Lewis, D.R. Comparing Auditory Brainstem Responses (ABRs) to Toneburst and Narrow Band CE-Chirp in Young Infants. *Int. J. Pediatr. Otorhinolaryngol.* **2013**, *77*, 1555–1560. [[CrossRef](#)]
26. Kristensen, S.G.B.; Elberling, C. Auditory Brainstem Responses to Level-Specific Chirps in Normal-Hearing Adults. *J. Am. Acad. Audiol.* **2012**, *23*, 712–721. [[CrossRef](#)]
27. Hopkins, W. Measures of reliability in sports medicine and science. *Sports Med.* **2000**, *30*, 1–15. [[CrossRef](#)]
28. Cobb, K.; Stuart, A. Neonate Auditory Brainstem Responses to CE-Chirp and CE-Chirp Octave Band Stimuli I: Versus Click and Tone Burst Stimuli. *Ear Hear.* **2016**, *37*, 710–723. [[CrossRef](#)]
29. Jerger, J. Comparative evaluation of some auditory measures. *J. Speech Hear. Res.* **1962**, *5*, 3–17. [[CrossRef](#)]
30. Henry, J.A.; Flick, C.L.; Gilbert, A.; Ellingson, R.M.; Fausti, S.A. Reliability of hearing thresholds: Computer- automated testing with ER-4B Canal Phone earphones. *J. Rehabil. Res. Dev.* **2001**, *38*, 567–581.
31. Schmuziger, N.; Probst, R.; Smurzynski, J. Test–retest reliability of pure-tone thresholds from 0.5 to 16 kHz using Sennheiser HDA 200 and Etymotic Research ER-2 earphones. *Ear Hear.* **2004**, *25*, 127–132. [[CrossRef](#)]
32. Stelmachowicz, P.G.; Beauchaine, K.A.; Kalberer, A.; Langer, T.; Jesteadt, W. The reliability of auditory thresholds in the 8- to 20-kHz range using a prototype audiometer. *J. Acoust. Soc. Am.* **1988**, *83*, 1528–1535. [[CrossRef](#)]
33. McMillan, G.P.; Hanson, T.E. Sample size requirements for establishing clinical test–retest standards. *Ear Hear.* **2014**, *35*, 283–286. [[CrossRef](#)]
34. Hatzopoulos, S.; Petruccioli, J.; Śliwa, L.; Jędrzejczak, W.W.; Kochanek, K.; Skarżyński, H. Hearing threshold prediction with Auditory Steady State Responses and estimation of correction functions to compensate for differences with behavioral data, in adult subjects. Part 1: Audera and CHARTR EP devices. *Med. Sci. Monit.* **2012**, *18*, MT47–MT53. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.