ASR
pISSN 2635-5019 / eISSN 2635-5027 / Audiol Speech Res 2024;20(3):152-163 / https://doi.org/10.21848/asr.240154
RESEARCH PAPER

Effectiveness of Sound Therapy Durations in Tinnitus Rehabilitation: A Systematic Review and Meta-Analysis

Woojin Jayce Baek, Junghwa Bahng

Department of Audiology and Speech Language Pathology, Hallym University of Graduate Studies, Seoul, Korea

Received: July 2, 2024
Revised: July 8, 2024
Accepted: July 8, 2024

Correspondence:
Junghwa Bahng, PhD
Department of Audiology and Speech Language Pathology, Hallym University of Graduate Studies, 405 Yeoksam-ro, Gangnam-gu, Seoul 06198, Korea
Tel: +82-70-8680-6933
Fax: +82-2-3453-6618
E-mail: bahng.jh@gmail.com

Purpose: This systematic review aimed to conduct a statistical analysis comparing subjective evaluations from patients to determine the efficacy of sound therapy durations for tinnitus. **Methods:** After defining search terms and establishing inclusion and exclusion criteria, 15 research articles from PubMed, ScienceDirect, KCI, and the Cochrane Library met all qualifications through a vetting process involving two reviewers. Extracted data were organized by collection time and analyzed using the R program (R foundation, Vienna, Austra) to calculate the standard mean difference with 95% confidence intervals. **Results:** The analysis of sound therapy efficacy showed statistically significant improvement in tinnitus symptoms between pre-treatment and 3 months, and between 3 months and 6 months. However, improvement stagnated and stabilized beyond 6 months of treatment. **Conclusion:** Sound therapy was most effective in providing significant relief and improvement in tinnitus symptoms during the first 6 months of treatment. Longer durations did not further improve symptoms but helped maintain a lower annoyance level.

Key Words: Tinnitus, Sound therapy, Masking therapy, Tinnitus retraining therapy.

INTRODUCTION

Tinnitus is phantom noise that patients perceive without an external source. Its identifying feature is simple but the potential impact is unproportional and global. In 2022, a systematic review and meta-analysis was conducted in regards to global prevalence of tinnitus. It is predicted that the pooled prevalence of tinnitus is approximately 9.8% with severe tinnitus being 2.3% which translates to about 740 million people have tinnitus and 120 million people consider tinnitus as a major problem in their lives (Jarach et al., 2022). The phantom noise is a global burden and greater efforts should be made to understand tinnitus. It is not only about the perceived noise but what that perception does to a person. For many patients, tinnitus is not a major concern as it serves as nothing more than a minor annoyance (Belli et al., 2008; Langguth et al., 2013). However, the impact can extend to make patients with severe tinnitus suffer from distress that leads to a myriad of psychological problems such as insomnia, anxiety, and inattentiveness as well as depression, hearing difficulties and hyperacusis (Belli et al., 2008; Jin et al., 2022; Langguth et al., 2013). These impacts raise urgency and importance to finding methods of treatment for tinnitus (Friberg et al., 2012).

Tinnitus remains elusive in the pursuit of comprehension and labeled as a scientific and clinical enigma (Cederroth et al., 2019). It is not a disease, but a symptom that accompanies other causes though there may be no obvious indication of what it is accompanying. It can be developed due to various reasons such as ear infections, acoustic interference (foreign objects or earwax), and exposure to loud noises or side effects of drugs (Belli et al., 2008). To further perplex researchers, tinnitus is also associated with psychological elements and stress (Hinton et al., 2006). These onsets involve changes anywhere within the auditory pathway, making tinnitus difficult to pinpoint its origin as the source seems to be multifaceted. It is not just the cause that varies, but also

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

 Copyright © 2024 Korean Academy of Audiology

compositions that may or may not make tinnitus permanent as observed with noise exposed patients. The pathology of tinnitus is not well understood due to the difference between temporary and permanent tinnitus, which remains unknown (Langguth et al., 2013). Tinnitus is formerly linked with hearing loss but it is a convoluted relationship. There are those with tinnitus that seem to have normal hearing thresholds by conventional means but perceive tinnitus due to damaged outer hair cells or presence of cochlear dead regions (Job et al., 2007; Weisz et al., 2006). Recently, it has been proposed that such perception is due to wrongful responses from the autonomic nervous systems (Jastreboff & Jastreboff, 2006). This discovery has led to construction of a neurological model which illustrates that tinnitus is the result of an incorrect trigger of the nervous system (Jastreboff & Jastreboff, 2006). This particular reaction creates negatively learned responses that are linked to the physiological responses (Formby et al., 2022). Despite these researches, tinnitus is still not understood because there are multiple forms with different pathophysiology involved. Because the pathophysiology is not understood fully, treatment options are quite limited in approach.

The treatment goal is to relieve symptoms and address patients' reactions to and perception of tinnitus (Belli et al., 2008; Jastreboff & Jastreboff, 2006). There are many methods proposed to treat and manage tinnitus, most of which focus on patients' behavior towards it rather than the cause (Jastreboff, 1990). Sound therapy (ST) refers to a treatment method in which small noise generators are assigned to patients and apply acoustic masking. Wide array of sounds (nature sounds, or white noise) that are more pleasant than the perceived tinnitus sounds are employed. Used stimuli cover a wide-frequency-bands to partially or fully mask the tinnitus, thereby decreasing patients' awareness (Hobson et al., 2012; Langguth et al., 2013). The end goal of ST is to habituate tinnitus. It is a combination of sound therapy and counseling, which does not aim to eliminate nor influence neural structure creating tinnitus sounds (Jastreboff & Jastreboff, 2006).

Although these methods are frequently employed clinically, there are not enough systematic reviews and meta-analysis of randomized controlled trials (RCTs) for tinnitus therapy (Henry et al., 2016). The aim of this review is to gather articles and perform a meta-analysis that attempts to analyze efficacy of ST by comparing repeated measures between durations. This attempt to analyze rates of improvements will establish an understanding of "minimum" amount of commitment required and add an additional repertoire for therapists to convince patients to carry on with therapy rather than quit prematurely.

MATERIALS AND METHODS

Prior to commencing data collection in earnest, we conducted a thorough search on PROSPERO to ascertain the existence of any systematic reviews and meta-analytic studies pertaining to redundant research ideas (CR42023390955).

Eligibility criteria

For an article to be eligible, publication year must be between 2000 and 2023 as well as having therapy that involved using sound stimulation or masking. Participants must have had tinnitus for 3 months or more. Types of stimulation employed in therapy were not limited as long as the therapy itself did not deviate from conventional masking methods. ST paired with counseling/education were considered acceptable. There were no restrictions to age or acoustic hearing ability.

As for the exclusion category, research that carried out sound therapy simultaneously with other methods such as pharmaceutical products, acupuncture or transcranial direct current stimulation were excluded. Studies that researched cognitive behavioral therapy that combine ST are also excluded. Any literature reviews, retrospective studies, meta-analysis or systematic review were excluded. No animal model studies were included.

Information sources and search strategy

An author conducted searches in four databases of the following: ScienceDirect, Cochrane Library, PubMed, and Korean Citation Index (KCI). Data collection was conducted between February and March of 2023. Search terms were decided on by determining two necessary terms; tinnitus and RCT as well as: ("tinnitus" AND ("masking therapy" OR "tinnitus retraining therapy" OR "music therapy" OR "sound therapy") AND ("randomized controlled trial" OR

"RCT" OR "randomized")). A different set of terms was used for KCI to broaden results by inputting only "tinnitus" AND "sound therapy" into KCI's search engine. Search terms were formulated by both reviewers beforehand.

Selection process

Conducted searches were collected per database into google spreadsheet tabs so that all articles could be traced back to the respective database. Collected information included title, author, year of publication, abstract and keywords if available. One reviewer (JWB) initiated exclusion of articles based on the year of publication and details based on titles and abstracts. Any systematic reviews or meta-analysis were filtered out. After the removal, remaining articles were put into one compiled list for the first screening which consisted of double checking the initial removal as well as to eliminate any duplicates.

Second screening was conducted to check for RCT. The second reviewer (JB) joined the screening process at this juncture. The remaining 105 articles were screened separately by the two reviewers to ensure that studies were RCT and had one of the approved treatment methods (ST, tinnitus retraining therapy or tinnitus masking) as well as availability. Both reviewers marked each article with Xs or Os. All rejections were debated before finalizing. Rejection elements during this screening included no access, not RCT, or that therapy was combined with another treatment. Thirty-five articles passed the second screening.

Third screening involved the process of double checking the second screening as well as a full body text check. Articles that did not use sound therapy as the main form of treatment were disqualified. All remaining articles were checked for participants, intervention, comparator, outcomes, and study type as well as available data. Articles that did not include extractable data, or feasibility as well as preliminary trials were excluded. Fifteen articles passed and were included in this analysis.

Data collection process

The collection process included checking method sections of each article for participants, level of hearing loss, used stimuli, outcome measures and mean values or presented graphics with survey outcome results. Many studies collected results in multiple formats and measured outcomes through various methods ranging from interviews, subjective evaluations to objective evaluation methods. All mean values (or mean differences), standard deviations and sample size from control groups and experimental groups were taken as data of interest. Control groups were defined as either standard of care or wait list control while experimental groups were any form of therapy that involved usage of masking sounds. These included sound therapy, tinnitus retraining therapy, placebo (in which sound was used), and tinnitus masking. The collected data values were outcomes measured through tinnitus handicap inventory (THI), tinnitus frequency inventory (TFI), tinnitus questionnaires (TQ), tinnitus handicap questionnaires (THQ), hospitals anxiety and depression scales (HADS), visual analogue scale (VAS), and World Health Organization quality of life brief version. However, to extract effect size without causing unit of analysis errors, only primary outcomes were extracted for computation. If a study had multiple primary measures for different sound therapy stimuli, then those were aggregated (Harrer et al., 2021) into one effect size.

If there were no mean values provided in a study and instead charts were provided as reported results, then a third-party program called Web Plot Digitizer (Automeris,

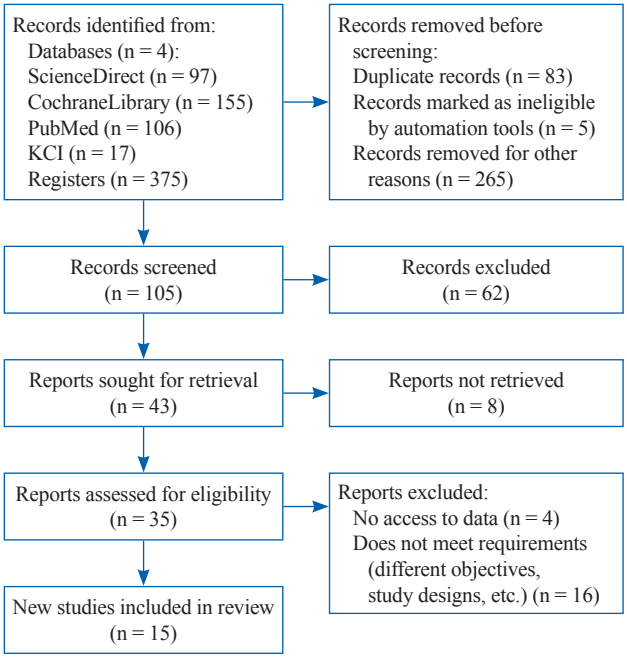

Figure 1. A flow diagram depicting the process of article selection.

Audistin, TX, US) was used to approximate pixel close value (Rohatgi, 2022). Although it would not provide an exact value as an outright reporting, the web program provided closer approximation than visual estimation (Figure 1).

Data items

Duration groups were determined based on frequently analyzed outcome measure intervals found on the gathered studies. Baseline group contained data of pre-treatment results. From pre-treatment measurements and measurements taken at 3 months were considered short groups. This group consisted participants who had received treatments for anywhere from 1 day to 3 months. Medium duration group consisted of measures taken anywhere between 3 months and a day to 6 months. Short and medium durations' outcome measures consisted of a wide range of time due to different intervals at which questionnaires were completed among studies. The long duration group contained measures taken at the 12 months mark with duration of therapy ranging from 6 months and a day to 12 months. Lastly, the extended groups consisted of outcome measures completed at 18 months.

The main goal of this analysis is to compare perceived improvements achieved by participants after certain durations and compare them to see if there were any notable clinically significant differences. As such, any mean value of outcome measures, standard deviations and sample sizes were taken as data items. If mean differences were presented instead of mean value of evaluations, then mean difference values were added back to the baseline results to unify data with mean values.

Publication bias

Publication biases were checked with the small-study effect methods and presented through funnel plots. In R package called meta (Balduzzi et al. 2019) has the funnel.meta function that generates a funnel plot by using the calculated effect size. Funnel plots were visually inspected and analyzed in order to assess publication bias.

Study risk of bias

All articles were evaluated independently in accordance with the physiotherapy evidence database (PEDro) scales by both reviewers in order to assess risk of bias. The PEDro scale is a tool composed of 11 items graded with 1 or 0 to evaluate risk of bias and completeness of research papers reporting statistics which is composed of inclusion criteria and source, random allocation, concealed allocation, similarity at baseline, subject blinding, therapist blinding, assessor blinding, completeness of follow up, intention-to-treat analysis, between-group statistical comparisons, and point measures and variability (Moseley et al., 2019). By answering 1 or 0 (yes or no), an article would be given a score of up to 11 possible points with the score of 8 or higher indicating higher reliability.

Heterogeneity

Heterogeneity was assessed through Higgins and Thompson's I^2 statistic, which was directly based on Cochran's Q test. With I^2, a general "rule of thumb" was followed where 25% or less to be considered low heterogeneity, 50% being moderate while 75% indicated substantial heterogeneity (Higgins et al., 2003). These percent ratings indicated approximately how much of the variance was caused by between-study heterogeneity.

Statistical analysis

All statistical analyses were performed by the R project version 4.3.0 (R foundation, Vienna, Austria) (Already Tomorrow) using an integrated program, R Studio (R foundation) "Cherry Blossom" along with multiple packages which were defined as part of the R source codes (R foundation) that allow it to function with dataset provisions, default statistics and graphics functions. Default packages used were: graphics, grDevices, methods, stats, and utils. Contributed packages were codes written by various authors to create specialized statistical methods that provided increased access to data or hardware and other functions to complement R. They were meant to enhance basic functions and help facilitate analysis further than possible with only basic functions. The basic packages turned on by default are of the following: base, graphics, grDevices, Matrix, methods, stats, and utils. Contributed packages listed as follow and not in any particular order: tidyverse 2.0.0, dplyr 1.1.2, forcats 1.0.0, ggplot2 3.4.2, lubridate 1.9.2, purrr 1.0.1, readr 2.1.4, stringr 1.5.0, tibble 3.2.1, tidyr 1.3.0, effectsize 0.8.3, ggvis ver.

0.4.8, pacman ver. 0.5.1, rappdirs ver. 0.3.3, rio ver. 0.5.29, FunnelPlotR ver. 0.4.1, Meta ver. 6.2-1, metadat ver. 1.2-0, metafor ver. 4.0-0, esc ver. 0.5.1, and rstatix ver. 0.7.2.

Standardized mean differences (SMD) were calculated as effect sizes with 95% confidence results by using Hedge's G formula for paired sample tests. An example code is of the following: m.sm = metacont (n.e = sm$n.e, mean.e = sm$mean.e, sd.e = sm$sd.e, n.c = sm$n.c, mean.c = sm$mean.c, sd.c = sm$sd.c, studlab = (sm$Author...1), data = sm, sm = "SMD", method.smd = "Hedges", fixed = FALSE, random = TRUE, method.tau = "REML", hakn = TRUE, title = "Short to Medium Duration").

The above example code calculated effect size between durations named short and medium. This method was applied to all comparisons. Effect sizes were then pooled by duration groups being compared. Baseline to short duration, short to medium durations, medium to long durations and long to extended durations were paired and pooled using a random effects model (Hedges and Vevea, 1998). The *p*-values that resulted from this input were collected.

The null hypothesis was set as there was no clinically significant improvement observed between repeated measures. The alternative hypothesis was set as there was a clinically significant perceived improvement difference between two measured outcomes. If the calculated effect size failed to be less than or equal to the *p*-value of 0.05, then the group successfully rejected the null hypothesis and proved the alternative. On the other hand, if the *p*-value was greater than 0.05, then said group failed to reject the null hypothesis (Flechner and Tseng, 2011).

RESULTS

Quality assessment of studies

PEDro scores were evaluated for all articles used in this review (Table 1). Seven studies out of 15 were rated "excellent" (Argstatter et al., 2015; Bauer et al., 2017; Hall et al., 2022; Li et al., 2017; Li et al., 2019; Searchfield and Sanders, 2022; Yoo et al., 2022). Six of the processed studies were rated "good" (Henry et al., 2016; Jin et al., 2022; Mahboubi et al., 2017; Schad et al., 2018; Scherer et al., 2019; Westin et al., 2011). The remaining two studies received "fair" grades (Davis et al., 2007; Henry et al., 2006).

Characteristics of studies

Participants

All participants were adults older than 18 years of age

Table 1. Table of PEDro scale completed for all articles

Study	1	2	3	4	5	6	7	8	9	10	11	Study quality
Argstatter et al.(2015)	1	1	1	1	1	0	1	1	1	1	1	10
Bauer et al.(2017)	1	1	1	1	0	1	0	1	1	1	1	9
Davis et al.(2007)	0	1	0	0	0	0	0	1	1	1	1	5
Hall et al.(2022)	1	1	1	1	0	1	1	1	1	1	1	11
Henry et al.(2006)	0	1	0	0	0	0	0	1	1	1	1	5
Henry et al.(2016)	1	1	0	1	0	0	0	1	1	1	1	7
Jin et al.(2022)	1	1	1	0	1	0	0	1	1	1	1	8
Li et al.(2019)	1	1	1	0	1	1	1	1	1	1	1	10
Li et al.(2017)	1	1	1	0	1	1	1	1	1	1	1	10
Mahboubi et al.(2017)	1	0	0	1	0	0	0	1	1	1	1	6
Schad et al.(2018)	1	1	0	1	1	0	0	1	1	1	1	8
Scherer et al.(2019)	1	1	0	1	1	0	0	1	1	1	1	8
Searchfield & Sanders(2022)	1	1	1	1	1	0	0	1	1	1	1	9
Westin et al.(2011)	0	1	0	0	0	0	1	1	1	1	1	6
Yoo et al.(2022)	1	1	1	0	1	1	0	1	1	1	1	9

PEDro: physiotherapy evidence database

with the mean age of each study being higher than mid 40's. They were all sufferers of subjective tinnitus with duration being longer than 3 months. The range of hearing loss varied from study to study. Most studies specified what level of hearing loss was acceptable by setting a limit of no more than certain decibel (Argstatter et al., 2015; Bauer et al., 2017; Davis et al., 2007; Hall et al., 2022; Li et al., 2017; Schad et al., 2018; Scherer et al., 2019). Some studies did not have any restriction (Henry et al., 2016; Li et al., 2019; Mahboubi et al., 2017). Others reported average hearing loss level and did not use degree of hearing loss as a cause of exclusion (Jin et al., 2022; Westin et al., 2011; Yoo et al., 2022). Lastly, only one study failed to specify (Henry et al., 2006). The details of characteristics for included studies are depicted in Appendix 1.

Interventions

All studies analyzed the effects of ST or tinnitus masking in some form with different stimuli, duration and format. Majority of studies used broadband noise and white noise type stimuli as the masker (Bauer et al., 2017; Davis et al., 2007; Hall et al., 2022; Henry et al., 2006; Henry et al., 2016; Jin et al., 2022; Li et al., 2019; Mahboubi et al., 2017; Schad et al., 2018; Searchfield & Sanders, 2022; Westin et al., 2011) while the remaining studies used altered music that were either tailored according to tinnitus frequency or placebo (Argstatter et al., 2015; Li et al., 2017; Yoo et al., 2022). Some studies also included counseling or education as an additional intervention to the sound therapy or tinnitus masking (Argstatter et al., 2015; Bauer et al., 2017; Henry et al., 2006; Henry et al., 2016; Scherer et al., 2019; Searchfield & Sanders, 2022; Westin et al., 2011).

Control

Since this review's objective was to compare effect sizes between durations, control groups did not serve any meaningful purpose unless two types of stimuli were compared. In such a case, the two variables were pooled together to be combined into one result. Otherwise, controls in each study were largely ignored for this review.

Outcome

All studies used subjective evaluations from participants as outcome measures. There were various measuring methods used among studies. The number of outcome measures differed among studies. A few studies only used one form of measure (Argstatter et al., 2015; Henry et al., 2016; Jin et al., 2022; Searchfield & Sanders, 2022). Vast majority used a combination of outcome measures that mainly consisted of one or multiple questionnaires (TFI, THI, VAS, TQ, tinnitus evaluation questionnaire, THQ, HADS, The World Health Organization of life-brief, and ISI etc.) as well as other forms of measure such as matched masking level, loudness discomfort level, loudness/pitch/bandwidth matching (Bauer et al., 2017; Davis et al., 2007; Hall et al., 2022; Henry et al., 2006; Jin et al., 2022; Li et al., 2017; Li et al., 2019; Mahboubi et al., 2017; Schad et al., 2018; Searchfield & Sanders, 2022; Westin et al., 2011; Yoo et al., 2022). However, most of these were secondary measures. For this review, each article's declared primary measures were used as long as mean differences were calculable. Lastly, this study's focus was on sound therapy. Any outcomes measured as results of sound therapy were pooled as long as the same measures were available.

Repeated measures

To analyze effects of therapies over time, outcomes were measured at certain intervals. Repeated measures analyze outcomes measured at different times on the same therapy or treatment (Singh et al., 2013). However, most studies did not extend their investigation to the 18 months mark. For this review, only studies with duration pairs available for effect size calculations were used for each group.

Study designs

Prior to article selection, it was agreed that all studies must be conducted as randomized controlled trials to be qualified. RCTs are designed with the objective of studying the effects of a new therapy or treatments by randomizing participants. The act of randomizing contributed to minimizing biases (Hariton & Locascio, 2018). Therefore, all studies were RCTs at varying levels of blinding.

Overall results

Table 2 presented the overall result comparisons of effect sizes and *p*-values calculated with Hedge's g method after mean values, mean differences and standard deviations were collected.

Majority of effect sizes were in the negatives, which meant that ST did prove beneficial to participants. This study, however, focused on attempting to find the diminishing return for time investment of ST. The duration group of baseline to short group displayed the largest absolute value, indicating the greatest difference between two intervals. The calculated *p*-value was smaller than 0.0001, meaning there was a meaningful significant difference (Figure 2). Short to medium duration's SMD exhibited enough difference to be clinically significant with the *p*-value of 0.0342 (Figure 3). However, this value was a noticeable jump in *p*-value compared to that of the previous group. As the therapy duration entered 6 months or more (medium to long and long to extended duration groups), outcome measures failed to reject the null hypothesis (Figures 4, 5). This meant that the diminishing return for patients' time investment occurs once therapy had been ongoing for 6 months.

Table 3 presented the heterogeneity calculated by using Higgin's and Thompson's I^2 for each comparison. Baseline to

Table 2. Results of effect size (SMD) and *p*-values calculated through Hedge's G

Duration group	Effect size (SMD)	*p*-value
Baseline to short	-0.6224	< 0.0001
Short to medium	-0.2381	0.0342
Medium to long	-0.1805	0.067
Long to extended	-0.1977	0.4265

SMD: standard mean difference

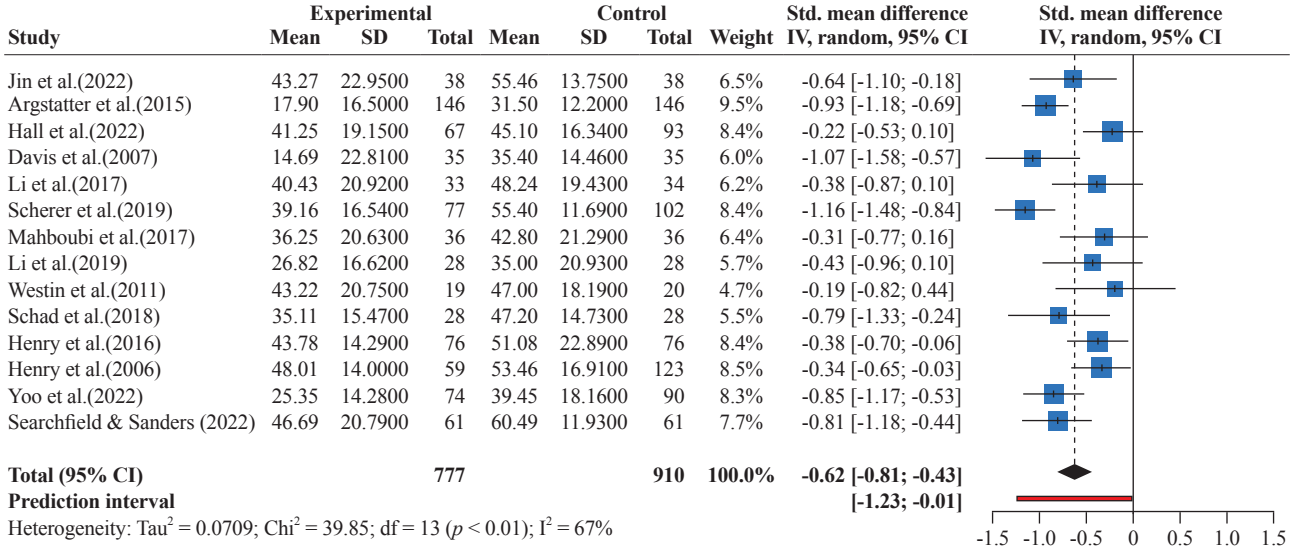

Figure 2. Forest plot of effect size for comparing baseline to short. SD: standard deviation, Std.: standard deviation, CI: confidential interval, df: degree of freedom.

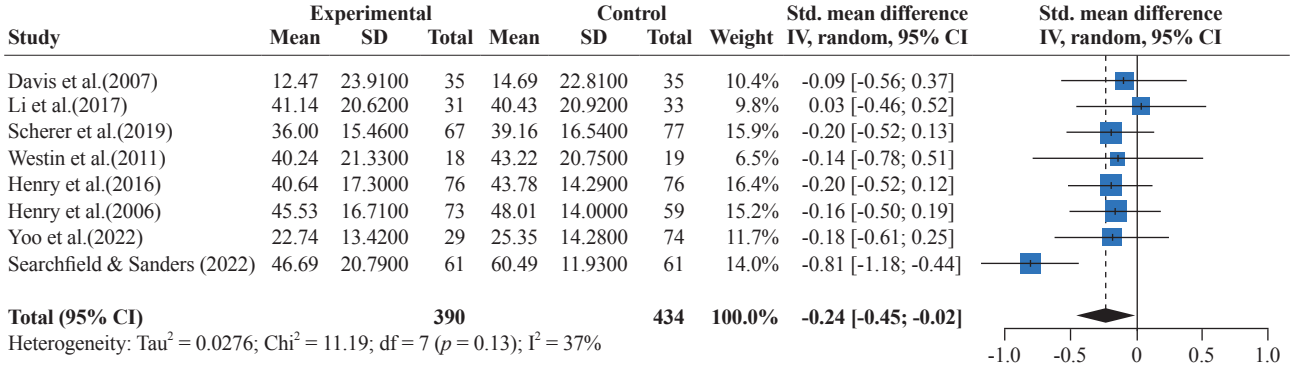

Figure 3. Forest plot of effect size for comparing short to medium duration. SD: standard deviation, Std.: standard deviation, CI: confidential interval, df: degree of freedom.

Study	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. mean difference IV, random, 95% CI
Bauer et al.(2017)	18.60	10.9000	19	26.10	14.1000	19	6.0%	-0.58 [-1.23; 0.07]
Davis et al.(2007)	11.21	23.9000	34	12.47	23.9100	35	11.4%	-0.05 [-0.52; 0.42]
Li et al.(2017)	40.22	26.2300	28	41.14	20.6200	31	9.7%	-0.04 [-0.55; 0.47]
Scherer et al.(2019)	34.53	16.0500	64	36.00	15.4600	67	21.6%	-0.09 [-0.44; 0.25]
Henry et al.(2016)	39.52	19.9100	76	40.64	17.3000	76	25.1%	-0.06 [-0.38; 0.26]
Henry et al.(2006)	38.31	20.0100	89	45.53	16.7100	73	26.1%	-0.39 [-0.70; -0.07]
Total (95% CI)			**310**			**301**	**100.0%**	**-0.18 [-0.38; 0.02]**
Prediction interval								**[-0.41; 0.05]**

Heterogeneity: $Tau^2 = 0$; $Chi^2 = 4.52$; df = 5 ($p = 0.48$); $I^2 = 0\%$

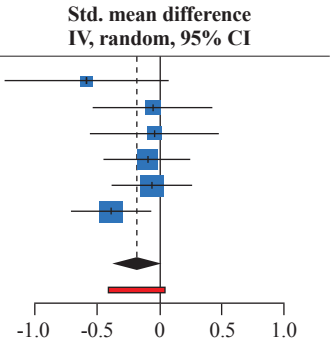

Figure 4. Forest plot comparing medium to long durations. SD: standard deviation, Std.: standard deviation, CI: confidential interval, df: degree of freedom.

Study	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. mean difference IV, random, 95% CI
Bauer et al.(2017)	17.30	12.3000	19	18.60	10.9000	19	19.0%	-0.11 [-0.75; 0.53]
Scherer et al.(2019)	37.33	17.1500	74	34.53	16.0500	64	26.6%	0.17 [-0.17; 0.50]
Henry et al.(2016)	39.04	19.9200	76	39.52	19.9100	76	27.0%	-0.02 [-0.34; 0.29]
Henry et al.(2006)	21.80	21.8400	94	38.31	20.0100	89	27.4%	-0.78 [-1.09; -0.48]
Total (95% CI)			**263**			**248**	**100.0%**	**-0.20 [-0.88; 0.49]**
Prediction interval								**[-2.18; 1.78]**

Heterogeneity: $Tau^2 = 0.1611$; $Chi^2 = 20.15$; df = 3 ($p < 0.01$); $I^2 = 85\%$

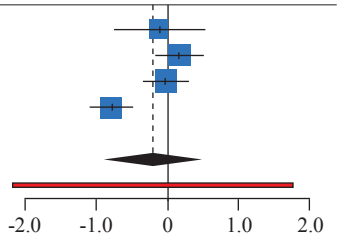

Figure 5. Forest plot comparing long to extended durations. SD: standard deviation, Std.: standard deviation, CI: confidential interval, df: degree of freedom.

Table 3. I^2 value of all groups

Duration group	I^2
Baseline to short	67.40
Short to medium	37.40
Medium to long	0.0
Long to extended	85.10

short group had a substantial heterogeneity at 67.40%. Short to medium group had a moderate heterogeneity at 37.40%. Medium to long group had no notable heterogeneity. Long to extended group had a considerable heterogeneity at 85.10%.

DISCUSSIONS

Previous studies attempted to prove efficacy of ST and produced mixed results. However, there were no reviews attempting to find effectiveness of ST over time. This analysis focused on finding the point of diminishing return by comparing outcome measures of studies that used ST. According to *p*-values calculated from paired sample *t*-tests (Figures 2-5), patients experienced clinically meaningful improvement up to the first 6 months of therapy and stagnated afterwards. The long to extended group (Figure 5) seemed to be the point of contention, however. Of the four studies, only one study reported a continued improvement at 18 months (Henry et al., 2006) while the other three studies reported a minor relapse in tinnitus symptom (Bauer et al., 2017; Henry et al., 2016; Scherer et al., 2019).

There were many limitations to this analysis. Primary limitation being the limited number of articles available that met the condition of only ST and RCT. After screening, less than 15 articles remained. Among the available articles, repeated outcome measures were inconsistent and spread out. Some articles only measured outcomes with increments of weeks (Hall et al. 2022; Li et al., 2019; Schad et al., 2018; Searchfield & Sanders, 2022) as well as a mix of weeks and months (Westin et al., 2011). Even among articles that reported findings in months, not all studies continued tracking participants past 6 or 12 months (Davis et al., 2007; Li et al., 2017; Yoo et al., 2022) and very few kept records at 18 months (Bauer et al., 2017; Henry et al., 2006; Henry et

al., 2016). One article did not measure outcome in between baseline and the next (Bauer et al., 2017). There was a general lack of standardization contributed to a decreasing credibility in pooled effect sizes for the longer duration groups. Additionally, this study analyzed the total duration of the intervention, but did not comparie the effects based on the amount of sound therapy conducted per day. For example, it could not analyzed the effects of performing ST for 3 hours a day for 3 months versus 1 hour a day for 6 months. Future research sound address this aspect.

This analysis concluded that patients could expect a high improvement rate within the first 6 months of sound therapy but diminished rate afterwards up to 12 months and eventually stagnate and relapse. However, this should not lead to patients making presumptions that sound therapy is not helpful past 6 months and onwards. This reaffirmed that sound therapy is an effective relief but not a permanent solution for tinnitus. As such, it is recommended that patients habituate sound therapy usage into their daily lives while researchers prioritize understanding physiology behind tinnitus and find an effective treatment.

Ethical Statement

N/A

Acknowledgments

N/A

Declaration of Conflicting Interests

There is no conflict of interest.

Funding

N/A

Author Contributions

Conceptualizing: Woojin, J. Baek, Junghwa Bahng, Data curation: Woojin, J. Baek. Methodolgy: Woojin, J. Baek. Writing original draft: Woojin, J. Baek. Writing-review & editing: Junghwa Bahng.

ORCID iDs

Woojin Jayce Baek https://orcid.org/0009-0000-7326-6481
Junghwa Bahng https://orcid.org/0000-0002-5265-3586

REFERENCES

Argstatter, H., Grapp, M., Hutter, E., Plinkert, P. K., & Bolay, H. V. (2015). The effectiveness of neuro-music therapy according to the Heidelberg model compared to a single session of educational counseling as treatment for tinnitus: A controlled trial. *Journal of Psychosomatic Research*, *78(3)*, 285-292.

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *BMJ Ment Health*, *22(4)*, 153-160.

Bauer, C. A., Berry, J. L., & Brozoski, T. J. (2017). The effect of tinnitus retraining therapy on chronic tinnitus: A controlled trial. *Laryngoscope Investigative Otolaryngology*, *2(4)*, 166-177.

Belli, S., Belli, H., Bahcebasi, T., Ozcetin, A., Alpay, E., & Ertem, U. (2008). Assessment of psychopathological aspects and psychiatric comorbidities in patients affected by tinnitus. *European Archives of Oto-Rhino-Laryngology*, *265*, 279-285.

Cederroth, C. R., Gallus, S., Hall, D. A., Kleinjung, T., Langguth, B., Maruotti, A., et al. (2019). Towards an understanding of tinnitus heterogeneity. *Frontiers in Aging Neuroscience*, *11*, 446253.

Davis, P. B., Paki, B., & Hanley, P. J. (2007). Neuromonics tinnitus treatment: Third clinical trial. *Ear and Hearing, 28(2)*, 242-259.

Flechner, L. & Tseng, T. Y. (2011). Understanding results: P:-values, confidence intervals, and number need to treat. *Indian Journal of Urology*, *27(4)*, 532-535.

Formby, C., Yang, X., & Scherer, R. W. (2022). Contributions of counseling and sound generator use in tinnitus retraining therapy: Treatment response dynamics assessed in a secondary analysis of a randomized trial. *Journal of Speech, Language, and Hearing Research*, *65(2)*, 816-828.

Friberg, E., Jansson, C., Mittendorfer-Rutz, E., Rosenhall, U., & Alexanderson, K. (2012). Sickness absence due to otoaudiological diagnoses and risk of disability pension: A nationwide Swedish prospective cohort study. *PLoS One*, *7(1)*, e29966.

Hall, D. A., Pierzycki, R. H., Thomas, H., Greenberg, D., Sereda, M., & Hoare, D. J. (2022). Systematic evaluation of the T30 neurostimulator treatment for tinnitus: A double-blind randomised placebo-controlled trial with open-label extension. *Brain Sciences*, *12(3)*, 317.

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). *Doing Meta-analysis with R: A Hands-on Guide*. New York: Chapman and Hall/CRC.

Hariton, E. & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology*, *125(13)*, 1716.

Hedges, L. V. & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3(4)*, 486.

Henry, J. A., Schechter, M. A., Zaugg, T. L., Griest, S., Jastreboff, P. J., Vernon, J. A., et al. (2006). Clinical trial to compare tinnitus masking and tinnitus retraining therapy. *Acta Oto-Laryngologica*, *126(sup556)*, 64-69.

Henry, J. A., Stewart, B. J., Griest, S., Kaelin, C., Zaugg, T. L., & Carlson, K. (2016). Multisite randomized controlled trial to compare two methods of tinnitus intervention to two control conditions. *Ear and Hearing*, *37(6)*, e346-e359.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327(7414)*, 557-560.

Hinton, D. E., Chhean, D., Pich, V., Hofmann, S. G., & Barlow, D. H.

(2006). Tinnitus among Cambodian refugees: relationship to PTSD severity. *Journal of Traumatic Stress*, *19(4)*, 541-546.

Hobson, J., Chisholm, E., & El Refaie, A. (2012). Sound therapy (masking) in the management of tinnitus in adults. *Cochrane Database of Systematic Reviews*, *(11)*, CD006371.

Jarach, C. M., Lugo, A., Scala, M., van den Brandt, P. A., Cederroth, C. R., Odone, A., et al. (2022). Global prevalence and incidence of tinnitus: A systematic review and meta-analysis. *JAMA Neurology*, *79(9)*, 888-900.

Jastreboff, P. J. (1990). Phantom auditory perception (tinnitus): Mechanisms of generation and perception. *Neuroscience Research*, *8(4)*, 221-254.

Jastreboff, P. J. & Jastreboff, M. M. (2006). Tinnitus retraining therapy: A different view on tinnitus. *ORL*, *68(1)*, 23-30.

Jin, I. K., Choi, S. J., Ku, M., Sim, Y., & Lee, T. (2022). The impact of daily hours of sound therapy on tinnitus relief for people with chronic tinnitus: A randomized controlled study. *Journal of Speech, Language, and Hearing Research*, *65(8)*, 3079-3099.

Job, A., Raynal, M., & Kossowski, M. (2007). Susceptibility to tinnitus revealed at 2 kHz range by bilateral lower DPOAEs in normal hearing subjects with noise exposure. *Audiology and Neurotology*, *12(3)*, 137-144.

Langguth, B., Kreuzer, P. M., Kleinjung, T., & De Ridder, D. (2013). Tinnitus: Causes and clinical management. *The Lancet Neurology*, *12(9)*, 920-930.

Li, S. A., Bao, L., & Chrostowski, M. (2017). Investigating the effects of a personalized, spectrally altered music-based sound therapy on treating tinnitus: A blinded, randomized controlled trial. *Audiology and Neurotology*, *21(5)*, 296-304.

Li, Y., Feng, G., Wu, H., & Gao, Z. (2019). Clinical trial on tinnitus patients with normal to mild hearing loss: Broad band noise and mixed pure tones sound therapy. *Acta Oto-Laryngologica*, *139(3)*, 284-293.

Mahboubi, H., Haidar, Y. M., Kiumehr, S., Ziai, K., & Djalilian, H. R. (2017). Customized versus noncustomized sound therapy for treatment of tinnitus: A randomized crossover clinical trial. *Annals of Otology, Rhinology & Laryngology*, *126(10)*, 681-687.

Moseley, A. M., Rahman, P., Wells, G. A., Zadro, J. R., Sherrington, C., Toupin-April, K., et al. (2019). Agreement between the Cochrane risk of bias tool and physiotherapy evidence database (PEDro) scale: A meta-epidemiological study of randomized controlled trials of physical therapy interventions. *PLoS One*, *14(9)*, e0222770.

Rohatgi, A. (2022, Jun 20). *Webplotdigitizer: Version 4.5 (2021)*. Automeris. Retrieved from https://automeris.

Schad, M. L., McMillan, G. P., Thielman, E. J., Groon, K., Morse-Fortier, C., Martin, J. L., et al. (2018). Comparison of acoustic therapies for tinnitus suppression: A preliminary trial. *International Journal of Audiology*, *57(2)*, 143-149.

Scherer, R. W., Formby, C., & Tinnitus Retraining Therapy Trial Research Group. (2019). Effect of tinnitus retraining therapy vs standard of care on tinnitus-related quality of life: A randomized clinical trial. *JAMA Otolaryngology–Head & Neck Surgery*, *145(7)*, 597-608.

Searchfield, G. D. & Sanders, P. J. (2022). A randomized single-blind controlled trial of a prototype digital polytherapeutic for tinnitus. *Frontiers in Neurology*, *13*, 958730.

Singh, V., Rana, R. K., & Singhal, R. (2013). Analysis of repeated measurement data in the clinical trials. *Journal of Ayurveda and Integrative Medicine*, *4(2)*, 77.

Weisz, N., Hartmann, T., Dohrmann, K., Schlee, W., & Norena, A. (2006). High-frequency tinnitus without hearing loss does not mean absence of deafferentation. *Hearing Research*, *222(1-2)*, 108-114.

Westin, V. Z., Schulin, M., Hesser, H., Karlsson, M., Noe, R. Z., Olofsson, U., et al. (2011). Acceptance and commitment therapy versus tinnitus retraining therapy in the treatment of tinnitus: A randomised controlled trial. *Behaviour Research And Therapy*, *49(11)*, 737-747.

Yoo, S., Yakunina, N., & Nam, E. C. (2022). Does listening to tinnitus frequency-filtered music relieve tinnitus? *Journal of Audiology & Otology*, *26(3)*, 147.

APPENDIX

Appendix 1. Details of all studies included in this systematic review

Study	Participants (age, hearing level)	Intervention	Control group	Study design	Outcome measures
Argstatter et al.(2015)	-N = 146 -Mean age = 45.1 -Hearing loss = no greater than 60 dB HL in the region of the center tinnitus frequency	-Duration: 18-50 (n = 3), 51-65 (n = 11), and 66-75 (n = 5) sessions of individualized -5 days for a week, twice a day -Training material neuro-music therapy and one individual counseling session	-N = 144 -Mean age = 53.2 -Hearing loss = no greater than 60 dB HL in the region of the center tinnitus frequency	Neuro-music stimuli + counseling compared to control group that only received counseling	-Measures: tinnitus questionnaire -Repeated measures: 1 week
Bauer et al.(2017)	-N = 19 -Mean age = 18-50 (n = 3), 51-65 (n = 14), and 66-75 (n = 2) -Hearing loss = lower than 100 dB SPL LDL	-Duration: 31 hours TRT directive counseling according -18 months of sound generator -Training material binaural combination devices	-N = 19 -Mean age = 18-50 (n = 3), 51-65 (n = 11), and 66-75 (n = 5) -Hearing loss = lower than 100 dB SPL LDL	Comparing TRT and standard care	-Measures: THI, TFI, TEQ, TIQ -Repeated measures: 6, 12, and 18 months
Davis et al.(2007)	-N = 19 -Mean age = 56.1 -Hearing loss = less than 50 dB	-Duration: 12 months -Training material -Sound player device and individually customized acoustic stimulus (commercially available music recordings with a specially designed BBN component	-N = 16 -Mean age = 61.3 -Hearing loss = less than 50 dB	Intermittent vs. complete coverage of perception	-Measures: TRQ, VAS, MML, LDL -Repeated measures: 2, 4, 6, and 12 months
Hall et al.(2022)	-N = 33 -Mean age = 49.1 -Hearing loss = less than 60 dB HL PTA	-Duration: 36 weeks -Training material: T30 neurostimulator device and headphones	-N = 37 -Mean age = 51.8 -Hearing loss = less than 60 dB HL PTA	T30 neurostimulator vs. placebo	-Measures: THQ, TFI, EEG, WHOQOL-BREF, VAS, loudness, pitch, and bandwidth -Repeated measures: 2, 4, 8, 12, 24, and 36 weeks
Henry et al.(2006)	-N = 64 -Mean age = 58.7 -Hearing loss = NA	-Duration: 18 months -Training material: ear-level devices (hearing aids, noise generators, or combinations)	-N = 59 -Mean age = 61 -Hearing loss = NA	TM vs. TRT	-Measures: THI, THQ, TSI, the TRT follow-up interview -Repeated measures: 3, 6, 12, and 18 months
Henry et al.(2016)	-TRT (n = 34) -TRT (mean age = 60.1) -TM (n = 42) -TM (mean age = 62.4) -TED (n = 39) -TED (mean age = 62.7) -Hearing loss = all participants varied in hearing level	-Duration: 18 months -Training material: ear-level devices (hearing aids, noise generators, or combinations)	-WLC (n = 33) -WLC (mean age = 61.2) -Hearing loss = varied in hearing levels	Comparison between TRT, TM, TED and WLC	-Measures: THI -Repeated measures: 3, 6, 12, and 18 months
Jin et al.(2022)	-1 hour (n = 20) -1 hour (mean age = 48) -1 hour average PTA = L, 14; R, 13.5 -3 hours (n = 19) -3 hours (mean age = 44.94) -3 hours average PTA = L, 11.66; R, 10.96 -5 hours (n = 19) -5 hours (mean age = 48.52) -5 hours average PTA = L, 15.35; R, 15.61	-Duration: 3 months -Training material: education, smart phone, and headphones	NA	Effect comparison between 1, 3, and 5 hours daily sound therapy	-Measures: K-TPFQ -Repeated measures: 3 months

Appendix 1. Continued

Study	Participants (age, hearing level)	Intervention	Control group	Study design	Outcome measures
Li et al.(2019)	-N = 14 -Mean age = 49 -Hearing loss = no restriction	-Duration: 3 months -Training materials: sound-isolation booth, and LabVIEW 15.0 to generate sounds, earphones and mobile phones to listen	-N = 14 -Mean age = 41 -Hearing loss = not restricted	Comparison BBN sound therapy between mixed pure tone sound therapy	-Measures: THI, VAS -Repeated measures: 2, 4, 8, and 12 weeks
Li et al.(2017)	-N = 15 -Mean age = 55.20 -Hearing loss = < 70 dB at all frequencies	-Duration: 12 months -Training material: MP3 players and earphones	-N = 19 -Mean age = 56.16 -Hearing loss = < 70 dB at all frequencies	Comparison between spectrally altered music sound therapy and unaltered classical music therapy	-Measures: THI, TFI, HADS -Repeated measures: 3, 6, and 12 months
Mahboubi et al.(2017)	-N = 18 -Mean age = 53 -Hearing loss = no restriction	Duration: 3 months, then switch with 3 weeks washout -Training material: MP3 player, customized stimuli	-N = 18 -Mean age = 53 -Hearing loss = no restriction	Comparison between customized sound therapy and BBN	-Measures: mean loudness, THI, BAI, and MML -Repeated measures: 3 months
Schad et al.(2018)	-Notched (n = 10) -Matched (n = 10) -Mean age of all groups = 58 -Hearing loss = no single frequency air-bone gap greater than 15 dB	-Duration: 1 month -Training material: iPod Nano with Ety-Kids 5 earphones	-Placebo (n = 10) -Mean age of all groups = 58 -Hearing loss = no single frequency air-bone gap greater than 15 dB	Comparison among notched, matched and placebo BBN	-Measures: VNS, LM, and TFI -Repeated measures: 2 and 4 weeks
Scherer et al.(2019)	-TRT (n = 49) -TRT (mean age = 51.1) -pTRT (n = 51) -pTRT (n = 50.9) -Hearing loss for all = normal to mild through 8,000 Hz	-Duration: 18 months -Training materials: sound generators, tinnitus-specific educational counseling	-SoC (n = 45) -SoC (mean age = 49.9) -Hearing loss for all = normal to mild through 8,000 Hz	Comparison among TRT, pTRT, and SoC	-Measures: TQ, TFI, THI, and VAS -Repeated measures: 3, 6, 12, and 18 months
Searchfield & Sanders (2022)	-N = 35 -Mean age = 53 -Hearing loss = yes, but majority do not use HA	-Duration: 16 weeks -Training materials: white noise and environmental sound, headphones and neck speaker	-N = 36 -Mean age = 53 -Hearing loss = yes, but majority do not use HA	Comparing UpSilent (USL) TRT and white noise	-Measures: TFI -Repeated measures: 6 and 12 weeks
Westin et al.(2011)	-ACT (n = 20) -ACT (mean age = 53.5) -TRT (n = 20) -TRT (mean age = 48.95) -Hearing loss = all participants averaged about 12.8 dB HL for the better ear	-Duration: 18 months -Training materials: open fitted sound generators, consultation	-WLC (n = 22) -WLC (mean age = 49.59) -Hearing loss = all participants averaged about 12.8 dB HL for the better ear	Comparison among ACT, TRT, and WLC	-Measures: THI, ISI, QOLI, HADS, and TAQ -Repeated measures: 10 weeks, 6 months, and 18 months
Yoo et al.(2022)	-N = 45 -Mean age = 42.7 -Hearing loss = PTA low: 11, PTA mid: 12.8, and PTA high: 35.2	-Duration: 6 months -Training materials: favorite music and smartphones	-N = 45 -Mean age = 42.2 -Hearing loss = PTA low: 9.2, PTA mid: 11.5, and PTA high: 25.6	Comparing music stripped of an octave band centered on the tinnitus frequency and music stripped of random frequency	-Measures: matched loudness (dB HL), MML (dB SL), VAS, and THI -Repeated measures: 3 and 6 months

N: number, HL: hearing loss, SPL: sound pressure level, LDL: loudness discomfort level, TRT: tinnitus retraining therapy, THI: tinnitus handicap inventory, TFI: tinnitus frequency inventory, TEQ: tinnitus cognitions questionnaire, TIQ: tinnitus Impact questionnaire, BBN: broadband noise, TRQ: tinnitus reaction questionnaire, VAS: visual analogue scale, MML: matched masking level, PTA: puretone avergage, THQ: tinnitus handicap questionnaires, EEG: electroencephalogram, WHOQOL-BREF: The World Health Organization Quality of Life-brief, NA: no applicable, TM: tinnitus masking, TSI: tinnitus severity intex, TED: technology, entertainment, design, WLC: wait list control, L: left, R: right, K-TPFQ: Korean version-tinnitus primary function questionnaire, LabVIEW: laboratory virtual instrument engineering workbench, MP3: MPEG audio layer 3, HADS: hospitals anxiety and depression scales, BAI: beck anxiety inventory, VNS: vagus nerve stimulation, LM: loudness matching, pTRT: pre tinnitus retraining therapy, SoC: standard of care, TQ: tinnitus questionnaires, HA: hearing aids, USL: upsilent, ACT: acceptance and commitment therapy, ISI: insomnia severity index, QOLI: quality of life inventory, TAQ: tinnitus acceptance questionnaire, SL: sensation level