J Hear Sci, 2020; 10(2): 55–59
DOI: 10.17430/JHS.2020.10.2.6

TEST-RETEST RELIABILITY OF VARIOUS PSYCHOACOUSTIC MEASURES USING THE MAXIMUM LIKELIHOOD PROCEDURE

Contributions:
A Study design/planning
B Data collection/entry
C Data analysis/statistics
D Data interpretation
E Preparation of manuscript
F Literature analysis/search
G Funds collection

Chandni Jain[A,C-F], **Kirti Joshi**[B-C,F]

Audiology, All India Institute of Speech and Hearing, India

Corresponding author: Chandni Jain, Audiology, All India Institute of Speech and Hearing, Naimisham Campus, 570006, Mysuru, India; email: chandni.aud@gmail.com, Phone: +918212502358

Abstract

Objective: Psychoacoustic abilities include the perception of frequency, intensity, and the temporal parameters of sound, which are necessary for speech perception. These abilities can be measured using adaptive and non-adaptive procedures. The maximum likelihood procedure (MLP) is an adaptive psychophysical procedure that is usually utilized in psychoacoustic tasks. The present study aimed to assess the test-retest reliability of various psychoacoustic measures assessed using the MLP toolbox implemented in Matlab.

Material and methods: A total of 20 participants with normal hearing sensitivity were selected for the study. The test–retest reliability of psychoacoustic measures was studied in terms of frequency difference limen, intensity difference limen, duration discrimination thresholds, gap detection thresholds, and modulation detection thresholds (4 Hz and 128 Hz). To check test–retest reliability all measures were assessed twice within a span of one day.

Results: The results showed no significant difference in the various psychoacoustic measures obtained across the two sessions. The reliability of each measure was tested using the intraclass correlation coefficient (ICC).The results indicated that the test-retest reliability of various psychoacoustic measures assessed using the MLP toolbox was fair to good.

Conclusions: It can be concluded that MLP is a reliable tool to assess various psychoacoustic abilities.

Key words: intensity • frequency • temporal • MLP

POWTARZALNOŚĆ TEST–RETEST RÓŻNYCH POMIARÓW PSYCHOAKUSTYCZNYCH PRZY UŻYCIU PROCEDURY NAJWIĘKSZEGO PRAWDOPODOBIEŃSTWA

Streszczenie

Wstęp: Zdolności psychoakustyczne obejmują: percepcję częstotliwości, natężenia i czasowych parametrów dźwięku, które są niezbędne do percepcji mowy. Zdolności te można mierzyć za pomocą procedur adaptacyjnych i nieadaptacyjnych. Procedura największego prawdopodobieństwa (MLP) to adaptacyjna procedura psychofizyczna, która jest zwykle stosowana w testach psychoakustycznych. Niniejsze badanie miało na celu ocenę powtarzalności test-retest różnych pomiarów psychoakustycznych określanych za pomocą zestawu narzędzi MLP wdrożonych w Matlabie.

Materiał i metody: W badaniu wzięło udział łącznie 20 uczestników z normalnym słuchem. Powtarzalność pomiarów psychoakustycznych została zbadana w odniesieniu do: progów różnicowania częstotliwości, progów różnicowania natężenia, progów dyskryminacji czasu trwania, progów wykrywania przerw i progów wykrywania modulacji (4 Hz i 128 Hz). W celu oceny powtarzalności test-retest wszystkie pomiary zostały ocenione dwukrotnie w ciągu jednego dnia.

Wyniki: Wyniki nie wykazały znaczącej różnicy w badanych pomiarach psychoakustycznych uzyskanych podczas dwóch sesji. Powtarzalność każdego pomiaru została przetestowana przy użyciu współczynnika korelacji wewnątrzklasowej (ICC). Wyniki wskazują, że rzetelność testu-retestu w odniesieniu do różnych pomiarów psychoakustycznych, ocenianych przy użyciu zestawu narzędzi MLP, była zadowalająca lub dobra.

Wnioski: Można stwierdzić, że MLP jest wiarygodnym narzędziem do oceny różnych zdolności psychoakustycznych.

Słowa kluczowe: intensywność • częstotliwość • tymczasowość • MLP

Introduction

Psychoacoustics is a branch of psychology that deals with understanding the perception of sound in the auditory system. Psychoacoustic abilities involve the perception of frequency, intensity, and the temporal parameters of sound that are important for speech perception [1,2]. The assessment of these abilities has generated strong interest among researchers in the field of audiology, psychology, and acoustics. Several studies have been carried out to assess psychoacoustical abilities in persons with normal hearing, and the findings have been compared with aging individuals, individuals with hearing impairment, auditory neuropathy, and auditory processing disorder, to name a few [3,4]. In these studies psyshoacoustic abilities were assessed using the maximum likelihood procedure (MLP); however there is no mention of its test–retest reliability.

Psychoacoustic research has been mainly done using various procedures, including adaptive and non-adaptive procedures. In adaptive procedures, the stimulus presented to the listener at each trial depends on their response to the previous trial [5]. In non-adaptive procedures, the stimulus is presented in a pre-set format, which is decided before the commencement of the actual experiment [6]. Studies have been done in the past where, to assess psychoacoustic abilities, researchers have used different adaptive procedures, namely simple up-down procedure/staircase procedure, transformed up-down procedure, MLP, parameter estimation by sequential testing (PEST), and so on [6].

MLP is an adaptive psychoacoustic procedure that is commonly used to measure various psychoacoustic abilities [7,8]. In this procedure, several psychometric functions, named hypotheses, are estimated. The highest likelihood of the obtained response, which is like the actual listener's psychometric function, is estimated trial by trial using the maximum likelihood algorithm. The most likely hypothesis in MLP is the threshold of the participant. The MLP tracks any point of the psychometric function and estimates threshold using either yes/no experiments or alternate force choice (nAFC). In a yes/no experiment, the participant is presented with a series of different stimulus levels and is asked whether he or she has detected the stimulus (yes) or not (no). In an nAFC task, the participant is presented with a series of *n* stimuli differing in level where one stimulus (the variable) changes its level across the trials, whereas the level of the others (the standards) is fixed. After each trial, the subject is asked to report which was the variable stimulus. It is reported that within 12 trials, the MLP usually reaches a reasonably stable estimation of the most probable psychometric function, which can be used to approximate thresholds [9,10].

This procedure has been widely used to assess psychoacoustical abilities and is claimed to have good reliability and validity [11]; however, no explicit measure to study the reliability of MLP has been reported. There is therefore a need to assess the test–retest reliability of MLP for various psychoacoustic measures. The psychoacoustic abilities tested here involve the perception of frequency, intensity, and the temporal parameters of sound. The present study aimed to assess the test–retest reliability of frequency difference limen (FDL), intensity difference limen (IDL), duration discrimination threshold (DDT), gap detection threshold (GDT), and modulation detection threshold (MDT) at 4 Hz and 128 Hz using the MLP toolbox implemented in Matlab.

Material and Methods

Participants

A total of 20 participants (6 males, 14 females) aged 17–28 years were selected for the study [12]. A routine audiological evaluation was done in an acoustically treated

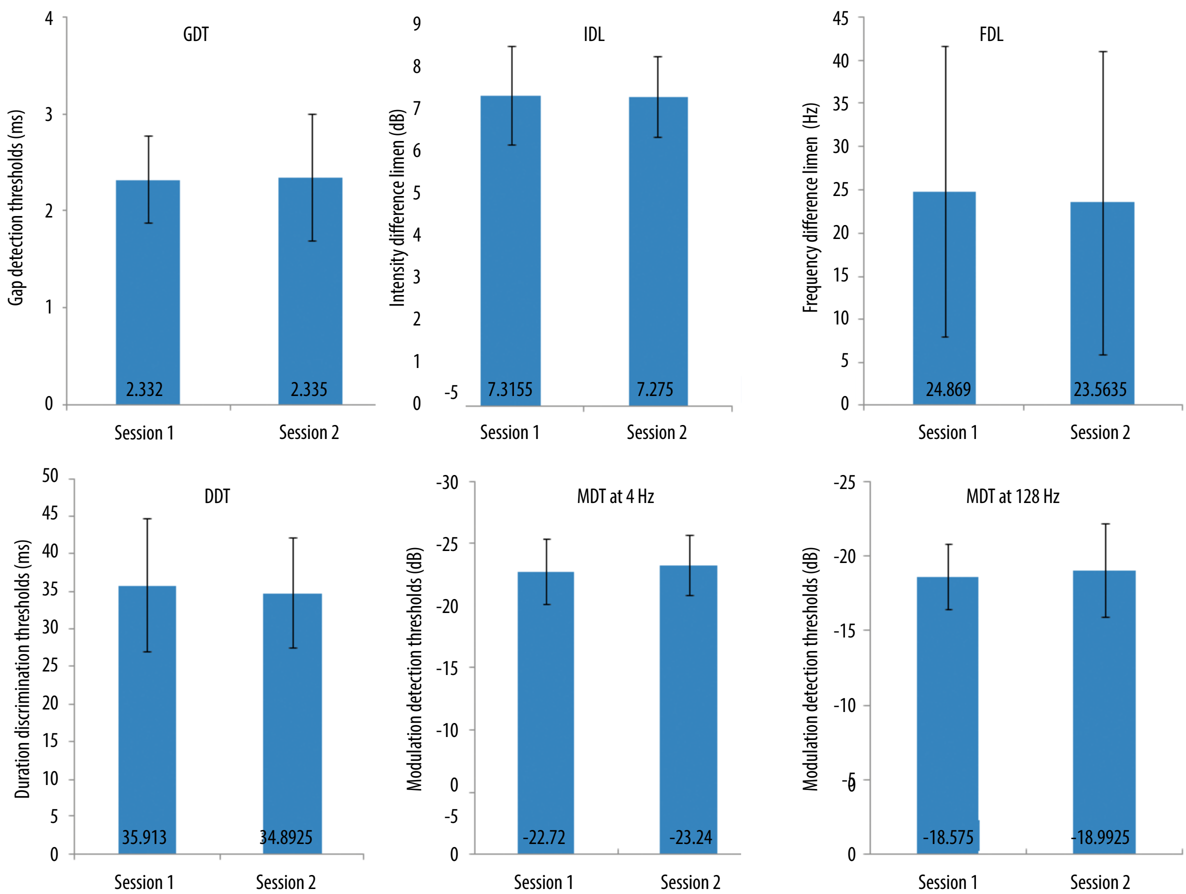

Figure 1. Mean and standard deviation of various psychoacoustic measures across two sessions

air-conditioned room, including pure-tone audiometry, speech audiometry, and immittance evaluation. All the participants had bilateral normal hearing sensitivity, defined as: pure-tone thresholds of less than 15 dB HL for octave frequencies ranging from 250 Hz to 8000 Hz for air conduction and 250 Hz to 4000 Hz for bone conduction; type A tympanogram in immittance evaluation; and acoustic reflexes present for 500, 1000, and 2000 Hz. Participants were selected using a purposively convenient sampling technique. None of the participants reported any illness during testing.

Procedure

The stimuli for all the psychoacoustic tests were generated through the MLP toolbox implemented in MATLAB version 2010a (Mathworks Inc., Natick, MA) [9,10]. All the tests were performed using a three-interval alternate forced-choice adaptive technique to estimate a 79.4% response criterion. Each test trial consisted of three blocks, in which two blocks had the standard stimulus, and the third block had the variable stimulus, which was chosen randomly in MLP. The participant was asked to identify the block which contained the variable stimulus. All the test stimuli were presented at 60 dB SPL binaurally, which was routed through the audiometer. The detailed procedure for each test is described below.

FDL. The minimum frequency difference requires the subject to discriminate two closely spaced frequencies. FDL was measured for a 1000 Hz pure tone at an anchor duration of 250 ms with onset and offset of 10 ms raised cosine ramps [3]. The minimum and maximum frequency deviation of the variable stimulus was 0.1 Hz and 100 Hz, respectively [4]. The participant was asked to identify the variable block.

*IDL***.** The minimum intensity difference requires the subject to discriminate two otherwise identical sounds. IDL was measured for a 1000 Hz pure tone at an anchor duration of 250 ms with onset and offset of 10 ms raised cosine ramps . The minimum and maximum intensity deviation used was 0.99 dB and 10 dB. The participant was asked to identify the variable block.

*DDT***.** The minimum difference in duration which a participant requires in order to discriminate was assessed. DDT was measured for a 1000 Hz tone at an anchor duration of 250 ms with onset and offset of 10 ms raised cosine ramps. The minimum and maximum value of the duration deviation used was 0.1 ms and 200.1 ms . The participant was asked to identify the variable block.

GDT. The minimum gap which a participant can identify in the middle of a 500 ms broadband noise was assessed The minimum and maximum duration of the gap used

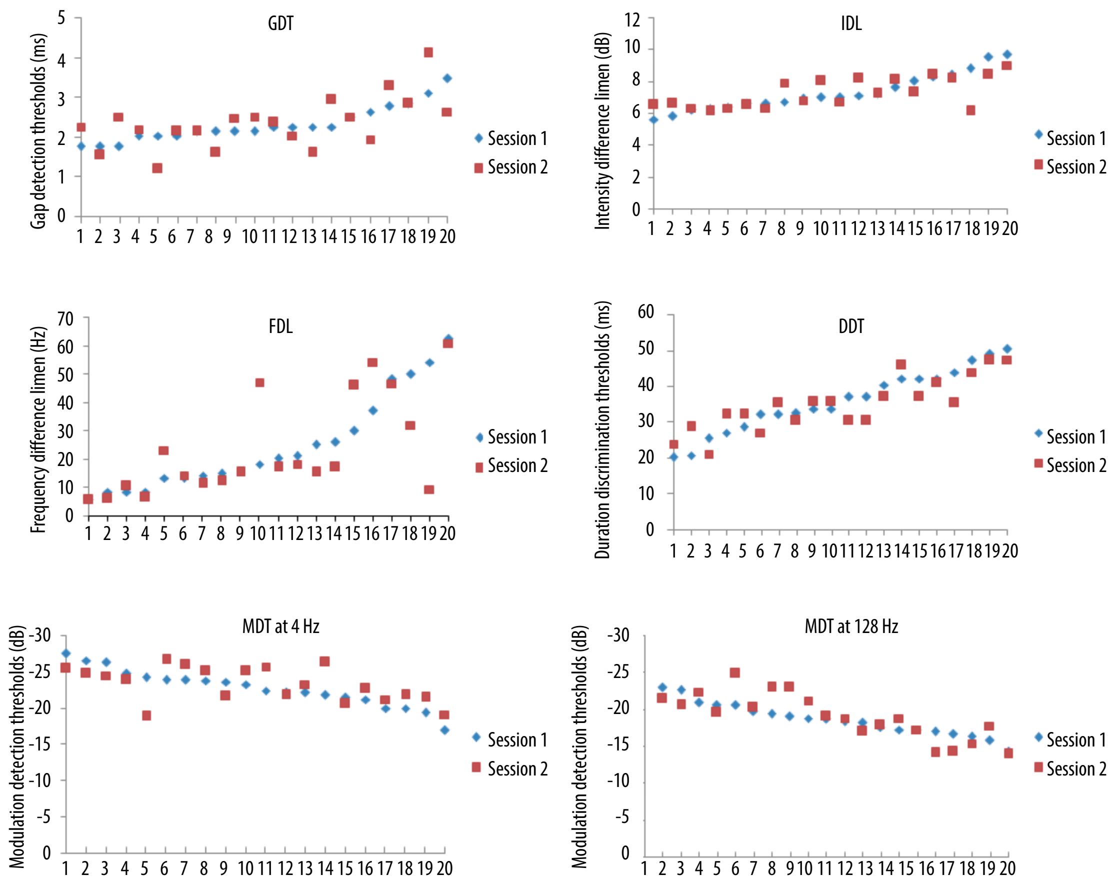

Figure 2. Scatter plot representing the individual scores for various psychoacoustic measures across two sessions; the participants are arranged in ascending order based on the scores of session 1

was 0.1 ms and 64 ms . The participant was asked to identify the variable block.

MDT. The minimum amplitude modulation necessary to identify amplitude-modulated noise from unmodulated white noise was assessed. A 1000 ms Gaussian noise was sinusoidally amplitude modulated at 4 Hz and 128 Hz. These modulation frequencies were selected as they cover both high and low modulation frequencies. The task of the participant was to identify the block which had the modulated noise. The minimum and maximum value of the amplitude modulation used was –30 dB and 0 dB .

All these measures were assessed for two sessions within one day to assess test–retest reliability.

Statistical analyses

The data of the present study was subject to statistical analyses using the Statistical Package for the Social Sciences (version 20). Descriptive statistics were done to assess the mean and standard deviation of all the parameters across the two sessions. The data was distributed normally across the trials (Shapiro–Wilks $p > 0.05$), and thus parametric statistics was performed on the data. A paired-sample t-test was done to estimate the significance of the difference between sessions on the psychoacoustical procedures. The test–retest results for each of these measures were analyzed using the intraclass correlation coefficient (ICC), which is a widely used reliability index in test–retest reliability analyses.

Results

The present study aimed to assess the test–retest reliability of the various psychoacoustic measures using MLP. Figure 1 shows the mean and standard deviation of the various psychoacoustic measures across two sessions, and Figure 2 shows the performance of individual participants on different psychoacoustic measures across two sessions. The participants in the figures are arranged in ascending order of the scores attained in session 1, and the performance for participants cannot be compared across each graph. From Figures 1 and 2 it can be inferred that there is minimal difference in scores across two sessions for most of the participants. However it can be noted that in FDL there are a few outliers with large differences between the scores of session 1 and 2 (participant 10 and 19).

Further, a paired t-test was done and the results showed no significant difference in scores for frequency difference limen ($t(19) = -0.192$, $p > 0.05$), intensity difference limen ($t(19) = -0.990$, $p > 0.05$), duration discrimination thresholds ($t(19) = 0.403$, $p > 0.05$), gap detection thresholds ($t(19) = 0.740$, $p > 0.05$), amplitude modulation detection at 4 Hz ($t(19) = -0.983$, $p > 0.05$) and for 128 Hz ($t(19) = 0.900$, $p > 0.05$) across two sessions. The test–retest reliability for each of these measures was analyzed using an interclass correlation test. The ICC values for each of these measures are depicted in Table 1. From the table, it can be inferred that the test–retest reliability of various psychoacoustic measures using MLP is between fair and good (ICC ranging from 0.618 to 0.83).

Table 1. ICC values across sessions for the various psychoacoustic measures

Psychoacoustic measure	ICC value	Reliability outcome
FDL	0.793	Good
IDL	0.790	Good
DDT	0.618	Fair
GDT	0.717	Fair
MDT at 4 Hz	0.709	Fair
MDT at 128 Hz	0.830	Good

Discussion

The present study aimed to assess the test–retest reliability of various psychoacoustic measures using the maximum likelihood procedure. Studies in the past have assessed psychoacoustic measures using MLP for various purposes; however, the test–retest reliability was not reported in these studies The results of the present study showed that there was no significant difference in scores of various psychoacoustical measures across sessions using MLP. The result of the present study is supported by the study done by Green However, in that study, the test–retest reliability was for pure tone thresholds at six audiometric frequencies using a yes/no procedure after 12 trials. Green reported that the mean threshold estimated through MLP showed a standard deviation of 3 dB during the five trials. In the present study, the test–retest reliability for various psychoacoustic measures was measured using the three-AFC method after 30 trials.

Further, studies in the past have shown that MLP is less time-consuming in threshold estimation compared to conventional procedures . In MLP, the threshold is estimated by varying the stimulus level over a broad range in the initial trials. Therefore, in MLP a short practice session can precede the actual testing, or the initial block of trials can be excluded from the statistical analysis. In the study by Green , it took around 5 min for the participants to complete the two sets of threshold estimates. In the present study, it took around 3 min to complete the assessment of one psychoacoustic test.

However, in the present study outliers were seen for the FDL procedure, wherein two participants (participant 10 and 19) performed differently across the two sessions. This could be because FDL is a difficult task compared to the other psychoacoustic procedure, as reported by the participants. It can also be noted from Table 1 that the ICC is best for MDT at 128 Hz and least for DDT.

Conclusion

It can be concluded from the present study that the test–retest reliability of various psychoacoustic measures using MLP is between fair and good. However, further study with more number of participants across various sessions would strengthen the present results.

References

1. Dreschler WA, Plomp R. Relation between psychophysical data and speech perception for hearing impaired subjects. I. J Acoust Soc Am, 1980; 68(6): 1608–15.
2. Glasberg BR, Moore BC. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scand Audiol Suppl. 1989; 32: 1–25.
3. Jain C, Sahoo JP. The effect of tinnitus on some psychoacoustical abilities in individuals with normal hearing sensitivity. Int Tinnitus J, 2014; 19(1): 28–35.
4. Jain C, Kumar AU. Relationship among psychophysical abilities, speech perception in noise and working memory in individuals with normal hearing sensitivity across different age groups (PhD in Audiology), 2016; University of Mysore, Mysore, India.
5. Leek MR. Adaptive procedures in psychophysical research. Perception & Psychophysics, 2001; 63(8):1279–92.
6. Soranzo A, Grassi M. Psychoacoustics: a comprehensive MATLAB toolbox for auditory testing. Front Psychol, 2014; 5:712.
7. Hall JL. Maximum likelihood sequential procedure for estimation of psychometric functions. J Acoust Soc Am, 1968; 44(1): 370.
8. Shen Y, Richards VM. A maximum-likelihood procedure for estimating psychometric functions: thresholds, slopes, and lapses of attention. J Acoust Soc Am, 2012; 132: 957–996.
9. Grassi M, Soranzo A. MLP: a MATLAB toolbox for rapid and reliable auditory threshold estimation. Behav Res Methods, 2009; 41(1): 20–8.
10. Green DM. A maximum likelihood method for estimating thresholds in a yes–no task. J Acoust Soc Am, 1993; 93(4): 2096–105.
11. Kumar AU, Ankmnal VS. Temporal processing abilities across different age groups. J Am Acad Audiol, 2011; 22(1): 5–12.
12. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Arch Orofac Sci, 2017; 12: 1–11.
13. Jain C, Mohamed H, Kumar AU. Short-term musical training and psychoacoustical abilities. Audiol Res, 2014; 4(1): 102.
14. Sao T, Jain C. Effects of hormonal changes in temporal perception, speech perception in noise and auditory working memory in females. Hearing Balance Commun, 2016; 14(2): 94–100.
15. Harris KC, Eckert MA, Ahlstrom JB, Dubno JR. Age-related differences in gap detection: effects of task difficulty and cognitive ability. Hear Res, 2010; 264(1–2): 21–9.
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med, 2016; 15(2): 155–63.
17. Green DM. A maximum-likelihood method for estimating thresholds in a yes-no task. J Acoust Soc Am, 1993; 93(4 Pt 1): 2096–2105.