Hearing Research 416 (2022) 108442

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Speech-in-noise perception in musicians and non-musicians: A multi-level meta-analysis

Sarah Hennessy[a], Wendy J. Mack[b], Assal Habibi[a,*]

[a] Brain and Creativity Institute, University of Southern California, Los Angeles, CA, United States
[b] Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, United States

ARTICLE INFO

Article history:
Received 17 May 2021
Revised 10 January 2022
Accepted 13 January 2022
Available online 15 January 2022

Keywords:
Speech-in-noise
Music training
Auditory perception
Meta-analysis

ABSTRACT

Speech-in-noise perception, the ability to hear a relevant voice within a noisy background, is important for successful communication. Musicians have been reported to perform better than non-musicians on speech-in-noise tasks. This meta-analysis uses a multi-level design to assess the claim that musicians have superior speech-in-noise abilities compared to non-musicians. Across 31 studies and 62 effect sizes, the overall effect of musician status on speech-in-noise ability is significant, with a moderate effect size ($g = 0.58$), 95% CI [0.42, 0.74]. The overall effect of musician status was not moderated by within-study IQ equivalence, target stimulus, target contextual information, type of background noise, or age. We conclude that musicians show superior speech-in-noise abilities compared to non-musicians, not modified by age, IQ, or speech task parameters. These effects may reflect changes due to music training or predisposed auditory advantages that encourage musicianship.

1. Introduction

Speech-in-noise (SIN) perception refers to the ability to hear a relevant voice in the context of a noisy background (e.g., a loud restaurant or gathering). Understanding speech in noisy environments is vital to successful communication. Deficits in SIN abilities are associated with reduced school performance in children (de Carvalho et al., 2017), and increased emotional and social loneliness in older adults (Stam et al., 2016). Speech-in-noise abilities decline with age (Pronk et al., 2013a) and are difficult to remedy with assistive technology (Chung, 2004; Killion, 1997). Important to our understanding of human auditory processing is untangling how speech-in-noise abilities vary across individuals and can be improved across the lifespan.

One group of individuals that is reported to have better SIN abilities in some studies is lifelong musicians (for review, see Coffey et al., 2017). Music training may improve auditory processing through repetitive practice of fine-tuned pitch discrimination and enhanced attendance to changes in acoustic features such as timbre and rhythm. According to the OPERA (overlap, precision, emotion, repetition, attention) hypothesis (Patel, 2011, 2014), music training may influence speech processing specifically because it places demands on shared sensory or cognitive processes and involves repetition, attentional focus, and is emotionally rewarding. Patel (2014) suggests that music training may drive neural plasticity by placing a higher demand on overlapping brain networks that process music and speech than in everyday speech communication. It may also be the case that individuals who choose to become musicians have pre-existing advantages in perceptual abilities (including, but not limited to speech-in-noise perception) that may help them to excel at music playing and thus continue to participate in musical activities. Thus, musicians may have better speech-in-noise abilities than non-musicians as a function of predisposed auditory advantages that encourage continued music training, rather than as a result of music training.

Several cross-sectional studies have reported differences in speech-in-noise performance between musicians and non-musicians (Parbery-Clark et al., 2009; Zendel et al., 2015), while others have observed no differences (e.g., Boebinger et al., 2015; Madsen et al., 2017, 2019). Discrepancies between studies may be due to differences in task selection; there is great variety in tasks used to assess speech-in-noise abilities – for example, many researchers (e.g., Zendel and Alain, 2012) have used QuickSIN (Etymotic Research, 2001; Killion et al., 2004) while others (e.g., Parbery-Clark et al., 2013) use Hearing In Noise Test (HINT) (Neff and Green, 1987), or Words in Noise (WIN) (e.g.: Slater and Kraus, 2016). Many studies report results from multiple tasks (e.g., Bidelman and Yoo, 2020; Escobar et al., 2019). Each task measures speech-in-noise perception in slightly different ways; QuickSIN (Killion et al., 2004), for example, assesses perception of full-length meaningful and grammatical sentences that are embedded

* Corresponding author.
E-mail address: ahabibi@usc.edu (A. Habibi).

https://doi.org/10.1016/j.heares.2022.108442
0378-5955/

in 4-talker speech babble that gradually increases in volume, while HINT (Nilsson et al., 1994) has similar, yet more highly predictable, sentences embedded in speech-shaped noise where the target volume is adaptive relative to participant performance. In contrast, WIN (Wilson, 2003) consists of monosyllabic words embedded in adaptive babble, without any grammatical or semantic context. It has been reported that some speech-in-noise tasks may be more sensitive than others. For example, HINT, because most users achieve higher overall performance (at ceiling), is poorer at discriminating between individuals with and without hearing loss than is QuickSIN and the WIN (Wilson et al., 2007). Relatedly, differences between studies may be due to target speech type, where some studies use tasks where the target speech is a sentence (e.g., Anaya et al., 2016; Başkent and Gaudrain, 2016), while others choose syllables (e.g., Du and Zatorre, 2017) or words (e.g., Fostick, 2019) as the target. These stimuli may be processed differently, relying on different encoding cues and cognitive resources. Sentences often contain semantic and grammatical information that allows the listener to make predictions and fill in gaps about missed information and words. Words, while not always embedded in contextual cues, follow predictable patterns of consonant and vowel combinations. In contrast, syllables and made-up words do not contain predictable information, and thus perception is less able to depend on top-down mechanisms. Type of background noise also varies across studies, as some researchers choose a single voice masking paradigm (e.g., Başkent and Gaudrain, 2016), while others use babble (e.g., Escobar et al., 2019), or speech-shaped noise (e.g., Fuller et al., 2014). SIN perception outside of the laboratory is supported by contextual cues and syntactical information and is typically in the presence of speech-related background noise, and thus the choice of speech target type, contextual information available, or background noise may lead to critical differences in performance in a measurement setting.

Lastly, in a recent meta-analysis, speech-in-noise abilities were found to be positively associated with cognitive abilities, including working memory and IQ (Dryden et al., 2017). This was suggested to be due to a cognitively-related ability to effectively use contextual cues, as this association was particularly strong in contextually-rich tasks. Musicians, as compared to non-musicians, have been reported to have higher auditory working memory (Chan et al., 1998; Moreno et al., 2011; Talamini et al., 2016) and verbal and nonverbal IQ (Schellenberg, 2011), although differences in intelligence may be simply due to differences in music aptitude rather than a result of training (Swaminathan et al., 2017). Thus, to assess whether musicians have superior speech-in-noise abilities compared to non-musicians independent of cognitive ability, it may be important to control for IQ and auditory working memory measures. While several studies have explored this idea and controlled for cognitive ability (e.g., Boebinger et al., 2015), many have not.

In this meta-analysis, we assess the hypothesis that musicians have superior speech-in-noise ability compared to non-musicians. We restrict our analysis to cross-sectional studies due to insufficient longitudinal studies and randomized control trials that would be necessary to conduct this analysis. To our knowledge, this is the first meta-analysis to explore this question. We explore four main questions: 1) do adult musicians perform better than non-musicians in tasks measuring speech-in-noise?, 2) how much variance can be attributed to within-study effects (specifically in studies with multiple SIN tasks) as compared to between-study effects?, 3) are observed effects dependent on the type of speech target (e.g., sentences vs. words), 4) are observed effects dependent on whether participants were explicitly identified as equivalent in cognitive ability? Given that speech-in-noise abilities decline with age (Pronk et al., 2013b) and that many studies have investigated SIN perception specifically in older adult participants, we additionally explore a fifth question: 5) are observed effects dependent on the age group assessed (e.g., older adults vs. younger adults)?

2. Methods

2.1. Literature search

A literature search using PubMed and ProQuest was conducted. The first author designed the search method, and terms used in each search are listed in (Supplementary Table 1). The search was conducted once in April 2020 and updated again in February 2021. After removal of duplicates, we retrieved 3011 records.

2.2. Inclusion criteria

Articles that met the following criteria were included in the meta-analysis:

1. Participants were adults with normal hearing thresholds defined as less than 20 dB HL from 250 to 4000 Hz.
2. The study was cross-sectional and included a long-term musically-trained (> 6 years of training) group and a musically-untrained control group (i.e., music training as a categorical, not continuous, variable).
3. The study was a peer-reviewed publication, published in English. Dissertations, theses, conference proceedings, abstracts, unpublished manuscripts, and case studies were not included.
4. The study reported behavioral outcomes of speech-in-noise (i.e., sentences, words, or syllables in noise). Studies that reported auditory stimuli unrelated to speech (i.e., tones) in noise were not included.
5. The study reported sufficient data to compute effect sizes.

Excluded studies were those that did not meet all five inclusion criteria. Records were evaluated by the first author for eligibility based on inclusion criteria (see Fig. 1), resulting in 31 studies and 62 effect sizes included in the meta-analysis. We conducted a Risk-of-Bias assessment using criteria from the ROBINS-I tool for non-randomized studies (Sterne et al., 2016), which was visualized using the *robvis* package (McGuinness, 2019) in R (Supplementary Figure 1). Specifically, we assessed: 1) bias due to confounding (are there confounding factors (e.g.: IQ differences, age, pure-tone average) present that may have influenced outcomes, and did authors control for such factors?), 2) bias in classification of groups (were music and control groups clearly defined at the beginning of the study?), 3) bias due to missing data (were outcome data available for all, or nearly all participants? Were participants excluded due to missing data?), 4) bias in measurement of outcomes (Were methods of outcome assessment comparable across groups?) 5) bias in selection of reported result (is the effect estimate likely to be selected from multiple analyses of the group-outcome relationship or different subgroups?). It should be noted that Risk-of-Bias assessments are not infallible assessments of a study's quality, as they do not provide information on the relative impact of one bias vs. another, or the magnitude or direction of a bias (Savitz et al., 2019).

2.3. Outcome measures

Multiple outcome measures of SIN perception were allowed for each study. Outcome variables were coded as the name (e.g., "Hearing in Noise Task") and category (e.g., "masked speech") of the task performed. See Table 1 for a complete list of outcome variables for each included study, and Table 2 for a brief description of common outcome variables.

Table 1
Studies included in meta-analysis.

Study, Journal	Number of Effect Sizes included	Outcome Measure	Speech Target/ Noise Type/ Context	Duration of Music Training and Age of Onset in the Musician Group	Duration of Music Training and Age of Onset in the Non-musician Group	Mean Age, total N (musician N)	IQ Measures
Anaya et al., 2016, *The Journal of the Acoustical Society of America*	1	*HINT* and *PRESTO composite*	Sentences/ speech-shaped noise / semantic	15.45, onset age 4.9	1.7 years, not currently playing instrument	20.7, $N = 22$ (11)	Nonverbal[*]
Başkent and Gaudrain, 2016, *The Journal of the Acoustical Society of America*	1	Masked sentences	Sentences/ single speaker/ semantic	At least 10 years, onset age < 7	< 10 years, no training within the past 7 years	22.3, $N = 38$ (18)	none
Bidelman and Yoo, 2020, *Frontiers in Psychology*	2	Masked sentences QuickSIN	Sentences /babble / none Sentences/ babble/ syntactic	15.1, onset age 7.2	0.89 years	24.2, $N = 28$ (14)	Nonverbal[*], AWM[*]
Boebinger et al., 2015, *The Journal of the Acoustical Society of America*	4	Masked BKB sentences	Sentences/ single speaker, speech-shaped noise, fluctuating speech-shaped noise/ semantic	22.7, onset age 5.9	0.2 years, onset age 12.8	27.2, $N = 50$ (25)	Verbal, Nonverbal, AWM
Clayton et al., 2016, *PloS ONE*	2	Masked sentences, co-located, separated	Sentences/ single speaker/ syntactic	14.4	< 3 years, no current instrument playing	21.5, $N = 34$ (17)	Nonverbal, AWM[*]
Du and Zatorre, 2017, *PNAS*	1	Syllable in Noise	Syllables/ white noise/ none	16.3, onset age 5.1	< 1 year, no training in the past year	21.8, $N = 30$ (15)	Verbal, AWM
Escobar et al., 2019, *Ear and Hearing*	3	QuickSIN HINT SPIN-R	Sentences/ babble / syntactic Sentences, speech-shaped noise, semantic Sentences, babble, semantic	13.4, onset age 8.2	< 3 years, no training in the past 7 years	21.4, $N = 49$ (27)	AWM
Fostick, 2019, *European Journal of Ageing*	2	AB- words task in speech noise	Words/ speech-shaped noise, white noise/ none	7 hrs/week with 3 h in orchestral rehearsal	No training, no current instrument playing	65.6[†], $N = 46$ (23)	Nonverbal, AWM
Fuller et al., 2014, *Frontiers in Neuroscience*	2	Sentences in noise Words in noise	Sentences/ speech-shaped noise/ semantic Words/ speech-shaped noise/ none	14.6, onset age 5.8	1.6, onset age 9.1	22.7, $N = 50$ (25)	none
Kaplan et al., 2021, *Frontiers in Psychology*	1	Masked sentences	Sentences/ 2-talker maskers/ semantic	13.1, onset age 6.2	2.05 years, onset age 14.4	26.7, $N = 33$ (16)	none
Madsen et al., 2017, *Scientific Reports*	2	Masked HINT sentences	Sentences/ babble, speech-shaped noise/ semantic	14.6, onset age 6.1	< 2 years, no training in the past 7 years	21.0, $N = 60$ (30)	Verbal, nonverbal
Madsen et al., 2019, *Scientific Reports*	2	Closed and open speech-on-speech task (Dantale II sentences), separated, co-located	Sentences/ babble/ semantic	15.3, onset age 5.4	< 2 years, no training in the past 7 years	22.9, $N = 64$ (32)	Verbal, nonverbal
Mankel and Bidelman, 2018, *PNAS*	1	QuickSIN,	Sentences/ babble/ syntactic	16, onset age 7.14	< 3 years, no training in the past 5 years	22.2, $N = 28$ (14)	none
Meha-Bettison et al., 2018, *International Journal of Audiology*	4	LiSN-S, low and high cue, co-located and separated	Sentences/ single speaker/ semantic	39.7, onset age 7.2	No training	45.9, $N = 20$ (10)	none
Morse-Fortier et al., 2017, *Trends in Hearing*	4	Masked words (natural, vocoded, spatial, nonspatial)	Words/ babble, fluctuating speech-shaped noise/ none	11.5	No training	21.3, $N = 40$ (20)	none

(*continued on next page*)

Table 1 (*continued*)

Study, Journal	Number of Effect Sizes included	Outcome Measure	Speech Target/ Noise Type/ Context	Duration of Music Training and Age of Onset in the Musician Group	Duration of Music Training and Age of Onset in the Non-musician Group	Mean Age, total N (musician N)	IQ Measures
Parbery-Clark et al., 2009, *Ear and Hearing*	2	QuickSIN	Sentences/ babble/ syntactic	16, onset age 4.7	2 years, onset age 10.8	23, $N = 31$ (16)	Nonverbal, AWM*
		HINT	Sentences/ speech-shaped noise/ semantic				
Parbery-Clark et al., 2011, *PLoS ONE*	3	QuickSIN,	Sentences, / babble/ syntactic	50, onset age 5.6	$<$ 3 years training	54.5†, $N = 37$ (18)	Verbal, Nonverbal, AWM*
		HINT	Sentences/ speech-shaped noise/ semantic				
		WIN	Words/ babble/ none				
Parbery-Clark et al., 2011, *Neuropsychologia*	1	HINT	Sentences/ speech-shaped noise/ semantic	16.4, onset age 5.1	$<$ 3 years training	22.4, $N = 31$ (16)	Nonverbal
Parbery-Clark et al., 2012, *Front. Aging Neurosci.*	1	HINT	Sentences/ speech-shaped noise/ semantic	49, onset age 6.5	$<$ 4 years training	56†, $N = 48$ (23)	Nonverbal
Parbery-Clark et al., 2012, *Neuroscience*	1	QuickSIN	Sentences/ babble/ syntactic	17.3, onset age 5.4	2.1 years, onset age 11.4	22, $N = 50$ (23)	Nonverbal
Parbery-Clark et al., 2013, *Journal of Neuroscience*	1	HINT	Sentences/ speech-shaped noise/ semantic	16.2, onset age 5.1	0.7 years, onset age 12.2	20, $N = 30$ (15)	Nonverbal
Ruggles et al., 2014, *PloS ONE*	2	QuickSIN,	Sentences/ babble/ syntactic	At least 10 years, onset age 6.9	$<$ 2 years training, not currently playing instrument	21.2, $N = 33$ (16)	none
		HINT	Sentences/ speech-shaped noise/ semantic				
Slater and Kraus, 2016, *Cognitive Processing*	2	QuickSIN	Sentences/ babble/ syntactic	15.7	$<$ 3 years music experience, no active music making within the last 3 years	23.9, $N = 54$ (37)	Nonverbal
		WIN	Words/ babble/ none				
Swaminathan et al., 2015, *Scientific Reports*	4	Masked speech (forwards, reversed, co-located, separated)	Sentences/ babble/ syntactic	13.8, onset age 8.5	No formal training, not currently playing music	21.7, $N = 24$ (12)	none
Vanden Bosch der Nederlanden et al., 2020, *Psychological Research*	2	SPIN-R high predictability	Sentences/babble/semantic	11.6, onset age 9.1	1.24 years of training, onset age 11.09	21, $N = 60$ (30)	Verbal, Nonverbal, AWM
		SPIN-R low predictability	Sentences/babble/syntactic				
Varnet et al., 2015, *Scientific Reports*	3	Nonwords in noise (correct, sensitivity, SNR)	Nonwords/white noise/none	15.8, onset age $<$13	No musical practice	22.8, $N = 38$ (19)	none
Yoo and Bidelman, 2019, *Hearing Research*	3	QuickSIN	Sentences, Words/ babble/ syntactic	15.8, onset age 7	$\leq$ 3 years training	25.4, $N = 31$ (16)	Nonverbal*, AWM*
		WIN	Words/ babble/ none				
		HINT	Sentences/ speech-shaped noise/ semantic				
Zendel and Alain, 2012, *Psychology and Aging*	2 (younger adults; older adults)	QuickSIN	Sentences/babble/syntactic	At least 6, onset age $<$ 16	$<$ 2 years formal or self-directed lessons, not currently playing instrument	66.3† (older), $N = 83$ (35); 28.0 (younger), $N = 75$ (37)	none
Zendel et al., 2015, *Journal of Cognitive Neuroscience*	1	Words in babble	Words/ babble/ none	15.5, onset age 7.8	$<$ 1 year training, not regularly playing instrument	22.7, $N = 26$ (13)	none
Zendel and Alexander, 2020, *Frontiers in Neuroscience*	1	QuickSIN	Sentences/ babble/ syntactic	23.5, onset age 11.4	Not currently playing an instrument, little to no previous training	32.0, $N = 37$ (19)	none
Zhang et al., 2019, *International Journal of Audiology*	1	QuickSIN	Sentences/ babble/ syntactic	At least 10 years, onset age $<$ 7	$<$ 3 years training, no training in the last 5 years	24, $N = 34$ (17)	none

* musicians significantly outperformed non-musicians on IQ measure (marked as "nonequivalent").
† marked as "older adults" in sub-group analysis.

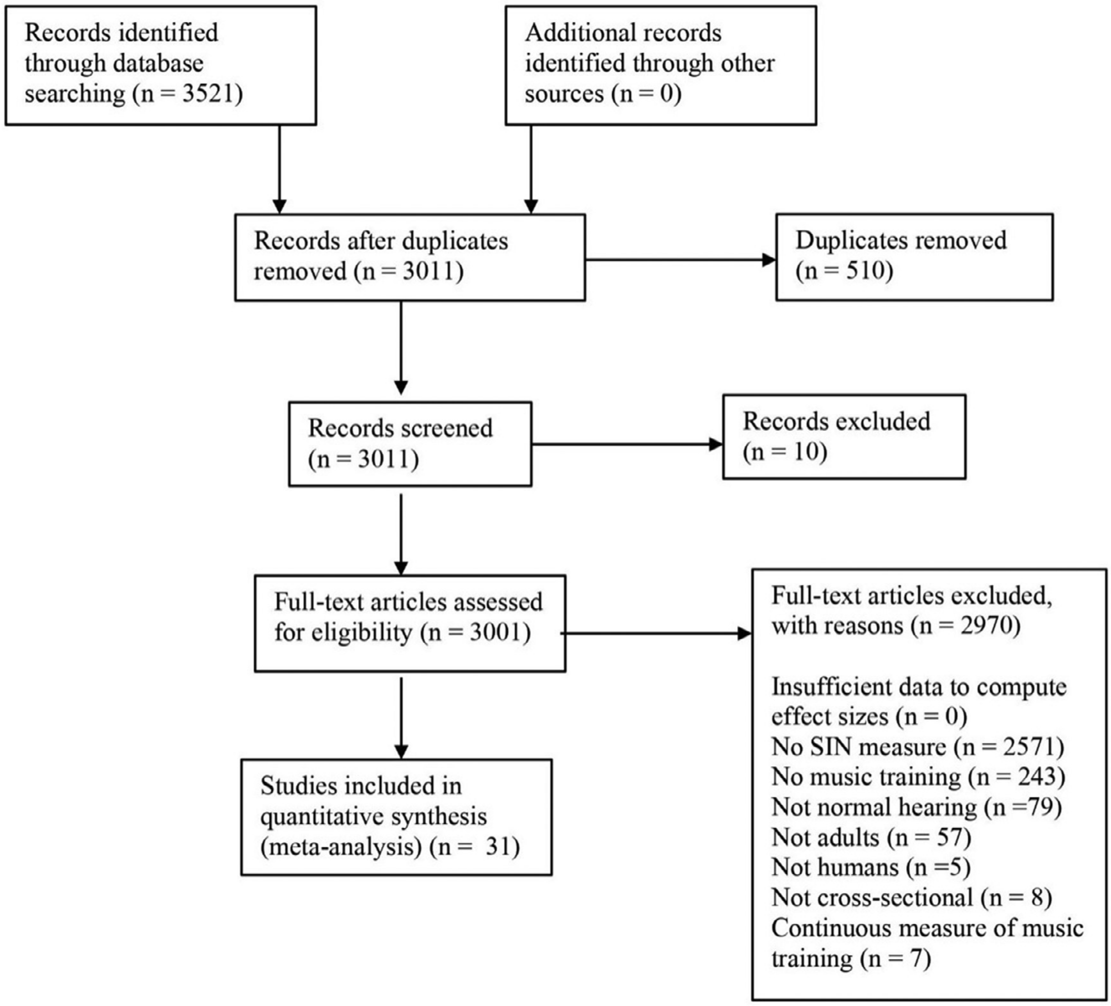

Fig. 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram, indicating records identified through database searching, screened, excluded, and final studies included in meta-analysis.

2.4. Data extraction

Study characteristics and outcome data were extracted manually from each study, first by the first author and then by two independent researchers using a spreadsheet form. During data extraction, researchers were blind to information extracted by other researchers. After finishing extraction, in the case of disagreement between researchers, the first author reviewed the disagreement and paper in question. Disagreements occurred 4 total times and, in all cases, numbers had been extracted incorrectly (e.g., researcher A mistyped a mean, researcher B correctly typed the mean, first author reviewed the disagreement by returning to the study and verifying the correct number). If a study included additional groups, only the data from the formally trained musician and non-musician control groups were extracted. If a study contained insufficient data to calculate effect size or to attain descriptive statistics, the authors were contacted via email up to two times by the first author (7 authors contacted). For each article, the following data points were extracted: publication details, speech-in-noise task description, *n* for each group, years of music training in the musician group, mean participant age, target speech type and context, whether the two groups were evaluated for equivalence on IQ, outcome means and standard deviations or standard errors for each group, and F or t statistics.

2.5. Power analysis

All analyses for this manuscript were conducted using R statistics (R Core, 2021) version 4.0.5 “Shake and Throw”. Packages used for data import, wrangling, and plotting include *readxl* (Wickham and Bryan, 2019), *dplyr* (Wickham et al., 2021), *ggplot2* (Wickham, 2016), *sjplot* (Lüdecke, 2020), and *ggrepel* (Slowikowski, 2021).

We conducted an a priori power analysis using the *dmetar* package (Harrer et al., 2019b) in R. To achieve 80% power, assuming moderate heterogeneity among studies and an alpha level of 0.05, at least 20 studies with an average of 15 participants in each group were necessary to detect a Cohen’s d effect size of 0.30. This medium effect size was chosen given the limited number of papers available in this subject to estimate an a priori effect size estimate. We additionally conducted a power analysis of subgroups using the *dmetar* package (Harrer et al., 2019b) in R, which indicated that an effect size difference of 0.40, assuming standard errors of 0.10 for each effect, was necessary to achieve power of 80% for moderator analyses.

2.6. Effect size calculation

Effect sizes were calculated for each outcome measure using the *esc* package (Lüdecke, 2019) in R (R Core, 2021). Hedge’s *g*, an

Table 2
Description of common outcome variables included in meta-analysis.

Task Name	Description	Outcome Measure
HINT	Participants are asked to repeat sentences (Bamford-Kowal-Bench sentences) that are presented against speech-shaped noise. Sentence level is adaptive to participant performance.	SNR-50: Signal-to-noise-ratio (SNR) required to correctly repeat 50% of target sentences.
QuickSIN	Participants are asked to repeat sentences (Institute of Electrical and Electronic Engineers sentences) that are presented against 4-talker babble at increasing noise level.	SNR loss: SNR required to repeat 50% of sentences (measured by accuracy of 5 key words per sentence)
SPIN-R	Participants are asked to repeat the last word of a sentence presented against 12-talker babble at multiple SNRs. Half of sentences have a high-predictable final word, and half of sentences have a low-predictable final word.	Percent correct: Percent of final words repeated correctly.
LiSN-S	Participants are asked to repeat sentences (Bamford-Kowal-Bench) presented against a distractor voice that is either 0 ° or −90 ° azimuth in relation to target sentence. Target sentence level is adaptive to participant performance.	SNR required to repeat at least 50% of words in target sentence.
WIN	Participants are asked to repeat monosyllabic words presented in 4-talker babble at decreasing SNR.	SNR score: SNR required to repeat 50% of words.

effect size measure that accounts for small study bias (Hedges and Olkin, 1985), was computed using the following formula, where *d* is Cohen's *d*, n_1 is the sample size of group 1 and n_2 is the sample size of group 2:

$$g \simeq d \ \times \ \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

A positive effect size indicated an advantage in the music compared to the control group.

2.7. Three-Level model

Given studies reported more than one SIN outcome, a three-level model (Assink and Wibbelink, 2016; Cheung, 2014; Hox, 2010) was employed using the *metafor* package (Viechtbauer, 2010) using guidelines from M. Harrer et al. (2019). Three-level models have been shown to perform well in meta-analyses involving multiple effect sizes within one study (Cheung, 2014). Here, meta-analysis variances are assessed across three levels: Level 1) sampling variance, Level 2) variance between effect sizes within a single study (i.e., different outcome measures), and Level 3) variance between studies. Model equations, as presented by (M M. Harrer et al., 2019) are as follows, where i is an individual effect size from study *j*. θ_{ij} and $\hat{\theta}_{ij,}$ are the true and estimated effect sizes i from study j, and ϵ_{ij} is the Level 1 error, $\zeta_{(2)ij}$ is the Level 2 error, $\zeta_{(3)j}$ is the Level 3 error, κ_j is the average effect size of study j, and β_0 is the effect size at the population level.

Level 1: $\hat{\theta}_{ij} = \theta_{ij} + \epsilon_{ij}$
Level 2: $\theta_{ij} = \kappa_j + \zeta_{(2)ij}$
Level 3: $\kappa_j = \beta_0 + \zeta_{(3)j}$

The combined equation is, therefore: $\hat{\theta}_{ij} = \beta_0 + \zeta_{(2)ij} + \ \zeta_{(3)j} + \epsilon_{ij}$

Heterogeneity measures (I^2) across model levels were calculated using the *dmetar* package in R (Harrer et al., 2019b)

2.8. Moderators

Five moderators (subgroups) were assessed:

1 *Type of target stimulus* coded the type of speech stimulus participants were asked to identify within noise (i.e., words, sentences, or syllables).
2 *Type of background noise* was coded as the type of noise in which speech targets were embedded (i.e., competing speaker, babble (more than one speaker), speech-shaped noise, fluctuating speech-shaped noise (speech-shaped noise that matched temporal envelope of speech noise), or white noise).
3 *Type of context* was coded as the type of contextual or syntactical cues within which a target stimulus was embedded, as a measure of target predictability. For example, if the target was a meaningful, syntactically correct sentence, it was coded as "semantic", but if the target was a meaningless but syntactically correct sentence, it was coded as "syntactic". Stimuli that contained no contextual cues were coded as "none".
4 *Age group of participants.* Effect sizes were coded as "younger adults" (mean age $<$ 45) or "older adults" (mean age $\geq$ 45).
5 *IQ equivalence.* Studies were coded as to whether the musically-trained group and the musically-untrained group were tested for and determined to be equivalent in Verbal IQ, Non-verbal IQ, or Auditory Working Memory (AWM). Specifically, if IQ was measured and there were no IQ differences between groups, studies were coded as "equivalent". If IQ was measured and there *were* differences between groups, or if IQ was not measured at all, studies were coded as "not equivalent". This information for each study is shown in Table 1.

A separate three-level model was fitted for each moderator variable, as the inclusion of multiple moderators has been shown to increase Type-II error of the moderator estimate (Raudenbush and Bryk, 2002). To correct for multiple comparisons, we used the Bonferroni adjustment, multiplying p values by the number of models assessed (8; target stimulus, background noise, context, age group, any IQ, Verbal IQ, Non-verbal IQ, AWM). If moderators were significant, they were included in the full model.

2.9. Publication bias

Research has shown that studies reporting statistically significant findings are published more often than studies reporting null results (Viechtbauer, 2007). Methods of assessing this publication bias in multi-level meta-analyses show inconsistent performance (Fernández-Castilla et al., 2020). Traditional methods of publication bias analysis include Egger's Test of the Intercept, a method of assessing the asymmetry of a funnel plot, which is a visual representation of individual study effect estimates as a function of standard error. In an unbiased meta-analysis, a funnel plot will form the shape of a roughly symmetrical upside-down funnel centered around a midline at the pooled effect size, with studies at the top of the plot (those with low standard errors) lying close to the pooled effect size, and studies at the bottom (those with increasing standard errors) scattered increasingly away from the pooled effect to both the left and right. If Egger's Test of the Intercept reveals significant asymmetry, Duval and Tweedie's Trim & Fill method may be used to identify outliers (trim), and to add a mirrored effect size to the other side of the funnel (fill), and to recalculate pooled effect size until symmetry is achieved. However, these methods show inflated Type I error rates when dependent effect sizes are ignored (for example, collapsing multiple effect sizes of one study into a single number, or randomly sampling one effect size from each study) (Rodgers and Pustejovsky, 2020). Therefore, we opted for a multi-level method of assessing funnel plot asymmetry, the Egger MLMA test, originally proposed by (Van den Noortgate et al., 2013), which demonstrates sufficient power to detect asymmetry (in which 80% power is obtained with moderate effect size and moderate heterogeneity when the probability of censoring non-significant effects is 0.8) (Rodgers and Pustejovsky, 2020) and does not inflate Type I error (Fernández-Castilla et al., 2020; Rodgers and Pustejovsky, 2020). This test regresses effect size precision on effect size, with slope estimated using weighted least-squares and intercept significance testing using a multi-level model approach (Fernández-Castilla et al., 2020). If the publication-bias effects related slope (effect size precision) is significant, the intercept of the model is interpreted as the adjusted pooled effect size after adjusting for asymmetry. It should be noted that, while the source of funnel plot asymmetry may be selective reporting, or publication bias, asymmetry may also result from a "small-study effect" (Rodgers and Pustejovsky, 2020), referring to the phenomenon that studies with small sample sizes often have large effect sizes (Ioannidis, 2008). While we refer to this analysis as a "publication bias analysis", we acknowledge that the source of asymmetry may derive from either scenario. The Egger MLMA test was performed in R using the *metafor* package and code adapted from Rodgers and Pustejovsky (Rodgers and Pustejovsky, 2020).

2.10. Influence analysis

An influence analysis was conducted to identify and remove leverage effect sizes. Leverage effect sizes are study effect sizes that unduly influence the pooled effect size; for example, a very large effect size drawn from a study that has a large sample size (in comparison to other included studies in the meta-analysis) will have a large influence on the pooled effect size estimate, distorting meta-analytic results. To identify and remove such leverage effect sizes, influence analysis was conducted using the *dmetar* (Harrer et al., 2019b) and *metafor* (Vietchbauer, 2010) packages. We conducted analysis using all effect sizes treated independently using the *meta* package (Balduzzi et al., 2019), as done in previous multi-level meta-analyses (Castillo-Eito et al., 2020; Parry et al., 2021). Effect sizes exerting high influence were detected using a leave-one-out method, as suggested by Viechtbauer and Cheung (2010). In the leave-one-out method, meta-analysis results are calculated K times (where K is the number of effect sizes included), leaving out one effect size on each iteration. This method can then assess if one effect size exerts influence on meta-analytic results such that, when removed, results differ substantially from when it is included. The quantified "influence" value is the standardized difference between the overall effect with the effect size compared to the pooled effect without the effect size. Results of the influence analysis were visualized using a Baujat plot (Baujat et al., 2002), a diagnostic plot used to detect studies that contribute unequally to the meta-analysis by plotting contribution of the study to between-study heterogeneity against the "influence" value, determined by the leave-one-out method. In a Baujat plot, studies falling to the right of the plot exert high heterogeneity contribution, studies in the upper portion of the plot exert high "influence", and studies in the upper right-hand corner are considered leverage points as they have a large impact on both heterogeneity and pooled effect size. Between-study heterogeneity in the Baujat plot is measured by Cochran's *Q*, a metric based on the deviation of each effect from the pooled effect, weighted by the inverse of the study variance.

3. Results

3.1. Study characteristics

Study characteristics, including mean age, outcome measures, and IQ equivalence (whether groups were explicitly tested for and demonstrated to be equivalent on IQ) are presented in Table 1. The mean participant age was 28.5 and the mean participant age, when weighted for the number of effect sizes included, was 28.7. Of the 62 effect sizes included, 11 were from studies whose participants were "older adults" (> age 45), while 51 had participants who were considered "younger adults". 13 effect sizes were from studies that showed equivalence between groups for verbal IQ, 27 showed equivalence between groups for nonverbal IQ, and 12 showed equivalence between groups for auditory working memory. 30 effect sizes were from studies that showed equivalence between groups for any of the 3 IQ measures. Length of music training was specified in 26 of the 31 studies, with an average of 19.2 years of training in the musician group. In the remaining studies, music training length was described as at least 10 years (Baskent and Gaudrain, 2016; Ruggles et al., 2014), at least 6 years (Zendel and Alain, 2012), and practicing at least 7 h per week regularly with 3 h in orchestral rehearsal (Fostick, 2019). Although not explicitly meeting more than 6 years of training criteria, the authors determined to include this study because an individual playing music for *at least 7 h per week regularly* is likely to be considered a "musician". Musician participants across studies were largely recruited from University-level music programs, were mostly classically trained, engaged in both private lessons and ensemble, and played instruments such as piano, violin, cello, oboe, french horn, bassoon, guitar, and voice.

3.2. Three-level model analysis

The first three-level model, without moderators and before assessing influence or publication bias, resulted in a pooled effect size (g) of 0.58 ($p < 0.0001$), with a 95% confidence interval of [0.42, 0.74]. 47.35% of the total model variance was attributed to Level 1 (sampling variance). $I^2_{\text{Level 2}}$ was ~0%, indicating no within-study heterogeneity. $I^2_{\text{Level 3}}$ was 52.65%, indicating moderate between-study heterogeneity. $I^2_{\text{Total,}}$ indicating the amount of heterogeneity not attributable to sampling error, was 52.65%. Effect sizes for each outcome variable and study are presented in Fig. 2.

We then assessed whether our three-level model was superior to a two-level model by removing one of the levels and comparing fit. When removing only Level 2 (within-study heterogeneity), the resulting model had a slightly lower AIC (77.44) and BIC (81.66) than the full model (79.44, 85.77), but this difference was not significantly different ($p = 1.00$), suggesting that including Level 2 was not necessary, and that within-study heterogeneity was not significant. When removing Level 3 (between-study heterogeneity), the resulting model had a higher AIC (89.74) and BIC (93.96) when compared to the full model and was significantly different ($p < 0.001$). This suggests that Level 3 was necessary to include in the full model for this analysis (and that between-study heterogeneity was significant). Given the model comparisons, our final model excluded Level 2. The resulting estimate effect size and level heterogeneity, however, did not differ from the full model, as the removed Level 2 variance was originally at 0%.

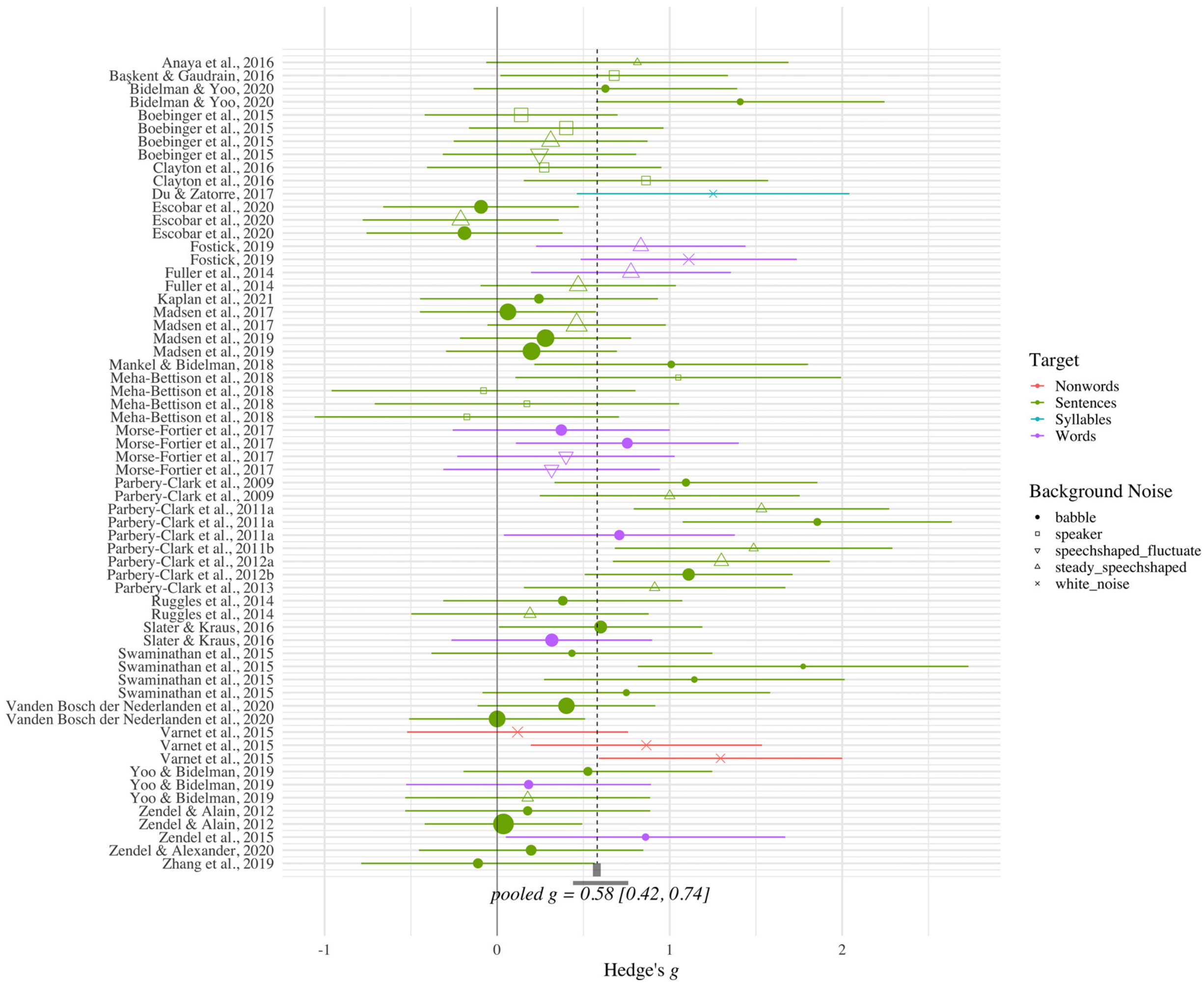

Fig. 2. Forest plot depicting all effect sizes and 95% CIs across studies and measures. The horizontal axis represents effect size (Hedge's g), where a positive effect size indicates a musician advantage. The vertical axis indicates each study, in alphabetical order, included in the meta-analysis, with a separate entry for each effect size included. For each effect size, speech target type is indicated by color and background noise type is indicated by shape. Line length indicates 95% confidence interval surrounding each effect size, and shape size indicates effect size weight (as determined by the inverse of the variance). Pooled effect size is shown as a black dashed line ($g = 0.58$) with a gray bar indicating 95% CI [0.42, 0.74].

3.2.1. Type of speech target

A test of moderation by speech target was conducted to compare effect sizes where the target stimulus was a sentence versus a word. Syllables and nonwords were excluded from this analysis as they did not contain at least 3 effect sizes. The test of moderators indicated that type of speech target was not a significant moderator ($F(1, 56) = 0.13$, $p = 0.72$, Bonferroni-adjusted $p = 1.00$).

3.2.2. Type of background noise

A test of moderation on type of background noise was conducted to compare effect sizes where the background noise was multi-talker babble, a single speaker, speech-shaped noise, fluctuating speech-shaped noise, or white noise. The test of moderators indicated that type of background noise was not a significant moderator ($F(4, 57) = 1.21$, $p = 0.31$, Bonferroni-adjusted $p = 1.00$).

3.2.3. Type of context

A test of moderation on type of contextual cues available to the listener was conducted to compare effect sizes where stimuli were placed within the context of semantic, syntactic, or no cues. The test of moderators indicated that type of context was not a significant moderator ($F(2, 59) = 0.23$, $p = 0.79$, Bonferroni-adjusted $p = 1.00$).

3.2.4. IQ equivalence

We conducted 4 separate models to assess the impact of IQ equivalence on effect size: non-verbal IQ, verbal IQ, auditory working memory, any IQ measure. Methods of assessing IQ equivalence were assessed independently, rather than in one model to account for the fact that IQ assessment methods were not mutually exclusive within a study (a single study could include multiple methods). IQ equivalence did not moderate effect size in any of these tests of moderation (any IQ $p = 0.26$, Bonferroni-adjusted $p = 1.00$; nonverbal IQ $p = 0.05$, Bonferroni-adjusted $p = 0.45$; verbal IQ $p = 0.37$, Bonferroni-adjusted $p = 1.00$, AWM $p = 0.30$, Bonferroni-adjusted $p = 1.00$).

3.2.5. Age group

The test of moderators including age group showed a significant difference between older adults and younger adults ($F(1, 60) = 5.95$, $p < 0.05$), indicating that studies with older adults had overall higher effect sizes than those with younger adults. A follow-up three-level model of only studies with younger adults indicated an overall effect size of ($g = 0.50$, $p < 0.001$, 95% CI: [0.36, 0.65]), whereas studies with older adults indicated an overall effect size of ($g = 1.04$, $p < 0.01$, 95% CI: [0.53, 1.56]). However, the test for age group differences was not significant after correcting for multiple comparisons (Bonferroni-adjusted $p = 0.14$).

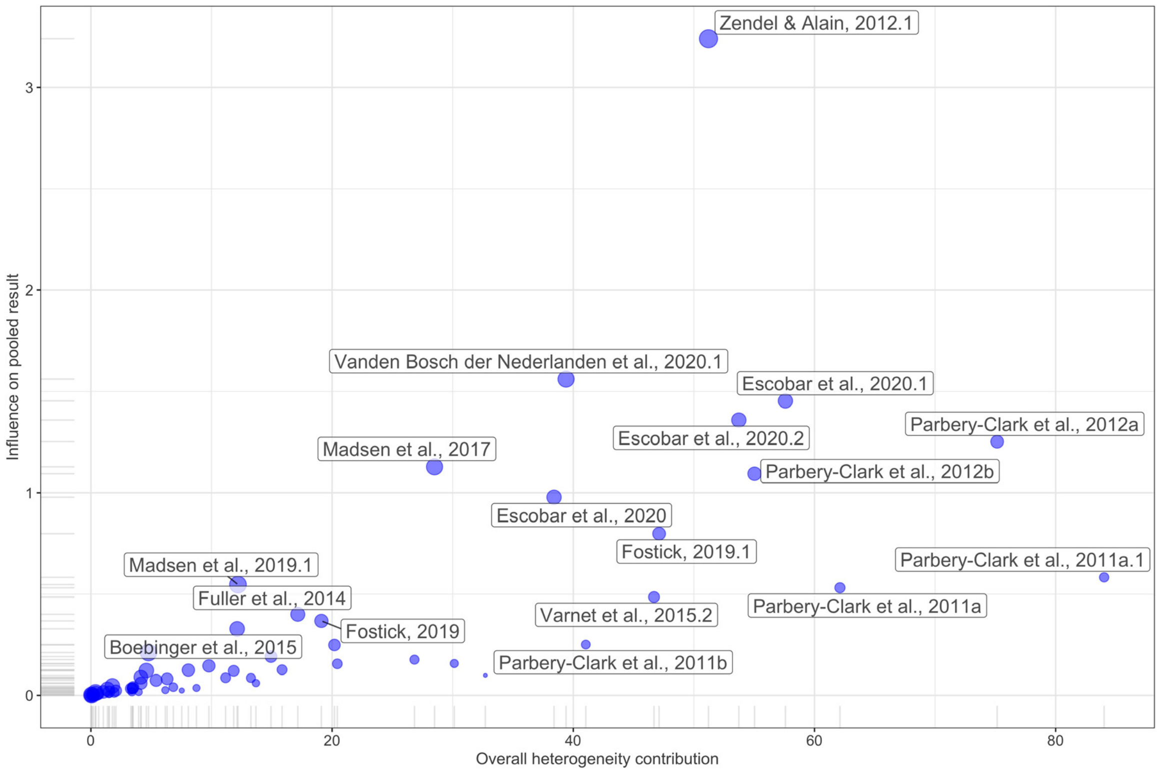

Fig. 3. Baujat plot depicting influence analysis results across all effect sizes. The horizontal axis represents the contribution of each study to overall heterogeneity (Cochran's Q). The vertical axis represents the "influence" value of a given effect size, calculated during leave-on-out analysis as the standardized difference between the overall meta-analysis effect with the effect size compared to without the effect size. Dot size represents inverse of the standard error of the study from which each effect size was taken. Labels indicate the specific study from which the effect size is taken (when shown; not all labels are included to maintain visibility of data points).

3.3. Influence analysis

Influence analysis indicated no leverage effect sizes, indicating no single effect size contributed disproportionately to the overall pooled effect size or heterogeneity (see Fig. 3).

3.4. Publication bias analysis

The Egger MLMA test for asymmetry on the intercept was not significant, indicating no significant funnel plot asymmetry due to selective reporting or small-study effects (β (SE) $= 2.69$ (1.42), $p = 0.063$), (see Fig. 4). This result indicates that meta-analytic results likely reflect a true pooled effect size and are not biased due to selective reporting or over-influence of studies with small sample sizes. We note, however, this effect did approach significance. If using a conservative significance cut-off, taken at the 0.1 level (if the alpha level was set to 0.1 rather than the traditional 0.05), the pooled effect size after adjusting for publication bias or small study effects, would be non-significant ($g = -0.29$, 95% CI [-1.22, 0.64], $p = 0.53$).

4. Discussion

In this multi-level meta-analysis, we investigated speech-in-noise perception abilities in adult musicians and non-musicians across 31 studies and 62 effect sizes. Results indicated a moderate effect size indicating a musician benefit for the ability to perceive speech in noisy environments, with moderate heterogeneity derived from between-study effects and virtually no heterogeneity derived from within-study effects. Age was initially identified as a significant moderator of effect size, suggesting that older adults, with more years of music training, showed greater advantages in speech-in-noise perception when compared to controls with potential age-related decrease in speech-in-noise perception; however, this was not significant after correcting for multiple comparisons among the 8 subgroup analyses conducted and should be interpreted with caution given the small number of studies including older adults. We additionally found that the overall musician benefit for SIN ability was not impacted by whether groups were assessed for and determined to have IQ-equivalence, the type of target stimulus, contextual cues, or the type of background noise. Overall, results from this meta-analysis indicate that a musician benefit for SIN abilities remains robust across studies independent of whether IQ was matched, age, and qualitative aspects of the speech task. While no statistically significant evidence of publication bias or small study effects were observed, these assessments approached significance, indicating a need for additional studies with larger sample sizes and reporting of non-significant effects.

We note that, while the observed pooled effect size is described as "moderate" based on the statistical effect size and naming conventions suggested by Cohen (1962), this may not necessarily translate to clinically relevant differences in speech-in-noise perception between musicians and non-musicians. For example, differences in SNR-50 between musicians and non-musicians in the QuickSIN and HINT tasks among included studies ranged from 0.19 to 1.2 dB, and the roughly estimated pooled mean difference between groups of these two tasks (calculated as the overall pooled effect size multiplied by the available pooled standard deviation for QuickSIN and HINT) was 0.40 dB. This is markedly lower than the just-meaningful difference in SNR reported by McShefferty and colleagues (McShefferty et al., 2016), where a 6 dB improvement was necessary for participants to seek hearing intervention. The estimated pooled mean difference between groups in two SIN tasks (0.40 dB) constitutes approximately half of the smallest detectable

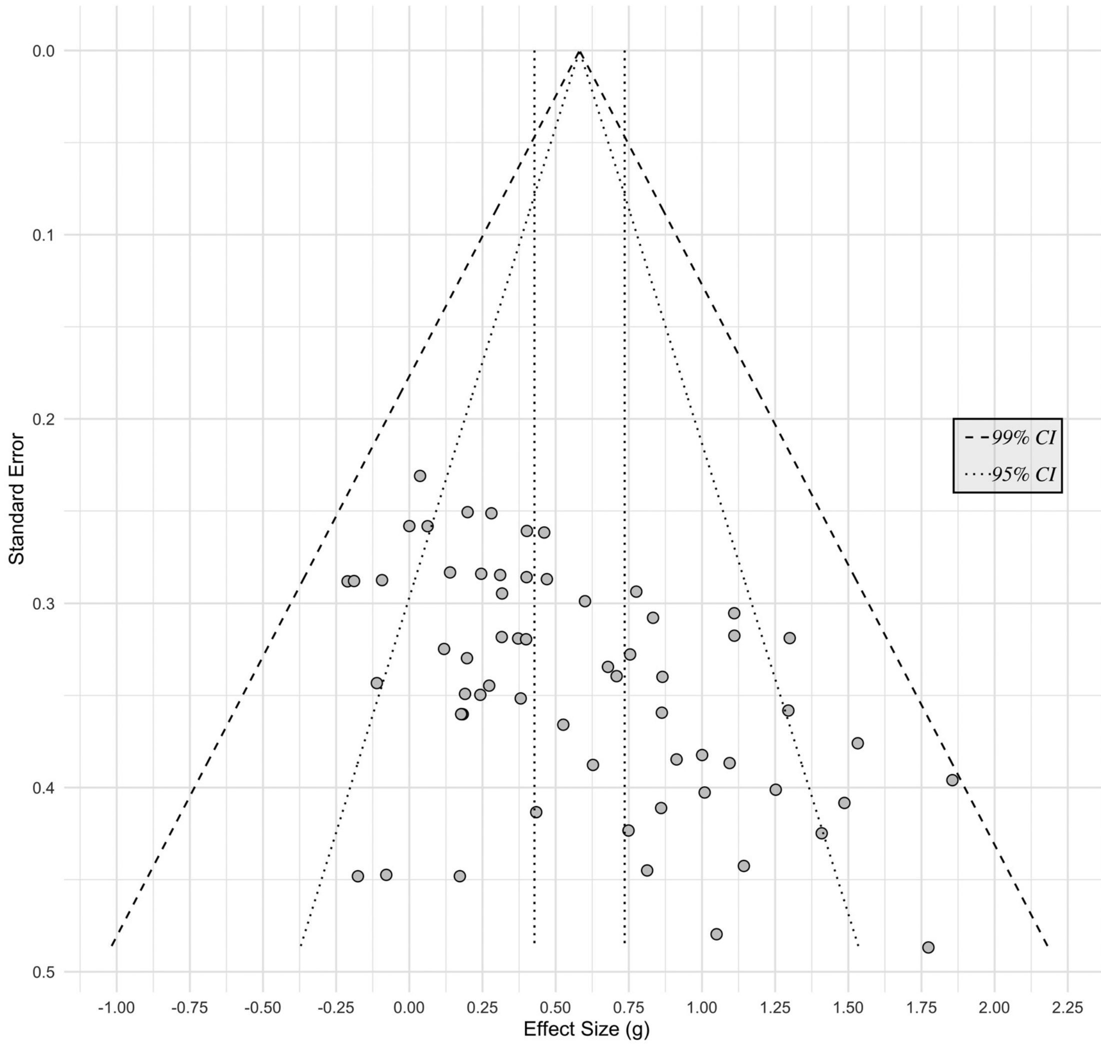

Fig. 4. Funnel plot to assess publication bias. The horizontal axis represents effect size (Hedge's *g*), where a positive effect size indicates a musician advantage. The vertical axis represents standard error of the effect size. Each dot on the plot represents a different effect size. Parallel vertical dotted lines indicate observed pooled effect size 95% CI ([0.42, 0.74]), and dotted and dashed slanted lines indicate 95% and 99% confidence intervals, respectively, for expected effect estimates.

difference (1 dB difference), which may not translate to meaningful change for individuals even without clinical hearing deficits. The present analysis provides evidence for a *statistically*, but not clinically or necessarily detectable, moderate effect of musician status on speech-in-noise perception for two of the most common SIN tasks included in this analysis.

4.1. Moderators of musician advantage for SIN perception

The type of target, context cues, and background noise used to assess SIN perception varied considerably across studies. Yet, despite these differences, neither speech target, background noise type, or contextual cues significantly moderated effect size between or within studies, indicating that the musician advantage for SIN perception is global, rather than specific to the semantic or syntactical richness of a speech cue (corresponding to its predictability) or the complexity of the background. This finding differs from observations made in individual studies included in this analysis; specifically, others have observed a musician advantage for perceiving speech specifically in informational, as opposed to energetic masking settings (e.g.: Morse-Fortier et al., 2017). We did not, however, observe a moderating effect of type of background noise on pooled effect size, indicating that the distinction between informational and energetic masking may not result in robust differences in effects among musicians and non-musicians after aggregating across studies. Yet, simply equating 'informational masking' to any speech-related background noise may be an overly broad generalization, as the amount of informational masking substantially differs between noise contexts of 1, 2, or 5 background speakers, for example. Given the scarcity of studies that aimed to explore informational masking explicitly (and thus, included a greater range of complexity), the present analysis was unable to assess differences between energetic and informational masking at this level. To adequately explore this concept in more detail, more studies are needed that include varying noise contexts, particularly those with solely informational masking.

Additionally, equivalence in cognitive performance did not significantly moderate the overall observed effect across studies. That is, while some studies made an effort to assess equivalence between musician and non-musician participants on verbal IQ, nonverbal IQ, or auditory working memory, whether a study chose to or was able to do so did not impact the overall difference observed between musicians and non-musicians in the meta-analysis. This was true for each cognitive measure independently (verbal, nonverbal, or AWM) and for assessing any of the three measures.

Notably, several (6) studies assessed IQ or auditory working memory and found significant differences between musicians and non-musicians (see Table 1, studies marked with asterisks). Each of these studies conducted follow-up analyses to explore the relationship between cognitive abilities and SIN perception. All six of these studies found significant correlations between scores on the cognitive assessment found to be different between musicians and non-musicians (Anaya et al., 2016: Boston Naming; Bidelman and Yoo., 2020, Clayton et al., 2016, Parbery-Clark et al., 2009, Parbery-Clark et al., 2011, Yoo and Bidelman, 2019: Auditory Working Memory) and at least one measure speech-in-noise perception abilities across participants. This correlation was not consistent across all tasks, however, in some studies that included multiple measures of speech-in-noise perception or multiple measures of IQ. Specifically, while Bidelman and Yoo (2020) found differences between groups in fluid IQ, auditory working memory, masked speech, and QuickSIN, but only auditory working memory was positively associated with performance on the masked speech task across participants. Similarly, Parbery-Clark et al. (2011) observed that auditory working memory was correlated with SIN performance on QuickSIN and HINT, but not the WIN. Lastly, while Yoo and Bidelman (2019) observed differences between musicians and nonmusicians on Raven's Matrices and auditory working memory, only auditory working memory was found to be correlated with speech-in-noise task performance. Additionally, several studies assessed the relationship between musician status and SIN abilities while controlling for cognitive abilities. Anaya et al. (2016) found that, when controlling for vocabulary knowledge in the Boston Naming task, no group difference between musicians and non-musicians was observed for speech-in-noise perception. Both Bidelman and Yoo (2020) and Yoo and Bidelman (2019) observed that, while controlling for auditory working memory, the relationship between musical training and QuickSIN performance remained significant, yet, in Bidelman and Yoo (2020) the relationship between musical training and masked speech performance did not remain after controlling for auditory working memory.

This partially mirrors findings from studies that assessed cognitive ability yet did *not* observe a significant difference between musicians and non-musicians; Boebinger et al. (2015) found that non-verbal IQ, but not musicianship, predicted speech perception thresholds, Escobar et al., 2019 found that those with higher working memory capacities had lower SNR thresholds across tasks, and Vanden Bosch der Nederlanden (2020) observed IQ was marginally correlated with speech-in-noise perception across participants. Yet, others in this category found no correlation between cognitive scores and speech-in-noise perception across participants (auditory working memory: Du and Zatorre, 2017, Slater and Kraus, 2016; collapsed IQ: Madsen et al., 2017; 2019)

Together, these follow-up analyses broadly demonstrate that cognitive ability, specifically auditory working memory, likely plays a role in differences between musicians and non-musicians in at least some tasks. If true, this would be consistent with Dryden et al. (2017)'s finding that cognitive abilities broadly predict SIN perception. While results from our moderator analyses here suggest that musicians' enhanced speech-in-noise abilities may be independent of any potential differences in cognitive ability, given the findings of the six above studies and the fact that not all studies in the present meta-analysis included a measure of cognitive ability, we interpret this finding with caution. It is likely that studies that did not include an IQ assessment that were thus coded as "not equivalent" contained participants that *were* matched on IQ, skewing the findings of our moderator analysis towards a null effect. We encourage future studies on this topic to include and report scores from a cognitive assessment so that future meta-analysts may more accurately capture the full effect of cognitive ability on speech-in-noise perception and may assess whether IQ abilities specifically predict differences between musicians and non-musicians (i.e.: meta-regression methods).

Lastly, we explored whether effect size magnitude was moderated by age group, comparing studies with younger adult participants with those with older adult participants. We initially, before correcting for multiple comparisons, observed that age group significantly impacted the effect of musician status on SIN perception. Specifically, while both subgroups demonstrated musician advantages independently, studies with older adults had an overall higher pooled effect size estimate than studies with younger adults. However, this effect did not remain significant after correcting for multiple comparisons. Additionally, only 4 of the 31 studies in this analysis (6 out of 62 effect sizes) included older adult participants. Given that the subgroups assessed in this study are highly uneven (there were many more studies with younger than older adults), it is difficult to confidently draw even trend-level conclusions taken from this analysis. Additionally, older adult musicians in this meta-analysis had roughly twice the amount of music training than did younger adults. Age differences observed in this analysis likely reflect simply length of training rather than other factors associated with age. Significantly more studies investigating older adult musicians and non-musicians are necessary to draw stronger conclusions regarding musician status, aging, and SIN perception. Additionally, given the age range available, we used 45 as a cut-off for "older adult". More studies with adults on the older end of the age spectrum are necessary.

We note that, based on the subgroup power analysis, the moderator analyses for this study were underpowered, which may have contributed to overall null findings across moderators. However, this subgroup power analysis was designed for a single-level analysis and thus was a more conservative estimate of difference between groups necessary for optimal power. We encourage researchers to include cognitive assessments and varying age groups, speech targets, and noise contexts so that more studies and datapoints may be included in future meta-analyses to reach full power to detect subgroup differences.

4.2. Defining "musician"

An important feature of this meta-analysis is that we included only studies that defined "musicians" and "non-musicians" as a binary variable, based on years of formal training or completion of conservatory degree. This decision was made to maximize comparability between studies, as at least a portion of the literature in this field defines musician status similarly. However, assessing differences between musicians and non-musicians in a binary fashion may exclude valuable information. Including music training as a continuous measure in a meta-regression analysis may provide a more nuanced look into the role of music training on speech-in-noise perception, allowing for a dose-response relationship to be investigated. For example, while Ruggles et al. (2014) did not find a significant overall difference between musicians and non-musicians in SIN abilities, they did observe a significant correlation between SIN perception and years of music training. Onset and duration of music training may also play an important role in speech-in-noise perception and examining this variable on a continuous scale may provide valuable insights on thresholds for a potential musician benefit in these abilities. Most of the included studies required musicians to have begun their studies before age 7 (e.g.: Du and Zatorre (2017); Boebinger et al. (2015)), but many musicians, for example, those who learn their primary instrument in public school curriculum or are self-taught, begin later in adolescence. Early trained musicians have been shown to perform better in rhythm tasks (Bailey and Penhune, 2009) and sensorimotor abilities (Watanabe et al., 2007), as compared to late-trained musicians with similar total years training. Examining age-of-onset as an additional variable in future analyses may shed light on whether such differences exist in relation to speech-in-noise perception among musicians. Additionally, examining music training as a binary variable does not allow for in-depth examination on the diversity of experiences encompassed even within "formal music training". For example, there may be differences in speech-in-noise perception abilities between musicians who were taught primarily by ear, versus those who were taught first to read music, or differences between individuals who play violin as compared to those who sing. Similarly, musicians who play primarily in ensemble settings may have different SIN abilities as compared to musicians who play primarily solo. Research on differences between training methods, settings, and instruments may provide valuable information on the mechanism supporting SIN perception in musicians. While, currently, an insufficient number of studies exist to conduct a meta-regression and additional meta-analysis on these topics, we encourage more incorporation of continuous measures of music training, and comparison between methods and instruments, as a compliment for future research.

Taking this one step further, it should be acknowledged that, while assessing formal training is a convenient way to measure musicianship on a binary or continuous scale, it is not representative of the range of musical experiences present in the population at large. Many individuals have had experience with music making in some informal capacity- for example, through religious services or family gatherings. More recently, the availability of tools for creating and "playing" music electronically (e.g., GarageBand, Logic), and learning instruments informally online (e.g., YouTube tutorials) has increased accessibility of music-making without formal instruction. (Zendel and Alexander, 2020) reported that self-taught musicians outperformed non-musicians and were outperformed by formally trained musicians on a melodic tone violation task. However, no SIN differences were observed between groups. To our knowledge, this is the only study that has included a group of self-taught musicians in assessments of SIN perception. Future studies could incorporate measures of informal music playing experience and assessment of musical abilities (rather than length of training) to further elucidate these findings.

4.3. A limitation of causality

Perhaps the most important caveat of the present meta-analysis is that our findings do not indicate a causal relationship between music training and improved speech-in-noise perception. Rather, we provide evidence for a moderate positive association between musicianship and SIN abilities. All studies included in the analysis were cross-sectional, and therefore it cannot be ruled out that differences between musicians and non-musicians are due to pre-existing biological traits rather than as a result of training. While cognitive abilities are related to accurate speech perception in noisy environments (Anderson et al., 2013; Dryden et al., 2017), we were unable to directly assess the role of cognitive performance on SIN perception. However, we did not find evidence that whether or not a study explicitly identified participants as equivalent in cognitive abilities accounted for SIN difference. It has been proposed that music aptitude, rather than music training, may account for many of the extra-musical benefits, including language abilities, observed in cross-sectional studies comparing musicians with non-musicians (Swaminathan et al., 2017). While most studies in the present meta-analysis did not assess the relationship between music aptitude and SIN perception, several did. Specifically, Slater and Kraus (2016) found that performance on melodic competence rhythm task predicted better performance on speech-in-noise abilities in both musicians and non-musicians, and Meha-Bettison et al. (2018) observed a significant correlation between pitch discrimination abilities and SIN perception for musicians, but not for non-musicians. However, Vanden Bosch der Nederlanden et al., (2020) observed that melody, tempo, rhythm, and tuning performance was not correlated with SPIN-R performance, and Madsen et al. (2019) observed that musical aptitude was not correlated with SIN perception. Including a music aptitude assessment, particularly in studies with continuous measures of music training, may help to separate these effects in the context of cross-sectional investigations. To address the role of training independent of pre-existing differences, several longitudinal studies have been conducted. Developmental work has demonstrated that, after two years of training, musically-trained children show enhanced maturity of early auditory evoked potentials and better ability to detect changes in tonal sequences as compared to sports-trained and children with no training (Habibi et al., 2016). Specific to SIN perception, Slater et al. (2015) found that, in a waitlist control study, children who received music training showed improved SIN abilities as compared to controls. Recent randomized-control trials with older adults show that 10 weeks of choir participation (Dubinsky et al., 2019), and 6 months of piano lessons (Zendel et al., 2019), produced improved performance in speech-in-noise tasks. Our present results are in line with longitudinal findings, suggesting that music training may induce neuroplasticity that supports speech-in-noise perception. To truly separate the effects of pre-existing differences from training, however, meta-analyses of longitudinal studies are needed.

4.4. Additional limitations

Several additional limitations of the present study must be considered. First, while the Egger MLMA analysis of publication bias was not significant at the 0.05 level it did approach significance ($p = 0.06$) and, if we were to conservatively correct for this bias, the pooled effect size would no longer be significant, indicating no musician advantage for speech-in-noise perception. While we used the traditional cut-off for significance at the 0.05 level for this analysis, we cannot ignore the near-significant effects of publication bias or small study effects observed. This trend-level finding indicates that, while aggregating many small studies may result in a clear effect of musician status on speech-in-noise perception, these differences may not reach significance when accounting for inflated effect sizes due to small sampling, or selective reporting. With that, we strongly advocate for publication of significant and null results, and accessibility to post null results on open science platforms to reduce effects of bias in future work. We additionally encourage the inclusion of larger and more representative samples in this field of research to combat small-study effects.

Additionally, we acknowledge that 7 of the 31 included studies (Parbery-Clark et al., 2009, 2011a, 2011b, 2012a, 2012b, 2013; Slater and Kraus, 2016), corresponding to 11 of the 61 effect sizes included in this analysis were conducted by the same research group. This is not surprising given that music science is a new field of study and laboratories with a focus on auditory science often have the expertise and necessary resources to conduct such studies. With the growth of the field however, we expect to see more research drawing from samples from different populations across the world that would provide additional valuable information on speech-in-nose perception in musicians and non-musicians.

4.5. Conclusions and future directions

Speech-in-noise abilities are important for successful communication, and understanding factors involved in enhanced SIN perception across the population provides insight into auditory processing as a whole. This meta-analysis utilized a multi-level design to assess whether musicians demonstrate superior processing of speech-in-noise when compared to non-musicians. A strength of the current investigation is its multi-level design, allowing for the incorporation of multiple effect sizes within a single study. We provide evidence for a moderate musician benefit for speech-in-noise abilities, supported by our overall effect size of g =0.58, 95% CI [0.42, 0.74]. This effect remained robust irrespective of speech target type, background noise, context cues, whether groups were explicitly identified as equivalent in cognitive ability, and age group, indicating that musicians experience advantages to SIN perception across a variety of contexts, cognitive abilities, and throughout the lifespan. We note that this musician benefit, while statistically significant and "moderate" by statistical conventions, may not necessarily reflect clinical or meaningful benefit in auditory perception across individuals.

Based on our analysis, we identify six key areas to address in future research. First, future studies should focus on elucidating a causal link between music training and SIN perception through the employment of randomized control designs. Secondly, more studies are needed that include varying age and hearing groups, including cochlear implant users, individuals with hearing loss, and children exposed to chronic noise. For example, musicians experience less age-related hearing decline, specifically in relation to speech-in-noise abilities (Zhang et al., 2020). While treatments for hearing loss, such as hearing aids, do not focus improving speech-in-noise perception (Killion, 1997), music training may serve as an option for complimentary treatment in those with age-related hearing loss (Dubinsky et al., 2019; Zendel et al., 2019). Third, future studies should consider controlling for socioeconomic status, which contributes to SIN abilities (Anderson et al., 2013), exposure to noise (Casey et al., 2017) (which in turn impacts SIN (Skoe et al., 2019)), and access to music lessons (Elpus and Abril, 2011). Fourth, more research is needed that explores the definition of a "musician", by including and comparing multiple types of training, instruments, onset age, and musical aptitude to assess whether different skill levels and training types may impact speech-in-noise abilities. Fifth, in light of the observed near-significant small-study effect, we strongly encourage researchers to include larger sample sizes in future studies. Finally, we believe focusing on continuous measures of music training years and onset age and experimental studies of music training and SIN abilities is necessary for future meta-analyses.

Data availability statement

The data that support the findings of this study are openly available on Open Science Framework at https://osf.io/w2cm8/.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Sarah Hennessy: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft. **Wendy J. Mack:** Validation, Writing – review & editing. **Assal Habibi:** Conceptualization, Writing – review & editing, Supervision.

Acknowledgements

We thank Alzyeh Hussain, Cassandra Liu, and Nina Tanaka for their assistance in extracting data for this project. We additionally thank Melissa Rodgers for her assistance in assessing publication bias in this analysis. This work was supported by grants UL1TR001855 and UL1TR000130 from the National Center for Advancing Translational Science (NCATS) of the U.S. National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.heares.2022.108442.

References

Anaya, E.M., Pisoni, D.B., Kronenberger, W.G., 2016. Long-term musical experience and auditory and visual perceptual abilities under adverse conditions. J. Acoust. Soc. Am. doi:10.1121/1.4962628.

Anderson, S., White-Schwoch, T., Parbery-Clark, A., Kraus, N., 2013. A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. Hear. Res. 300, 18–32. doi:10.1016/j.heares.2013.03.006.

Assink, M., Wibbelink, C., 2016. Fitting three-level meta-analytic models in R: a step-by-step tutorial. The Quant. Methods Psychol. 12 (3), 154–174. doi:10.20982/tqmp.12.3.p154.

Balduzzi, S., Rücker, G., Schwarzer, G., 2019. How to perform a meta-analysis with R: a practical tutorial. Evid. Based Ment. Health 22, 153–160.

Başkent, D., Gaudrain, E., 2016. Musician advantage for speech-on-speech perception. J. Acoust. Soc. Am. 139 (3), EL51–EL56. doi:10.1121/1.4942628.

Baujat, B., Mahé, C., Pignon, J.P., Hill, C., 2002. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat. Med. 21 (18), 2641–2652. doi:10.1002/sim.1221.

Bidelman, G.M., Yoo, J., 2020. Musicians Show Improved Speech Segregation in Competitive, Multi-Talker Cocktail Party Scenarios. Front Psychol 11. doi:10.3389/fpsyg.2020.01927.

Boebinger, D., Evans, S., Rosen, S., Lima, C.F., Manly, T., Scott, S.K., 2015. Musicians and non-musicians are equally adept at perceiving masked speech. J. Acoust. Soc. Am. 137 (1), 378–387. doi:10.1121/1.4904537.

Casey, J.A., Morello-Frosch, R., Mennitt, D.J., Fristrup, K., Ogburn, E.L., James, P., 2017. Race/ethnicity, socioeconomic status, residential segregation, and spatial variation in noise exposure in the contiguous United States. Environ. Health Perspect. doi:10.1289/EHP898.

Castillo-Eito, L., Armitage, C.J., Norman, P., Day, M.R., Dogru, O.C., Rowe, R., 2020. How can adolescent aggression be reduced? A multi-level meta-analysis. Clinical Psychology Review. NLM (Medline) Vol. 78 doi:10.1016/j.cpr.2020.101853.

Chan, A.S., Ho, Y.C., Cheung, M.C., 1998. Music training improves verbal memory [7]. In: Nature, 396, p. 128. doi:10.1038/24075.

Cheung, M.W.L., 2014. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. Psychol. Methods 19 (2), 211–229. doi:10.1037/a0032968.

Chung, K., 2004. Challenges and recent developments in hearing aids: part i. speech understanding in noise, microphone technologies and noise reduction algorithms. Trends Amplif. 8 (3), 83–124. doi:10.1177/108471380400800302.

Clayton, K.K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A.D., Kidd Jr., G., 2016. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. PLoS ONE 11 (7), e0157638.

Coffey, E.B.J., Mogilever, N.B., Zatorre, R.J., 2017. Speech-in-noise perception in musicians: a review. In: Hearing Research. Elsevier B.V, pp. 49–69. doi:10.1016/j.heares.2017.02.006 Vol. 352.

Cohen, J., 1962. The statistical power of abnormal social psychological research: a review. J. Abnorm. Soc. Psychol. 65, 145–153. doi:10.1037/h0045186.

de Carvalho, N.G., Novelli, C.V.L., Colella-Santos, M.F., 2017. Evaluation of speech in noise abilities in school children. Int. J. Pediatr. Otorhinolaryngol. 99. doi:10.1016/j.ijporl.2017.05.019.

der Nederlanden, C.M.V.B., Rubio-Garcia, A., Clarkson, E., Snyder, J.S., 2020. Change detection in complex auditory scenes is predicted by auditory memory, pitch perception, and years of musical training. Psychol. Res. 84 (3), 585–601.

Dryden, A., Allen, H.A., Henshaw, H., Heinrich, A., 2017. The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic literature review and meta-analysis. Trends Hear. 21. doi:10.1177/2331216517744675.

Du, Y., Zatorre, R.J., 2017. Musical training sharpens and bonds ears and tongue to hear speech better. Proc. Natl. Acad. Sci. U.S.A. 114 (51), 13579–13584. doi:10.1073/pnas.1712223114.

Dubinsky, E., Wood, E.A., Nespoli, G., Russo, F.A., 2019. Short-term choir singing supports speech-in-noise perception and neural pitch strength in older adults with age-related hearing loss. Front. Neurosci. 13, 1153. doi:10.3389/fnins.2019.01153.

Elpus, K., Abril, C.R., 2011. High school music ensemble students in the United States: a demographic profile. J. Res. Music Education doi:10.1177/0022429411405207.

Escobar, J., Mussoi, B.S., Silberer, A.B., 2019. The effect of musical training and working memory in adverse listening situations. Ear Hear. doi:10.1097/AUD.0000000000000754.

Research, Etymotic, 2001. QuickSIN Speech-in-Noise Test (Version 1.3) User Manual. Etymotic Research Inc.

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S.N., .Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: a systematic review. Behav. Res. Methods, 52(5), 2031–2052. https://doi.org/ 10.3758/s13428-020-01373-9

Fostick, L., 2019. Card playing enhances speech perception among aging adults: comparison with aging musicians. Eur. J. Ageing 16 (4), 481–489. doi:10.1007/s10433-019-00512-2.

Fuller, C.D., Galvin, J.J., Maat, B., Free, R.H., Başkent, D., 2014. The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? Front. Neurosci. 8. doi:10.3389/fnins.2014.00179, 8 JUN.

Habibi, A., Cahn, B.R., Damasio, A., Damasio, H., 2016. Neural correlates of accelerated auditory processing in children engaged in music training. Dev. Cogn. Neurosci. doi:10.1016/j.dcn.2016.04.003.

Harrer, M., Cuijpers, P., Furukawa, T., Ebert, D.D., 2019a. Doing Meta-Analysis in R: A Hands-On Guide. PROTECT Lab Erlangen doi:10.5281/zenodo.2551803.

Harrer, Mathias, Cuijpers, P., Furukawa, T., & Ebert, D.D. (2019b). dmetar: Companion R Package For The Guide "Doing Meta-Analysis in R."

Hedges, L., Olkin, I., 1985. Statistical Models for Meta-Analysis. Academic Press.

Hox, J.J., 2010. Multi-level analysis: Techniques and Applications. Routledge.

Ioannidis, J.P., 2008. Why most discovered true associations are inflated. Epidemiology 640–648.

Kaplan, E.C., Wagner, A.E., Toffanin, P., Başkent, D., 2021. Do musicians and non-musicians differ in speech-on-speech processing? Front. Psychol. 12, 281.

Killion, M.C., 1997. Hearing aids: past, present, future: moving toward normal conversations in noise. In: British Journal of Audiology. Whurr Publishers Ltd, pp. 141–148. doi:10.3109/03005364000000016 Vol. 31, Issue 3.

Killion, M.C., Niquette, P.A., Gudmundsen, G.I., Revit, L.J., Banerjee, S., 2004. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. doi:10.1121/1.1784440.

Lüdecke, D. (2019). esc: effect size computation for meta analysis (Version 0.5.1). https://doi.org/ 10.5281/zenodo.1249218

Lüdecke D. (2020). sjPlot: data visualization for statistics in social science. R package version 2.8.5, URL: https://CRAN.R-project.org/package=sjPlot

Madsen, S.M.K., Marschall, M., Dau, T., Oxenham, A.J., 2019. Speech perception is similar for musicians and non-musicians across a wide range of conditions. Sci. Rep. doi:10.1038/s41598-019-46728-1.

Madsen, S.M.K., Whiteford, K.L., Oxenham, A.J., 2017. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. Sci. Rep. doi:10.1038/s41598-017-12937-9.

Mankel, K., Bidelman, G.M., 2018. Inherent auditory skills rather than formal music training shape the neural encoding of speech. Proceedings of the Natl. Acad. Sci. 115 (51), 13129–13134.

Meha-Bettison, K., Sharma, M., Ibrahim, R.K., Vasuki, Mandikal, P, R, 2018. Enhanced speech perception in noise and cortical auditory evoked potentials in professional musicians. Int. J. Audiol. 57 (1), 40–52.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E.G., Cepeda, N.J., Chau, T., 2011. Short-term music training enhances verbal intelligence and executive function. Psychol. Sci. 22 (11), 1425–1433. doi:10.1177/0956797611416999.

Morse-Fortier, C., Parrish, M.M., Baran, J.A., Freyman, R.L., 2017. The effects of musical training on speech detection in the presence of informational and energetic masking. Trends Hear. 21, 2331216517739427.

Neff, D.L., Green, D.M., 1987. Masking produced by spectral uncertainty with multicomponent maskers. Percept. Psychophys. doi:10.3758/BF03203033.

Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. 95 (2), 1085–1099.

McShefferty, D., Whitmer, W.M., Akeroyd, M.A., 2016. The just-meaningful difference in speech-to-noise ratio. Trends Hear. 20, 2331216515626570.

Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009. Musician enhancement for speech-In-noise. Ear Hear. 30 (6), 653–661. doi:10.1097/AUD.0b013e3181b412e9.

Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N., 2012a. Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. Front. Aging Neurosci. 4, 30.

Parbery-Clark, A., Tierney, A., Strait, D.L., Kraus, N., 2012b. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience 219, 111–119.

Parbery-Clark, A., Strait, D.L., Hittner, E., Kraus, N., 2013. Musical training enhances neural processing of binaural sounds. J. Neurosci. doi:10.1523/JNEUROSCI.5700-12.2013.

Parbery-Clark, A., Strait, D.L., Anderson, S., Hittner, E., Kraus, N., 2011a. Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. PLoS ONE 6 (5), e18082.

Parbery-Clark, A., Strait, D.L., Kraus, N., 2011b. Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. Neuropsychologia 49 (12), 3338–3345.

Parry, D. A., Davidson, B. I., Sewall, C. J., Fisher, J. T., Mieczkowski, H., Quintana, D. S., 2021. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nature Human Behaviour 1–13.Chicago.

Patel, A.D., 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Front. Psychol. 2. doi:10.3389/fpsyg.2011.00142, JUN.

Patel, A.D., 2014. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. Hear. Res. 308, 98–108. doi:10.1016/j.heares.2013.08.011.

Pronk, M., Deeg, D.J.H., Festen, J.M., Twisk, J.W., Smits, C., Comijs, H.C., Kramer, S.E., 2013a. Decline in older persons' ability to recognize speech in noise: the influence of demographic, health-related, environmental, and cognitive factors. Ear Hear. doi:10.1097/AUD.0b013e3182994eee.

Pronk, M., Deeg, D.J.H., Festen, J.M., Twisk, J.W., Smits, C., Comijs, H.C., Kramer, S.E., 2013b. Decline in older persons' ability to recognize speech in noise. Ear & Hearing doi:10.1097/aud.0b013e3182994eee.

Ruggles, D.R., Freyman, R.L., Oxenham, A.J., 2014. Influence of musical training on understanding voiced and whispered speech in noise. PLoS ONE 9 (1), e86980.

R Core, T., 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing https://www.r-project.org/.

Raudenbush, S., Bryk, A., 2002. Hierarchical Linear models: Applications and Data Analysis Methods, Vol. 1. SAGE.

Rodgers, M.A., Pustejovsky, J.E., 2020. Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. Psychol. Methods doi:10.1037/met0000300.

Savitz, D.A., Wellenius, G.A., Trikalinos, T.A., 2019. The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: assessing the impact of specific sources of potential bias. Am. J. Epidemiol. 188 (9), 1581–1585.

Schellenberg, E.G., 2011. Music lessons, emotional intelligence, and IQ. Music Percept. doi:10.1525/mp.2011.29.2.185.

Skoe, E., Camera, S., Tufts, J., 2019. Noise exposure may diminish the musician advantage for perceiving speech in noise. Ear Hear. 40 (4), 782–793. doi:10.1097/AUD.0000000000000665.

Slater, J., Kraus, N., 2016. The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. Cogn. Process doi:10.1007/s10339-015-0740-7.

Slater, J., Skoe, E., Strait, D.L., O'Connell, S., Thompson, E., Kraus, N, 2015. Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. Behav. Brain Res. 291, 244–252. doi:10.1016/j.bbr.2015.05.026.

Stam, M., Smit, J.H., Twisk, J.W.R., Lemke, U., Smits, C., Festen, J.M., Kramer, S.E., 2016. Change in psychosocial health status over 5 years in relation to adults' hearing ability in noise. Ear & Hearing 37 (6), 680–689. doi:10.1097/AUD.0000000000000332.

Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V., Kidd Jr,, G., Patel, A.D., 2015. Musical training, individual differences and the cocktail party problem. Sci. Rep. 5 (1), 1–11.

Swaminathan, S., Schellenberg, E.G., Khalil, S., 2017. Revisiting the association between music lessons and intelligence: training effects or music aptitude? Intelligence doi:10.1016/j.intell.2017.03.005.

Talamini, F., Carretti, B., Grassi, M., 2016. The working memory of musicians and non-musicians. Music Percept. doi:10.1525/MP.2016.34.2.183.

Van den Noortgate, W., López-López, J.A., Marín-Martínez, F., Sánchez-Meca, J., 2013. Three-level meta-analysis of dependent effect sizes. Behav. Res. Methods 45 (2), 576–594. doi:10.3758/s13428-012-0261-6.

Varnet, L., Wang, T., Peter, C., Meunier, F., Hoen, M., 2015. How musical expertise shapes speech perception: evidence from auditory classification images. Sci. Rep. 5 (1), 1–13.

Viechtbauer, W., 2007. Publication bias in meta-analysis: prevention, assessment and adjustments. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments doi:10.1002/0470870168.

Viechtbauer, Wolfgang, Cheung, M.W.-L., 2010. Outlier and influence diagnostics for meta-analysis. Res. Synth. Methods doi:10.1002/jrsm.11.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3), 1–48. URL https://www.jstatsoft.org/v36/i03/.

Watanabe, D., Savion-Lemieux, T., Penhune, V.B., 2007. The effect of early musical training on adult motor performance: evidence for a sensitive period in motor learning. Exp. Brain Res. 176 (2), 332–340.

Wickham, H., François, R., Henry, L., and Müller, K. (2021). dplyr: a grammar of data manipulation. R package version 1.0.6. https://CRAN.R-project.org/package=dplyr

Wickham, H., and Bryan, J. (2019). readxl: read excel files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl

Wickham, H., 2016. ggplot2: Elegant Graphics For Data Analysis. Springer-Verlag, New York 2016.

Slowikowski, K. (2021). ggrepel: automatically position non-overlapping text labels with 'ggplot2'. R package version 0.9.1. https://CRAN.R-project.org/package=ggrepel

Sterne, J.A., Hernán, M.A., Reeves, B.C., Savović, J., Berkman, N.D., Viswanathan, M., … Higgins, J.P., 2016. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. bmj 355.

Wilson, R.H., 2003. Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. J. Am. Acad. Audiol. doi:10.1055/s-0040-1715938.

Wilson, R.H., McArdle, R.A., Smith, S.L., 2007. An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. J. Speech, Language, and Hearing Res. doi:10.1044/1092-4388(2007/059).

Yoo, J., Bidelman, G.M., 2019. Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. Hear. Res. 377, 189–195.

Zendel, B.R., Alain, C., 2012. Musicians experience less age-related decline in central auditory processing. Psychol. Aging 27 (2), 410–417. doi:10.1037/a0024816.
Zendel, B.R., Tremblay, C.D., Belleville, S., Peretz, I., 2015. The impact of musicianship on the cortical mechanisms related to separating speech from background noise. J. Cogn. Neurosci. 27 (5), 1044–1059.
Zendel, B.R., Alexander, E.J., 2020. Autodidacticism and music: do self-taught musicians exhibit the same auditory processing advantages as formally trained musicians? Front. Neurosci. 14, 752. doi:10.3389/fnins.2020.00752.
Zendel, B.R., West, G.L., Belleville, S., Peretz, I., 2019. Musical training improves the ability to understand speech-in-noise in older adults. Neurobiol. Aging 81, 102–115. doi:10.1016/j.neurobiolaging.2019.05.015.
Zhang, F., Roland, C., Rasul, D., Cahn, S., Liang, C., Valencia, G., 2019. Comparing musicians and non-musicians in signal-in-noise perception. Int. J. Audiol. 58 (11), 717–723.