AJA

Research Article

Reliability of the Minimum Masking Level as Outcome Variable in Tinnitus Clinical Research

Patricia C. Mancini,[a,b] Richard S. Tyler,[a,c] Hyung Jin Jun,[a] Tang-Chuan Wang,[a] Helena Ji,[a] Christina Stocking,[d] Carrie Secor,[d] Eveling Rojas-Roncancio,[a] and Shelley Witt[a]

Purpose: The minimum masking level (MML) is the minimum intensity of a stimulus required to just totally mask the tinnitus. Treatments aimed at reducing the tinnitus itself should attempt to measure the magnitude of the tinnitus. The objective of this study was to evaluate the reliability of the MML.
Method: Sample consisted of 59 tinnitus patients who reported stable tinnitus. We obtained MML measures on two visits, separated by about 2–3 weeks. We used two noise types: speech-shaped noise and high-frequency emphasis noise. We also investigated the relationship between the MML and tinnitus loudness estimates and the Tinnitus Handicap Questionnaire (THQ).
Results: There were differences across the different noise types. The within-session standard deviation averaged across subjects varied between 1.3 and 1.8 dB. Across the two sessions, the Pearson correlation coefficients, range was r = .84. There was a weak relationship between the dB SL MML and loudness, and between the MML and the THQ. A moderate correlation (r = .44) was found between the THQ and loudness estimates.
Conclusions: We conclude that the dB SL MML can be a reliable estimate of tinnitus magnitude, with expected standard deviations in trained subjects of about 1.5 dB. It appears that the dB SL MML and loudness estimates are not closely related.

The measurement of tinnitus is a required primary factor to establish the effectiveness of tinnitus treatment. Tinnitus measurement was largely divided into two categories. According to the model of Dauman and Tyler (1992), it is important to distinguish the reaction to tinnitus (which is measuring how tinnitus affects a person's life) from the magnitude or internal level of the tinnitus. The primary reactions to tinnitus have been distinguished into the categories of thoughts and emotions, hearing, sleep, and concentration (Tyler et al., 2014). This measures how tinnitus affects a person's life. The other is measuring tinnitus based on characteristics of the perceived sound. Thus, it has characteristics of loudness, pitch, and quality. Conceptually, the tinnitus must have some "internal level" or "magnitude." The magnitude of the tinnitus could be measured from an estimate of its loudness or the level of a sound required to mask it (Tyler, 2000; Vernon & Meikle, 2003).

We have argued that treatments aimed at reducing the tinnitus should measure the magnitude of the tinnitus (Tyler, 1985; Tyler et al., 2007). Tinnitus loudness is an estimate of the intensity of patient's tinnitus using a bracketing procedure. The intensity of a pulsed, 500-Hz tone is increased in 2-dB steps until the patient reports that the pulsed tone is at the same intensity of their tinnitus (Tyler et al., 2014). The minimum masking level (MML) has been used as a metric in clinical trials of tinnitus for many years (e.g., Harker et al., 1987; Jastreboff et al., 1994; Vernon, 1977). Davis et al. (2007) used MML for evaluating Neuromonics tinnitus treatment in a clinical trial. They showed 5–dB SL reduction of MML 6 months after treatment and 12.1–dB SL reduction of MML 12 months after (Davis

[a]Department of Otolaryngology—Head and Neck Surgery, The University of Iowa, Iowa City
[b]Department of Speech-Language Pathology and Audiology, Universidade
Federal de Minas Gerais, Belo Horizonte, Brazil
[c]Department of Communication Sciences and Disorders, The University of Iowa, Iowa City
[d]University at Buffalo Speech-Language and Hearing Clinic, NY
Correspondence to Richard S. Tyler: rich-tyler@uiowa.edu

Editor-in-Chief: Ryan W. McCreery
Editor: Jamie Bogle
Received June 17, 2019
Revision received April 14, 2020
Accepted April 22, 2020
https://doi.org/10.1044/2020_AJA-20-00047

Disclosure: The authors have declared that no competing interests existed at the time of publication.

et al., 2007). Teggi et al. (2009) used MML for outcome evaluation for their laser treatment of tinnitus. They compared MML outcomes between the laser treatment group and the placebo group. There was no significant difference in MML of two groups (Teggi et al., 2009). Tyler et al. (1984) showed good correspondence between the MML and the loudness estimates within two individuals who showed clear changes over time during treatment.

To determine if an individual's reduction in tinnitus magnitude following treatment is significant or might result from test–retest variability, it is important to know what the test–retest variability is, preferably for the individual (Tyler et al., 2007).

Mitchell et al. (1993) reported test–retest data for the MML reported a difference of 4.8 dB SL. The test–retest correlation for MML was found to be 0.75. Henry et al. (2013) reported test–retest reliability of binaural MMLs, with the level adjusted independently in each ear. They measured MMLs on two visits for 21 participants. Their subject MML showed test–retest reliability of $r = .808$ (Henry et al., 2013).

In Henry et al.'s (2013) study, there is no mention about the consistency of participant's tinnitus perception. Furthermore, they measured MMLs using a binaural stimuli with levels adjusted separated for each ear. This is not representative of how MMLs are measured. Furthermore, there are binaural tinnitus masking effects that could be level dependent and confuse the interpretation of this procedure (Tyler et al., 1984).

The procedure for determining the MML is different in different studies. There exists no single standard or norm. However, the basic procedure is always the same (Vernon & Meikle, 2003), with variants such as:

- The type of masking noise (narrow band or broadband, broadband relatively flat-spectrum noise, or speech-shaped noise). Speech-shaped noise is intended to reflect the long-term average spectrum of ongoing speech. It is primarily a low frequency emphasis broadband noise. It is used in speech audiometry as the most effective masker for speech and is therefore widely available on clinical audiometers worldwide.
- The step size of the level used in measurement (5 dB, although smaller step sizes such as 2 dB would enable greater accuracy).
- The starting level (increasing masking noise until the tinnitus is no longer audible or starting with complete masking and then decreasing the masking noise until the tinnitus becomes audible again). High levels might change the tinnitus, so starting at a low level is desirable.
- The pyschoacoustical procedure (bracketing or ascending method of limits).

The Tinnitus Handicap Questionnaire (THQ; Kuk et al., 1990) measures emotional, behavioral, and general health effects of tinnitus. The THQ consists of 27 items and has a high reliability of .94. A total score is derived from the average of all 27 items, and three subscale scores can also be calculated. The subscales relate to the physical, emotion, social consequences of tinnitus, and hearing-related changes. The psychometric validation study found that the THQ correlated moderately high with similar measures of tinnitus distress, including tinnitus loudness magnitude (0–100 rating scale), average hearing thresholds, and scales of life satisfaction, depression, and health status (Kuk et al., 1990). The test–retest reliability of the THQ was independently tested and found to be high (Newman et al., 1995).

The aim of this study is to evaluate the reliability of the MML as a measure of the magnitude of tinnitus for its potential use as outcome variable in tinnitus clinical trials and to evaluate differences across three different noise types. The study assesses the aspect of intrasession and intersession test–retest reliability. We also investigated the relationship between the dB SL MML and loudness estimates.

Method

Study Subjects

Two study sites were involved, 40 subjects were enrolled at that University of Iowa and 19 subjects were enrolled at the University of Buffalo. Participants were recruited through advertisements on the study sites' websites, as well as from regular patients at the study sites' otolaryngology departments. To qualify, we asked participants the following questions:

1. When you stop and listen to your tinnitus, please rate the stability of your tinnitus loudness over the last 2 weeks (0 = changes all the time, 100 = tinnitus never changes).
2. When you stop and listen to your tinnitus, please rate the stability of your tinnitus pitch over the last 2 weeks (0 = changes all the time, 100 = tinnitus never changes).
3. When you were thinking of your tinnitus in the past 2 weeks, could you
   - always hear it?
   - only sometimes?
4. Did your tinnitus in the past 2 weeks occur
   - all of the time?
   - most of the time?
   - a good bit of the time?
   - some of the time?
   - a little of the time?
   - none of the time?

Only tinnitus patients who rated their tinnitus stability at 75 or more on both Questions 1 and 2 and who responded to Question 3 with "always hear it" and Question 4 with “all of the time,” “most of the time,” or “a good bit of the time” were considered “stable” tinnitus that we had defined and further assessed for inclusion in the study. Further inclusion and exclusion criteria included:

Inclusion criteria:

- Stable sensorineural tinnitus (unilateral or bilateral) —being “stable” defined by questions above about stability of tinnitus and “sensorineural tinnitus” being the tinnitus associated with a sensorineural hearing loss
- Tinnitus onset not more than 10 years ago
- Age ≥ 18 years and ≤ 70 years

Although tinnitus is a significant problem in older people (Aazh et al., 2017), in this study, we focused on establishing the reliability without the influence of cognitive-aging issues.

Exclusion criteria:

- Tinnitus that is not completely maskable at 70 dB HL
- Fluctuating tinnitus (varying in pitch and/or loudness from day to day)
- Intermittent tinnitus (present on certain days, but not on others)
- Tinnitus originating in the middle ear
- Present middle ear disease
- Ménière’s disease
- Air–bone gap of ≥ 15 dB in three most affected frequencies
- Any ongoing therapy known as potentially tinnitus-inducing (e.g. aminoglycosides, cisplatin, loop diuretics, high doses of aspirin, quinine)
- Any therapy for tinnitus that is (a) currently ongoing and was initiated only in the past 2 weeks, (b) expected or planned to be initiated only during the course of the study, or (c) expected or planned to be stopped, interrupted, resumed, or, otherwise, modified during the course of the study
- Presence of clinically significant depression (Beck Depression Inventory score ≥ 20) or other psychiatric disorder
- History of drug abuse or alcoholism
- Concurrent participation in another clinical trial
- Involvement in legal action
- Tinnitus patients sometimes suffer from severe depression (Aazh et al., 2019; Aazh & Moore, 2017). We did not want this to interfere with our metrics of typical reliability in these patients.

Patients who did not meet all inclusion criteria and/or who met one or more of the exclusion criteria were excluded. Fifty nine patients that met the inclusion/exclusion criteria and agreed to participate were included in this study. All study procedures were conducted after approval from the institutional review boards of the University of Iowa and University of Buffalo.

Study Design

Each subject attended three sessions, each about 2 weeks apart. The first visit was treated as a practice session. In clinical trials, we want the metrics to be as accurate as possible. Therefore, providing a practice session is desirable. MML measures were obtained only at the second and third participants’ visits, and we consider them as Visits 1 and 2, respectively. Patients were advised not to listen to loud music and to avoid noisy situations during the study. All testing was done in a double-walled sound-attenuated audio booth.

MML Method

In case of bilateral tinnitus, only the ear with the louder tinnitus (as determined by loudness magnitude estimate) at baseline was assessed, or if equally loud, the right ear was assessed. The noises used were audiometric standards for speech-shaped and for narrow-band noise (NBN; American National Standards Institute, 1996). A high-frequency emphasis noise was created based on the inverse of the average audiogram in patients with bilateral tinnitus (Pan et al., 2009). The threshold for these noises was measured first using an ascending method of limits and a 2-dB step size. To measure the noise-masked threshold, the noise level was increased in 2-dB steps. Each of these measures was obtained 3 times. If the three measures were ≤ 5-dB apart, then the average of the three values was taken as the threshold. If not, then a fourth measure was obtained. The average of the three closest values was taken as the threshold. It is important to appreciate that we measured the threshold to the noise and the noise level required to mask the tinnitus (which we refer to here as the “dB MML”) and then subtracted the two to get the MML dB SL.

Results

The characteristics of the subjects are shown in Table 1. They are representative of the tinnitus patients at large. For the MML, we analyzed the data separately for the Iowa and Buffalo data and found no differences. At Iowa, the Pearson correlation coefficients between Visit 1 and Visit 2 were $r = .84$ for the NBN, $r = .83$ for the speech-shaped noise, and $r = .84$ for the high-frequency noise ($p < .001$ for all). At Buffalo, the Pearson correlation coefficients were $r = .81$ for the NBN, $r = .84$ for the speech-shaped noise, and $r = .87$ for the high-frequency noise ($p < .001$ for all). Hereafter, the results are simply combined.

Figure 1 shows the MML in dB SL for Visits 1 and 2, for each of the three noises. The Pearson correlation coefficients were $r = .84$ for the NBN, $r = .84$ for the speech-shaped noise, and $r = .85$ for the high-frequency noise ($p < .001$ for all). Note the wide range of MMLs from a few dB above

Table 1. Demographics of subjects (age and duration of tinnitus are based on the practice visit dates).

Gender	No. of patients	Descriptive data
		Age (years)
Female	19	*M* = 53.8
Male	40	range: 21–70
		SD = 11.1
Site of tinnitus		
Left ear	5	
Right ear	8	
Both ears, equally	22	
Both ears, but worse in left ear	13	
Both ears, but worse in right ear	8	
More in the right side of head	1	
More in the left side of head	1	
Middle of head	1	
Loudness rating (0–100)		*M* = 60.7 range: 4–100 *SD* = 22.4
Fluctuation of tinnitus (0–100)		(*M*, range, *SD*)
–Loudness		(88.3, 75–100, 9.4)
–Pitch		(92.4, 75–100, 8.1)
Qualities of tinnitus		
Ringing or whistling	37	
Cricket-like	10	
Roaring, “Shhh”, or rushing	2	
Buzzing	4	
Hissing	4	
Other	2	
Duration of tinnitus (years)		*M* = 4.5 range: 0.5–10 *SD* = 2.9

threshold and several with over 60 dB above threshold. With about 2 weeks between sessions, most showed values within 3–5 dB, but a few had values that were over 9 dB different.

Figure 2 compares the standard deviations (calculated from the three replications) obtained from each individual for Visit 1 and Visit 2. This is shown for each of the noise types. The *SD* was typically less than 2 dB for most subjects. There were a few subjects, however, whose *SD* exceeded 3 dB. There were differences across the different noise types, but none showed smaller deviations. For the speech-shaped noise, the average within-session *SD* was 1.8 dB for the first test and 1.7 dB for the second test. For the high-frequency noise, the average within-session *SD* was 1.6 dB for the first test and 1.7 dB for the second test. For the NBN noise, the average within-session *SD* was 1.7 dB for the first test and 1.3 dB for the second test.

Figure 3 shows the relationship between MML and loudness. The correlation is weak with all of the noises (high-frequency noise $r = –.07$, NBN noise $r = –.09$, speech-shaped noise $r = .02$; $p > .05$ for all). Similarly, Figure 4 shows a weak correlation between the MML and the THQ (high-frequency noise $r = .18$ [$p > .05$], NBN noise $r = .13$ [$p > .05$], speech-shaped noise $r = .33$ [$p = .02$]). Figure 4 shows the relationship between the THQ and the loudness ratings for Visit 2 ($r = .44$, $p < .01$).

Discussion

It is critical to have a measure of the magnitude of the tinnitus in order to evaluate studies that propose to reduce the magnitude of the tinnitus. The MML is an important candidate for this. One advantage is that subjects are only required to respond whether they hear their tinnitus or not. In this study, MML generally showed good test–retest reliability. As expected, differences were observed across subjects.

The test–retest between Sessions 1 and 2 for those subjects was generally very reliable. The dB SL MML is a reliable metric of tinnitus magnitude with an *SD* of about 1.5 dB.

One source of variability is simply the test–retest associated with listening carefully for the tinnitus and judging whether it is present or not. A second source of variability is that a patient’s tinnitus actually might change

Figure 1. Minimum masking level (MML) dB SL for Visits 1 and 2 for each of the noise types.

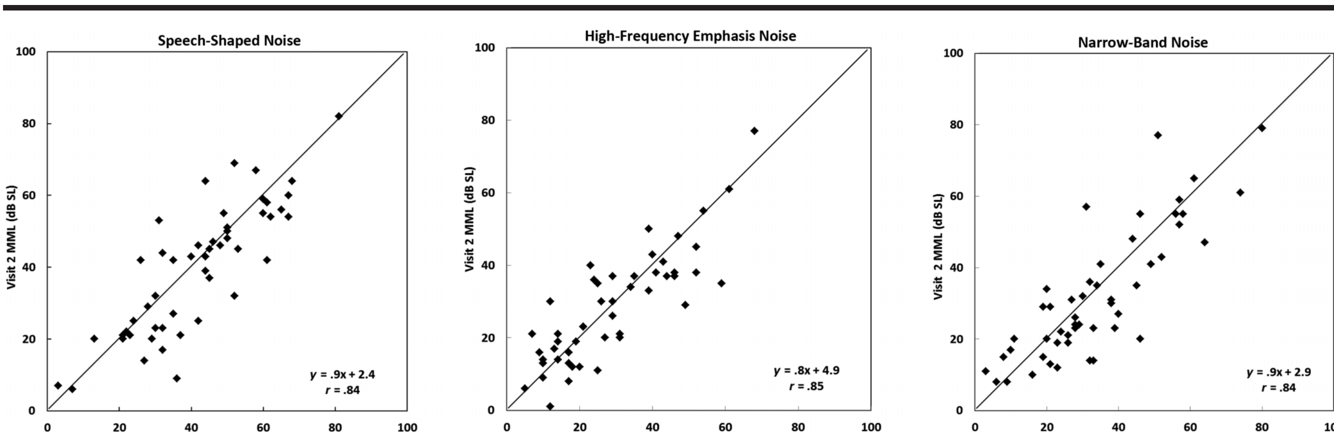
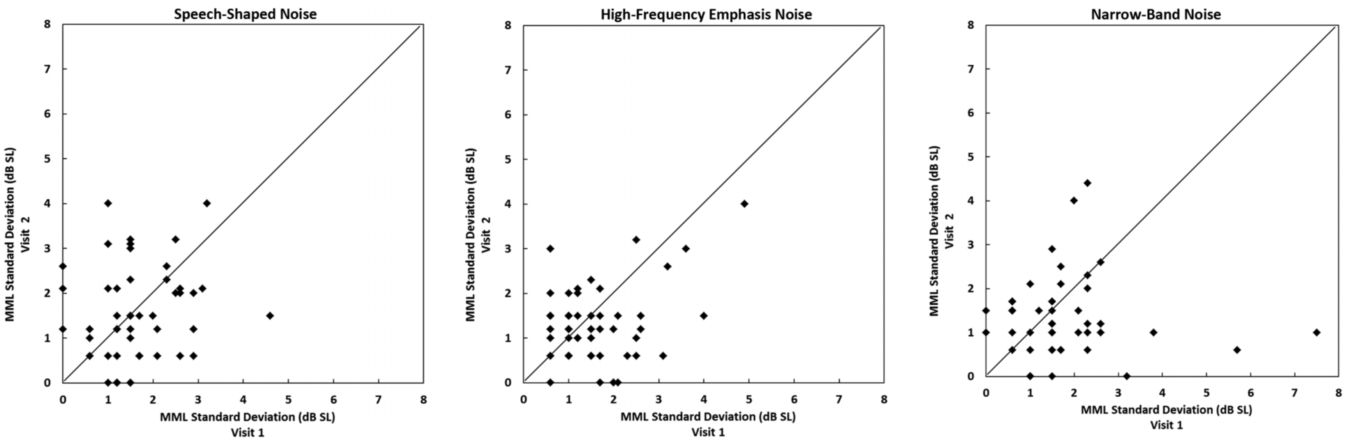


Figure 2. Standard deviations (in dB sensation level) for Visit 1 and Visit 2 for each of the noise types. MML = minimum masking level.



from day to day (Stouffer & Tyler, 1990). There are many tinnitus sufferers that their tinnitus changes in pitch and loudness. We attempted to control for this effect by requiring a stable tinnitus in our study population. However, it might be that our criteria were too lenient. If we examined only subjects who reported a stability estimate of 90% (instead of 75%), 16 participants replied their stability of tinnitus loudness above 90%. Intraclass coefficient of correlation for two sessions of these participants was .859. Intraclass coefficient of correlation between Sessions 1 and 2 is $r = .903$. A third source of variability is that the masking noise itself might change the tinnitus. We attempted to avoid this by using an ascending method.

According to the psychological model (Dauman & Tyler, 1992), the loudness of the tinnitus is a factor contributing to the overall annoyance of the tinnitus. Generally, within an individual, the louder the tinnitus, the more annoying it will be. However, when examining correlations across subjects, different people have different coping strategies. So, tinnitus loudness is only one factor.

The low correlation between MML and the THQ is generally consistent with other studies (Nyenhuis et al., 2013). Figueiredo et al. (2010) reported no significant relation between tinnitus handicap inventory and psychoacoustic measure, such as tinnitus pitch matching and MML. We emphasize again the importance of examining individual performance. Furthermore, treatments that attempt to decrease tinnitus magnitude should be evaluated by measuring tinnitus itself (such as the MML), and treatments that attempt to change the reaction of tinnitus should be evaluated by measuring tinnitus reaction, such as THQ or the newly developed Tinnitus Primary Functions Questionnaire (Tyler et al., 2014).

In conclusion, the dB SL MML can be a reliable estimate of tinnitus magnitude. The test–retest reliability might be established for each subject prior to participation in a study. It is important to stress that the magnitude of the tinnitus as measured by the MML will not correlate highly with metrics of annoyance when comparing across subjects. This does not invalidate either metric, it only

Figure 3. The relationship between minimum masking level (MML) and loudness for each of the noise types.

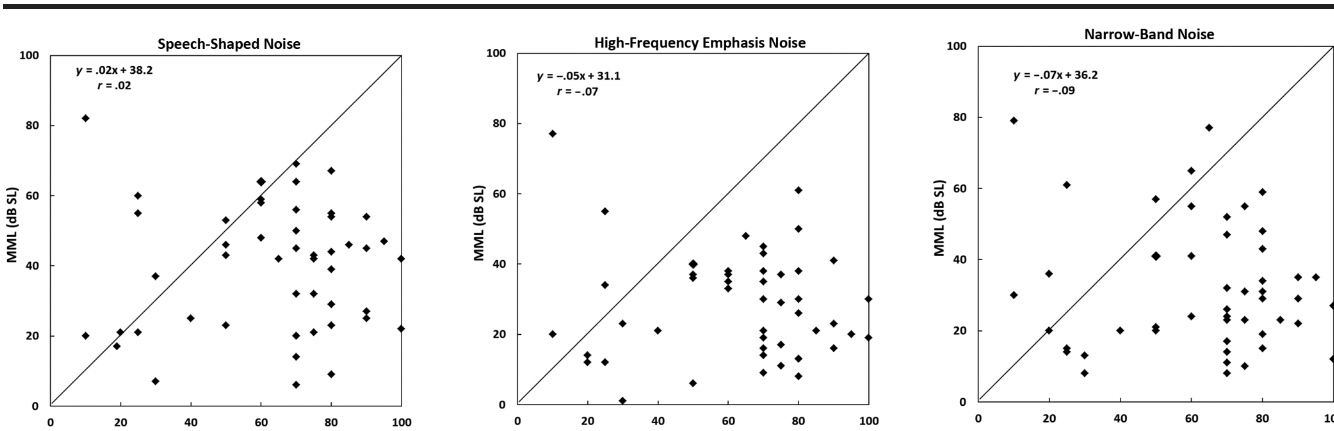
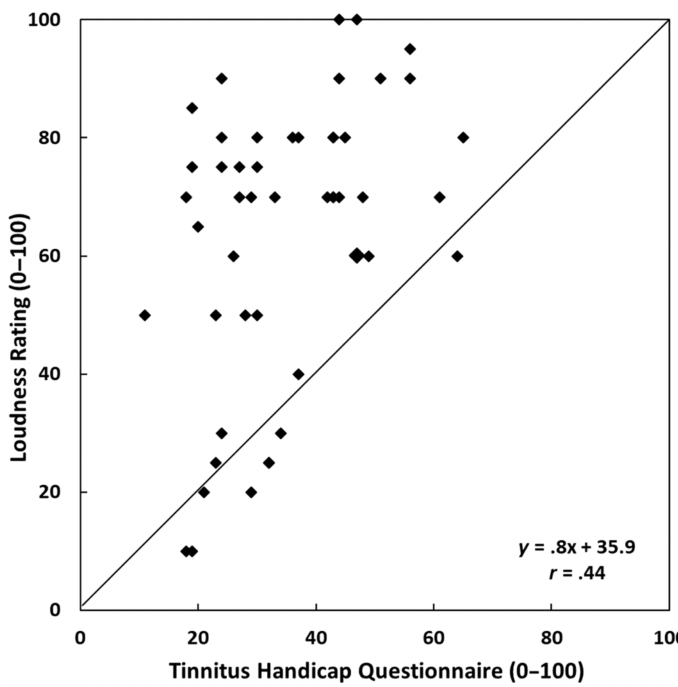


Figure 4. The relationship between the Tinnitus Handicap Questionnaire and the loudness ratings for Visit 2. MML = minimum masking level.



indicates that overall annoyance is impacting by the psychological makeup of the individual. Actually, changes within an individual can show consistency between magnitude and reactions metrics (Harker et al., 1987).

We note that, within-session test–retest reliability was very good, but that intersession test–retest reliability can be high for some subjects. One approach to this is that test–retest across and within sessions be monitored for each individual subject. For those that show greater variability, additional training or test sessions can be included. Additionally, a practice session could also be included if there is some improvement in test–retest reliability across sessions.

Acknowledgments

This study was partially funded by Auris Medical. The Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq), Brazil, provided a scholarship for P. C. Mancini.

References

Aazh, H., Danesh, A. A., & Moore, B. C. J. (2019). Parental mental health in childhood as a risk factor for anxiety and depression among people seeking help for tinnitus and hyperacusis. *Journal of the American Academy of Audiology*, *30*(9), 772–780. https://doi.org/10.3766/jaaa.18001

Aazh, H., Lammaing, K., & Moore, B. C. J. (2017). Factors related to tinnitus and hyperacusis handicap in older people. *International Journal of Audiology*, *56*(9), 677–684. https://doi.org/10.1080/14992027.2017.1335887

Aazh, H., & Moore, B. C. J. (2017). Factors associated with depression in patients with tinnitus and hyperacusis. *American Journal of Audiology*, *26*(4), 562–569. https://doi.org/10.1044/2017_AJA-17-0008

American National Standards Institute. (1996). *Specifications for audiometers* (ANSI S3.6-1996).

Dauman, R., & Tyler, R. S. (1992). Some considerations on the classification of tinnitus. In J.-M. Aran & R. Dauman (Eds.), *Tinnitus 91—Proceedings of the fourth international tinnitus seminar* (pp. 225–229). Kugler Publications.

Davis, P. B., Paki, B., & Hanley, P. J. (2007). Neuromonics tinnitus treatment: Third clinical trial. *Ear & Hearing*, *28*(2), 242–259. https://doi.org/10.1097/AUD.0b013e3180312619

Figueiredo, R. R., Rates, M. A., Azevedo, A. A., Oliveira, P. M., & Navarro, P. B. (2010). Correlation analysis of hearing thresholds, validated questionnaires and psychoacoustic measurements in tinnitus patients. *Brazilian Journal of Otorhinolaryngology*, *76*(4), 522–526. https://doi.org/10.1590/S1808-86942010000400018

Harker, L. A., Tyler, R. S., & Fredell, P. A. (1987). Evaluation of flecainide acetate (Tambocor) R as a treatment for tinnitus. In H. Feldmann (Ed.), *Proceedings of the III international tinnitus seminar* (pp. 322–325). Harsch Verlag.

Henry, J. A., Roberts, L. E., Ellingson, R. M., & Thielman, E. J. (2013). Computer-automated tinnitus assessment: noise-band matching, maskability, and residual inhibition. *Journal of the American Academy of Audiology*, *24*(6), 486–504. https://doi.org/10f.1016/0378-5955(94)90113-9

Jastreboff, P. J., Hazell, J. W., & Graham, R. L. (1994). Neurophysiological model of tinnitus: Dependence of the minimal masking level on treatment outcome. *Hearing Research*, *80*(2), 216–232. https://doi.org/10.1097/00003446-199012000-00005

Kuk, F. K., Tyler, R. S., Russell, D., & Jordan, H. (1990). The psychometric properties of the Tinnitus Handicap Questionnaire. *Ear and Hearing*, *11*(6), 434–445. https://doi.org/10.1097/00003446-199012000-00005

Mitchell, C. R., Vernon, J. A., & Creedon, T. A. (1993). Measuring tinnitus parameters: Loudness, pitch, and maskability. *Journal of the American Academy of Audiology*, *4*(3), 139–151. https://doi.org/10.1177/000348949510400910

Newman, C. W., Wharton, J. A., & Jacobson, G. P. (1995). Retest stability of the Tinnitus Handicap Questionnaire. *Annals of Otology, Rhinology & Laryngology*, *104*(9), 718–723. https://doi.org/10.1177/000348949510400910

Nyenhuis, N., Golm, D., & Kroner-Herwig, B. (2013). A systematic review and meta-analysis on the efficacy of self-help interventions in tinnitus. *Cognitive Behavior Therapy*, *42*(2), 159–169. https://doi.org/10.1080/16506073.2013.803496

Pan, T., Tyler, R. S., Ji, H., Coelho, C., Gehringer, A. K., & Gogel, S. A. (2009). The relationship between tinnitus pitch and the audiogram. *International Journal of Audiology*, *48*(4), 277–294. https://doi.org/10.1080/14992020802581974

Stouffer, J. L., & Tyler, R. S. (1990). Characterization of tinnitus by tinnitus patients. *Journal Speech and Hearing Disorders*, *55*(3), 439–453. https://doi.org/10.1044/jshd.5503.439

Teggi, R., Bellini, C., Piccioni, L. O., Palonta, F., & Bussi, M. (2009). Transmeatal low-level laser therapy for chronic tinnitus with cochlear dysfunction. *Audiology and Neurotology*, *14*(2), 115–120. https://doi.org/10.1159/000161235

Tyler, R. S. (1985). Pyschoacoustical measurement of tinnitus for treatment evaluations. In E. Myers (Ed.), *New dimensions in otorhinolaryngology head and neck surgery* (pp. 455–458). Elsevier.

Tyler, R. S. (2000). The psychoacoustical measurement of tinnitus. In R. S. Tyler (Ed.), *Tinnitus handbook* (pp. 149–179). Singular.

Tyler, R. S., Babin, R. W., & Niebuhr, D. P. (1984). Some observations on the masking and post-masking effects of tinnitus. *The Journal of Laryngology & Otology* (Suppl. 9), 150–156. https://doi.org/10.1017/S1755146300090363

Tyler, R., Ji, H., Perreau, H., Witt, S., Noble, W., & Coelho, C. (2014). Development and validation of the Tinnitus Primary Function Questionnaire. *American Journal of Audiology*, *23,* 260–272. https://doi.org/10.1044/1059-0889.0104.36

Tyler, R. S., Oleson, J., Noble, W., Coelho, C., & Ji, H. (2007). Clinical trials for tinnitus: Study populations, designs, measurement variables, and data analysis. *Progress in Brain Research*, *166,* 499–509. https://doi.org/10.1016/S0079-6123(07)66048-8

Vernon, J. (1977). Attempts to relieve tinnitus. *Journal of the American Audiology Society*, *2*(4), 124–131.

Vernon, J. A., & Meikle, M. B. (2003). Tinnitus: Clinical measurement. *Otolaryngologic Clinics of North America*, *36*(2), 293–305, vi. https://doi.org/10.1016/S0030-6665(02)00162-7