



Investigating the Effect of One Year of Learning to Play a Musical Instrument on Speech-in-Noise Perception and Phonological Short-Term Memory in 5-to-7-Year-Old Children

Douglas MacCutcheon^{1,2*}, Christian Füllgrabe³, Renata Eccles^{2,4}, Jeannie van der Linde^{2,4}, Clorinda Panebianco² and Robert Ljung¹

¹ Department of Building, Energy and Environmental Engineering, Högskolan i Gävle, Gävle, Sweden, ² Department of Music, University of Pretoria, Pretoria, South Africa, ³ School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom, ⁴ Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa

OPEN ACCESS

Edited by:

Mary Rudner,
Linköping University, Sweden

Reviewed by:

Lina Motlagh Zadeh,
Cincinnati Children's Hospital Medical
Center, United States
Stefanie Andrea Hutka,
University of Toronto, Canada

*Correspondence:

Douglas MacCutcheon
Douglas.MacCutcheon@hig.se

Specialty section:

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

Received: 22 June 2019

Accepted: 03 December 2019

Published: 10 January 2020

Citation:

MacCutcheon D, Füllgrabe C,
Eccles R, van der Linde J,
Panebianco C and Ljung R (2020)
Investigating the Effect of One Year
of Learning to Play a Musical
Instrument on Speech-in-Noise
Perception and Phonological
Short-Term Memory in
5-to-7-Year-Old Children.
Front. Psychol. 10:2865.
doi: 10.3389/fpsyg.2019.02865

The benefits in speech-in-noise perception, language and cognition brought about by extensive musical training in adults and children have been demonstrated in a number of cross-sectional studies. Therefore, this study aimed to investigate whether one year of school-delivered musical training, consisting of individual and group instrumental classes, was capable of producing advantages for speech-in-noise perception and phonological short-term memory in children tested in a simulated classroom environment. Forty-one children aged 5–7 years at the first measurement point participated in the study and either went to a music-focused or a sport-focused private school with an otherwise equivalent school curriculum. The children's ability to detect number and color words in noise was measured under a number of conditions including different masker types (speech-shaped noise, single-talker background) and under varying spatial combinations of target and masker (spatially collocated, spatially separated). Additionally, a cognitive factor essential to speech perception, namely phonological short-term memory, was assessed. Findings were unable to confirm that musical training of the frequency and duration administered was associated with a musicians' advantage for either speech in noise, under any of the masker or spatial conditions tested, or phonological short-term memory.

Keywords: speech in noise, phonological short-term memory, musical training, children, cognition

INTRODUCTION

Children receive their education in acoustic environments in which background noise is nearly always present. Classroom noise is known to cause distraction and annoyance in children, but its primary effect is a reduction in speech intelligibility (for reviews, see Shield and Dockrell, 2003; Klatte et al., 2013), with a consequently negative impact on academic achievement (Shield and Dockrell, 2008). In typically developing children, the ability to cope with speech in noise (SiN) has been linked to individual differences in cognitive and language

abilities (Nelson et al., 2005; Strait et al., 2012; MacCutcheon et al., 2019), age (Corbin et al., 2016), gender (Prodi et al., 2019), and supra-threshold auditory processing abilities (Lorenzi et al., 2000), as well as environmental factors, including reverberation and the spatial, spectral and temporal characteristics of the background noise (MacCutcheon et al., 2018, 2019; McCreery et al., 2019).

Many studies have focused on how manipulating the acoustic environment can improve children's attention to verbal instructions, self-rated ability to cope with noise, speech reception thresholds (SRTs) and cognitive performance (DiSarno et al., 2002; Purdy et al., 2009; Dockrell and Shield, 2012; Prodi et al., 2019). Contrastingly, the aim of the present study is to investigate whether musical training can improve individual characteristics of the listener that contribute to speech perception (e.g., auditory, linguistic and cognitive abilities) and thereby mitigate speech-intelligibility challenges posed by noise.

Musical training has been suggested as a possible candidate for improving auditory, linguistic and cognitive abilities (Patel, 2011; Tallal, 2014) because a multitude of studies indicate that adults and children with musical training show greater motor, cognitive, linguistic and auditory skills (for a review, see Benz et al., 2016), referred to as the "musicians' advantage" (Başkent and Gaudrain, 2016; Talamini et al., 2017). Indeed, a musicians' advantage for SiN perception has been reported by a number of studies in adults and children (Parbery-Clark et al., 2009; Strait et al., 2012, 2013; Bidelman et al., 2014; Kraus et al., 2014; Slater et al., 2015; Başkent and Gaudrain, 2016). However, there are also a substantial number of studies that failed to find strong evidence in favor of advantages in musicians (Strait et al., 2012; Fuller et al., 2014; Ruggles et al., 2014; Boebinger et al., 2015; Fleming et al., 2019; Zendel et al., 2019).

Despite diverging findings, there is a compelling theoretical basis for the possibility that musical training could improve speech perception. Indeed, due to the similarity of the acoustic features of music and speech, these stimuli are processed by the same brain networks (Patel, 2011). For example, both music and speech perception require the processing of fluctuations in the amplitude envelope of the acoustic signal (Patel, 2011) to discriminate musical notes and phrases and segments of syllables and words, respectively. Additionally, pitch processing (the ability to perceptually discriminate between frequencies) is both an essential aspect of the emotional and linguistic content of speech as well as the harmonic and melodic content of music.

How and why abilities developed through musical training might lead to improvements in SiN processing is currently still unknown. In this study, we consider three possibilities. The first is that musical training confers benefits for dealing with energetic and/or informational maskers; the second is that musical training improves spatial listening; and the third is that musical training confers benefits for SiN perception by improving mediating cognitive processes.

Noise presents a challenge for speech perception as a consequence of the acoustic and spatial characteristics of

the masker. Energetic maskers reduce speech intelligibility, while informational maskers reduce speech perception due to acoustic similarity with the target speech, resulting in perceptual confusion (Brungart, 2001; Wightman and Kistler, 2005; Wightman et al., 2006; MacCutcheon et al., 2019), and informational interference (Dole et al., 2012; Stone et al., 2012). Meanwhile, localization cues provided by the spatial separation of the target speech from the masker can improve intelligibility because timing and level differences between the two ears assist with sound segregation (Litovsky, 2005; Johnstone and Litovsky, 2006); referred to as "spatial release from masking" (Freyman et al., 1999; Hawley et al., 2004). However, assessments of the potential for musical training to help speech perception under these acoustic and spatial conditions have produced mixed results (Parbery-Clark et al., 2009; Strait et al., 2012; Swaminathan et al., 2015) and there is a dearth of longitudinal studies in children in the literature.

The development of SiN perception occurs in conjunction with cognitive development (Hall et al., 2002; Bradley and Sato, 2008; Neuman et al., 2010). According to the Ease of Language Understanding model (Rönnberg et al., 2008), noise places demands on cognitive processing of speech as working memory resources are required for assisting with the matching of incoming phonological information with phonological representations stored in long term memory. Meanwhile, explicit processing resources are also used for making guesses (informed by prior knowledge and experience as well as contextual factors) that might provide clues as to the nature of the missing input. This turns a relatively automatic task into a cognitively demanding, effortful task. Both cross-sectional and longitudinal studies have shown musical-training-induced improvements in cognitive functioning in adults and children (Benz et al., 2016). In particular, phonological short-term memory processes essential for SiN perception seem to be higher in child and adult musicians than in non-musician controls (Chan et al., 1998; Lee et al., 2007; Franklin et al., 2008; Strait et al., 2012, 2013; Bergman Nutley et al., 2014; Roden et al., 2014).

The present study builds longitudinally on a previous cross-sectional study by MacCutcheon et al. (2019). The study investigated whether individual differences in linguistic and cognitive abilities contribute to SiN perception in a variety of listening conditions, composed of different masker types and spatial configurations of the target speech and masker. Participants were typically developing children in early stages of development that are critical to the co-development of language (Rhyner, 2009) and speech perception (Johnstone and Litovsky, 2006). The results of MacCutcheon et al. (2019) indicated that, under certain listening conditions, memory span and expressive language provided benefits for SiN perception. The present study adds to these findings by longitudinally assessing the effect of 1 year of musical training on SiN perception and phonological short-term memory. Children attended one of two schools with equivalent academic curriculums, except that one school offered additional music lessons as part of the school curriculum while the

other school offered additional sports activities. Based on the published literature, it was hypothesized that musical training minimizes the effect of energetic and/or informational masking on speech perception and maximizes the use of spatial cues, resulting in improved speech perception relative to the control group. An additional hypothesis was that musical training improves speech perception via improvements in phonological short-term memory.

Previous studies reporting evidence for a musicians' advantage provided a higher frequency and longer duration of musical training for their participants than the present study. For example, Kraus et al. (2014)'s and Slater et al. (2015)'s children received up to 4 h of musical training per week for up to 2 years before a musicians' advantage was discernible. Although lesson frequencies and lengths for beginners learning an instrument are by no means standardized, norms suggest that children who show an interest in music will initially receive a lesson in their primary instrument once per week. Beginner instrumental lesson times for young children are generally 30–60 min depending on the child's innate musical abilities and attentional capacity as well as practicalities such as parental preferences and resources. As this range is more representative of what the majority of children engaging in musical activities at that age receive under "normal" circumstances, the present study hoped to ascertain a musicians' advantage within a shorter timeframe and with a lower intensity of musical training than previous studies.

MATERIALS AND METHODS

Participants

A total of 41 typically developing male school children participated in the study. On average, they were aged 6.3 years (standard deviation = 0.5 years, range: 5–7 years) at the start of the study, and had no history of cognitive, sensory or behavioral deficits, according to parental report. Parents of children in the participating schools received an information letter through the schoolteacher and agreed for their children to participate by providing written consent. Ethical approval for the study was granted by the University of Pretoria Research Ethics Committee, Approval 25071999 (GW20171130HS).

Prior to participation, all children were screened for hearing deficits. Normal hearing function was established using the smartphone hearing-screening application hearScreen™ that detects hearing losses in excess of 20 dB Hearing Level at 1, 2, and 4 kHz with 97.8% reliability compared to standard manual audiometric procedures (Swanepoel et al., 2014). The application was run on Samsung Galaxy J2 mobile phones connected to Sennheiser HD280 Pro headphones.

Musical Training and Control Groups

Twenty-six participants attended a music-focused school (the musical-training group) where they received up to 1 h per week of instrumental training over the course of a 38-week school year. The training was delivered by a qualified music teacher

who used a combination of Kodaly and Orff methodologies.¹ All children attended a 30-min group recorder lesson, and twelve (29%) children received a further 30-min individual piano or violin lesson. The remaining fifteen participants attended a sports-focused school (the control group) where they participated in extra-curricular sports (e.g., football, cricket, hockey and swimming) for 2–5 h per week. Both schools otherwise followed an equivalent Independent Examinations Board academic curriculum. As part of this curriculum, all children attended a weekly 30-min general group music lesson that did not involve instrumental training. None of the participants received additional musical training outside school.

The musical-training and control groups did not differ in age [$t(39) = 1.38, p = 0.177$, two-tailed], and socio-economic status as measured by maternal education level [$t(39) = 0.39, p = 0.695$, two-tailed]. Both groups were tested on the SiN and FDS tasks twice: once at the first assessment point (T1) when none of the participants had received any formal musical training, and then again at the second assessment point (T2) after attending their respective schools for 1 year. Between-group differences in language ability were also measured using the Renfrew Action Picture Test (RAPT; Renfrew, 1980). This test consists of 10 pictures that must be verbally described (e.g., a girl hugging a teddy-bear), and the information and grammar content of the responses are scored out of 40 and 35 points, respectively. No group differences in language ability were detected at T1 [$t(39) = -0.10, p = 0.922$, two-tailed].

Design

A 2 Groups (musical training vs. control) \times 2 Assessment points (T1 vs. T2) \times 2 Masker types [speech-shaped noise (SSN) vs. single talker] \times 2 Spatial locations (collocated vs. spatially separated) mixed design was used. Speech-in-noise intelligibility was analyzed separately for each group at the two assessment points in each of the four listening conditions obtained by combining masker type and spatial location, as well as averaged across listening conditions.

Tasks

Speech-in-Noise Perception

The SiN test was run on a DELL Latitude E6430 laptop, and the auditory stimuli were presented to the participants through a Focusrite Scarlett 2i2 audio interface and Sennheiser HD 650 headphones. All stimuli were pre-recorded and acoustics were simulated in a virtual classroom with a mean mid-frequency reverberation time T_{30} of 0.6 s using the software Room Acoustics for Virtual Environments (RAVEN; Schröder, 2011). Binaural room impulse responses were simulated based on a head-related transfer function measured from a child dummy head so that the

¹The musical training taught the following musical concepts: pitch (identify and produce high and low pitches, identify and produce pitch contours), duration (identify and produce long and short sounds), beat (keeping steady beat to music through movement and instrumental play), timbre (identify sounds through aural cues, identify instrument families), dynamics (getting louder and softer), form structure (introducing common form structures including AB, ABA, and Rondo form), rhythm (producing crotchet and quaver rhythmic patterns, creating own rhythmic patterns) and creativity (creating a "sound story").

virtually simulated environment was appropriate for the sample under investigation (Fels et al., 2004). Further details about the masker and the simulation of the virtual acoustic environment are reported in MacCutcheon et al. (2019). Speech identification was assessed using an adaptation of the “Children’s Coordinate Response Measure” software described in Vickers et al. (2016). The task was to identify two target words in the carrier sentence “show the dog where the [number word] [color word] is,” spoken by an adult male with an English accent. The color word was one of six colors (black, red, green, white, blue or pink) and the number word was a number between one and nine, with the exception of the disyllabic number seven. The location of the target talker was simulated to be at 0° azimuth. The target speech was accompanied by either a single male adult talker reading fictitious news items, or SSN with the same long-term average speech spectrum as the masking talker. The masker started and ended with the target sentence. Within the simulated virtual environment, each masker was either collocated with the target talker, or spatially separated to the right of the target talker, at +90° azimuth. SRTs for identifying the two target words correctly 50% of the time were assessed. The presentation level of the masker was fixed at 55 dB(A) while the presentation level of the target speech, initially set to 68 dB(A), was adaptively varied, using a 1-up, 1-down procedure (Levitt, 1971). Until the first incorrect response, the presentation level for the target speech was decreased by 8 dB. Then, a step size of 4 dB was used until the second incorrect response occurred. Thereafter, the step remained fixed to 2 dB. Each threshold run was composed of 48 sentences, corresponding to all possible color-number combinations. The SRT was computed as the mean of the final four reversals for a given threshold run.

Phonological Short-Term Memory Capacity

The “Number Repetition – Forward” subtest from the Clinical Evaluation of Language Fundamentals (CELF-4; Semel et al., 2003) was used to assess phonological short-term memory capacity. This version of a forward digit span (FDS) test required the participant to recall number sequences of varying length

(from two to nine digits) in serial order. Initially, the sequence was composed of two digits and the sequence length was increased by one digit after two sequences of the same length were presented. The test was terminated once the participant incorrectly recalled two sequences of the same sequence length in a row, or completed all the lists. Each correctly recalled sequence was awarded a point, resulting in a maximum score of 16 points. Raw scores were converted to age-normed standard scores provided in the CELF-4 manual and all further analyses were conducted using standard scores.

Experimental Procedure

Testing was conducted in a sound-isolated music room of one of the participating schools in the presence of an experimenter. For the SiN test, the graphical user interface showed a photograph of a dog beside six colored panels, each subdivided into nine numbered buttons representing all possible number and color combinations. Given their young age, participants were asked to repeat verbally the number and color they had heard, and the experimenter entered the responses for them by clicking the appropriate buttons on the user interface. The order of the four listening conditions was counterbalanced using a Latin square design. The FDS test was administered according to the protocol provided in the manual of the CELF-4.

RESULTS

Results for the two groups on the short-term memory task and the speech-perception task in the four different listening conditions and on average are given in **Table 1** for the first and second assessment point.

Baseline Performance

At the start of the study (i.e., at T1), the two groups did not differ significantly in SRTs averaged across the four listening conditions [$t(39) = 0.017$, $p = 0.987$, two-tailed]. However, there

TABLE 1 | Group summary statistics in terms of mean, standard deviation (SD), and the lower and upper range of the 95% confidence interval (CI 95%) for performance on the forward digit span (FDS) test and speech-in-noise perception (SiN) test in each listening condition and on average.

Test	Assessment point	Control group			Musical training group		
		Mean (dB)	SD	CI 95% [lower, upper]	Mean (dB)	SD	CI 95% [lower, upper]
FDS	T1	6.4	2.1	[5.2, 7.6]	7.9	2.0	[7.0, 8.7]
	T2	7.1	1.9	[6.1, 8.2]	8.2	1.5	[7.6, 8.8]
SiN: Collocated * SSN	T1	-3.1	3.9	[-5.3, -0.9]	-3.1	4.7	[-5.0, -1.2]
	T2	-4.1	6.2	[-7.6, -0.7]	-5.6	2.8	[-6.7, -4.4]
SiN: Separated * SSN	T1	-4.7	4.1	[-7.0, -2.4]	-3.8	5.3	[-5.9, -1.6]
	T2	-5.5	3.3	[-7.3, -3.7]	-7.1	2.5	[-8.1, -6.1]
SiN: Collocated * single talker	T1	4.6	3.7	[2.5, 6.6]	5.2	3.6	[3.8, 6.7]
	T2	1.8	2.4	[0.5, 3.1]	1.8	3.8	[0.3, 3.4]
SiN: Separated * single talker	T1	0.9	5.0	[-1.8, 3.7]	-0.7	5.5	[-3.0, 1.5]
	T2	-4.4	4.0	[-6.7, -2.2]	-4.4	3.6	[-5.9, -3.0]
SiN: Average	T1	-0.6	2.5	[-2.0, 0.8]	-0.6	3.3	[-1.9, 0.8]
	T2	-3.1	2.5	[-4.5, -1.7]	-3.8	1.9	[-4.6, -3.1]

was a significant group difference on the FDS task [$t(39) = -2.49$, $p = 0.013$, two-tailed].

Effect of Musical Training, Noise-Type, Spatial Factors and Time on Speech-in-Noise Perception

To determine whether additional musical training over 1 year yielded improvements in SiN perception, a repeated-measures analysis of variance (ANOVA) was conducted on the SRTs, with Group as the between-subjects factor, and Assessment point, Masker type and Spatial location as within-subjects factors. Estimated marginal means for all main effects and interactions are provided in **Table 2**.

The main effect of Assessment point indicated that both groups' SiN perception was significantly better by 2.9 dB after 1 year [$F(1,39) = 33.54$, $p < 0.001$, $\eta_p^2 = 0.46$] consistent with findings that SiN perception improves with age (Hall et al., 2002). The significant main effect of Masker type [$F(1,39) = 123.68$, $p < 0.001$, $\eta_p^2 = 0.76$] indicated that the presence of a single talker led to an increase in SRTs by 5.2 dB compared to spectrally matched noise, across both groups and assessment points. The relative increase in perceptual difficulty experienced when the masker was a single talker is attributable to the acoustic similarity of the target and the masker with resulting informational interference (Dole et al., 2012; Stone et al., 2012), as well as the audible semantic content of the masker, which effectively captures attention in children (Cowan et al., 1999). The significant main effect of Spatial location [$F(1,39) = 59.25$, $p < 0.001$, $\eta_p^2 = 0.60$] indicated that across Group, Assessment point and Masker type factors, the average SRT in the collocated listening conditions was 3.4-dB higher compared to spatially separated listening conditions. This corroborates studies with adults and children indicating a benefit of spatially separating target and maskers (Litovsky, 2005; Johnstone and Litovsky, 2006).

The interaction between Assessment point and Group was not significant [$F(1,39) = 0.59$, $p = 0.448$, $\eta_p^2 = 0.018$], suggesting that the two groups did not differ in SiN perception, neither at baseline nor after providing additional musical training to one of the groups.

An interaction between Masker type and Spatial location and subsequent simple-effects analysis indicated that when the masker was SSN, speech in the collocated condition was significantly harder to perceive by 1.3 dB than in the spatially separated condition. When the masker was a single talker, this difference increased to 5.3 dB. This 4-dB difference in spatial release from masking shows that spatial cues are more helpful for children's speech perception when dealing with realistic changing-state maskers that would often be present in the classroom environment. Furthermore, SRTs for the collocated condition were 7.3 dB higher in the presence of a single talker than in SSN, indicative of the burden that masker-target similarity and attention capture place on auditory stream segregation in children.

A significant interaction was found between Masker type and Spatial location [$F(1,39) = 15.38$, $p < 0.001$, $\eta_p^2 = 0.28$]. A simple-effect analysis revealed that spatially separating the masker from the target resulted in better SiN perception regardless of the type of masker: when the masker was SSN, speech in spatially separated conditions was significantly easier to perceive by 1.3 dB than when collocated [$F(1,39) = 4.12$, $p = 0.05$, $\eta_p^2 = 0.095$], but when the masker was a single talker, this increase between separated and collocated conditions grew to 5.5 dB [$F(1,39) = 54.61$, $p < 0.001$, $\eta_p^2 = 0.5$]. Furthermore, under both spatial conditions, speech masked by SSN was more intelligible than when masked by the single talker: when the masker was spatially separated, speech perception masked by SSN was 7 dB easier to discern than the single talker [$F(1,39) = 21.39$, $p < 0.001$, $\eta_p^2 = 0.35$], but this difference decreased to 3 dB when the masker was collocated but remained significant [$F(1,39) = 94.91$, $p < 0.001$, $\eta_p^2 = 0.71$].

Another significant interaction was found between Masker type and Assessment point [$F(1,39) = 7.79$, $p = 0.008$, $\eta_p^2 = 0.17$]. The simple effects analysis indicated that at both assessment points, SSN was the less challenging masker: SRTs at T1 were 6.2 dB better for SSN than for the single talker masker [$F(1,39) = 102.02$, $p < 0.001$, $\eta_p^2 = 0.72$], and at T2, the difference was reduced to 4.3 dB but remained significant [$F(1,39) = 62.43$, $p < 0.001$, $\eta_p^2 = 0.62$]. Furthermore, the improvement between the two assessment points was greater for the single talker than SSN: when the masker was SSN, the significant increase from T1 to T2 was almost 2 dB [$F(1,39) = 9.04$, $p = 0.005$, $\eta_p^2 = 0.19$], and this increase between assessment points grew to 3.8 dB when the masker was a single talker [$F(1,39) = 47.41$, $p < 0.001$, $\eta_p^2 = 0.55$]. This suggests that there are different developmental trajectories for coping with energetic and informational maskers. While the effect of the energetic masker (SSN) takes place in the auditory periphery, the effect of the informational masker (single talker) is located more centrally and probably involves cognitive processes. That the developmental effect was larger in the single-talker masker indicates that cognitive abilities which assist with SiN

TABLE 2 | Estimated marginal mean speech-reception thresholds (SRTs), standard error (SE) and the lower and upper range of the 95% confidence intervals (CI 95%) for the main effects and interactions.

Factors	Mean (dB)	SE	CI 95% [lower, upper]
T1	-0.6	0.5	[-1.6, 0.4]
T2	-3.5	0.4	[-4.2, -2.7]
SSN	-4.6	0.4	[-5.5, -3.8]
Single talker	0.6	0.4	[-0.2, 1.4]
Collocated	-0.3	0.4	[-1.1, 0.5]
Separated	-3.7	0.5	[-4.6, -2.8]
Collocated * SSN	-4.0	0.6	[-5.1, -2.8]
Collocated * Single talker	3.4	0.5	[2.4, 4.3]
Separated * SSN	-5.3	0.5	[-6.3, -4.3]
Separated * Single talker	-2.2	0.6	[-3.4, -0.9]
SSN * T1	-3.7	0.6	[-4.9, -2.4]
SSN * T2	-5.6	0.4	[-6.5, -4.7]
Single talker * T1	2.5	0.5	[1.4, 3.6]
Single talker * T2	-1.3	0.4	[-2.2, -0.4]

perception develop faster than those attributable to peripheral auditory processing.

Effect of Musical Training on Phonological Short-Term Memory

A repeated-measures ANOVA, with the between-subjects factor Group and the within-subjects factor Assessment point, was conducted on the FDS scores to determine whether additional musical training yielded improvements in phonological short-term memory. There was a significant effect of Group [$F(1,39) = 9.54, p = 0.004, \eta_p^2 = 0.197$], with higher FDS score in the musical training group at both baseline and T2 [$t(39) = -1.84, p = 0.022$, two-tailed]. Within-subject effects indicated that, relative to T1, the average FDS score increased from 10.2 (SD = 3.1) to 10.4 points (SD = 2.5) at T2, but this increase was not significant [$F(1,39) = 0.17, p = 0.684, \eta_p^2 = 0.004$]. The interaction between Assessment point and Group was also not significant [$F(1,39) = 0.41, p = 0.528, \eta_p^2 = 0.01$]. Therefore, neither age-related development nor musical training produced improvements in FDS score in relation to baseline performance.

Correlations Between Speech-in-Noise Perception and Phonological Short-Term Memory

The relationship between FDS scores and SRTs at T1 and T2 was assessed using two-tailed Pearson correlations. Results indicated significant covariance in only one of the listening conditions, namely when the SSN was collocated with the target speech at both T1 ($r = -0.35, p = 0.026$) and T2 ($r = -0.45, p = 0.003$). Correlations between FDS scores and SRTs in the other three conditions were non-significant (all $p > 0.07$).

DISCUSSION

Effect of Musical Training on Speech-in-Noise Perception

The primary aim of this study was to assess whether additional musical instrument training provided over the course of 1 year improves speech perception under the sorts of challenging acoustic conditions children could realistically expect to experience in a classroom. Namely, environments in which energetic and informational maskers in various spatial relationships with the target speech would tax speech perception. However, there was no significant interaction between Assessment point and Group; that is, musical training was not associated with changes in SiN perception. Interactions that were predicted to show a musicians' advantage for SRTs under various masker and spatial manipulations were also not significant (Group \times Assessment point \times Masker type; Group \times Assessment point \times Spatial location). No other study to date has compared effects of musical training on SRTs in children using different masker types and target-masker spatial combinations in 5- to 7-year-old children. Therefore, in what follows, findings from previous cross-sectional and longitudinal studies which show parallels with the present study

but were conducted with children of various ages as well adults will be considered.

In a cross-sectional study by Strait et al. (2012), 7- to 13-year-old children with at least 4 years of musical training or no musical training were tested on different SiN perception tasks. Consistent with the present study's observations, the authors found no evidence for a musicians' advantage for speech perception in collocated babble or SSN. However, there was an advantage for musicians' speech perception when the SSN was spatially separated from the target speech. The masker and spatial conditions used in both studies had the potential to indicate whether musical training improves either peripheral auditory processing, cognition, or both. If the benefits of musical training were for peripheral auditory processing, speech perception under separated and energetic masker conditions would have been predicted because these conditions rely more on peripheral auditory processing than cognition. If benefits of musical training were cognitive, however, speech perception under the more cognitively demanding collocated and informational masker conditions would have been predicted in the musical-training group. In the case that both these processes were improved through musical training, both spatial and masker conditions would have shown improvement. As the cumulative findings of Strait et al. (2012) and the present study indicate no musicians' advantage for collocated conditions accompanied by informational maskers (i.e., babble noise or a single talker, respectively), a cognitive advantage of musical training cannot be concluded. Although Strait et al. (2012) found a musicians' advantage for speech perception under spatially separated energetic masker (i.e., SSN) conditions, the present study failed to demonstrate such trends longitudinally. Therefore, a benefit for musical training for peripheral auditory processing remains to be conclusively established.

A longitudinal musical-training study with children aged 6–9 years conducted by Slater et al. (2015) investigated whether musical training of up to 4 h per week over 2 years improves speech perception in collocated SSN compared to controls who received no musical training. After 1 year, the two groups did not perform significantly differently but a musicians' advantage was found after the second year of training. The discrepancy between this observation and the present study's findings might result from the considerable difference in the amount of the musical training provided in the two studies. However, cross-sectional studies with at least 4 years of musical training (Strait et al., 2012) and adults with over 10 years of musical training (Ruggles et al., 2014; Boebinger et al., 2015) reported no benefits for speech perception in collocated SSN for children either. Further longitudinal investigations are warranted to interpret these conflicting results.

Effect of Musical Training on Phonological Short-Term Memory

A secondary aim of this study was to test if musical training improved phonological short-term memory, which, in turn, could mediate improvements in SiN perception. At baseline, the musical-training group showed significantly higher FDS scores

and this advantage was maintained over time. Although groups were not equally matched at baseline, the ANOVA indicated whether the increase relative to baseline scores over time was greater in the musical training group than controls. The main effect of Assessment point indicated that FDS did not improve significantly over the course of 1 year across groups, and the non-significant interaction between Assessment point and Group meant that the relative increase in FDS was not higher in either group.

These findings contrast with results of Lee et al. (2007) who showed that 12-year-old children with an average of 6 years of musical training had better FDS than non-musicians, and results of Strait et al. (2012) who reported better auditory working memory in musically trained children aged 7 to 13 years. Strait et al. (2012) further reported that the correlation between the number of years of musical training received and auditory working memory ability was “marginally significant” ($r = 0.38$, $p = 0.08$), strongly implying that musical training was causally responsible for the measured between-group difference. Since the studies by Lee et al. (2007) and Strait et al. (2012) were cross-sectional, it cannot be excluded that these findings might be due to pre-existing between-group differences.

However, longitudinal evidence indicates that musically trained children’s phonological short-term memory advantage, indicated by cross-sectional studies, are not necessarily due to pre-existing differences masquerading as training effects. A study by Roden et al. (2014) showed that 45 min of weekly musical training over 1 year in 7- to 8-year-old children significantly improved phonological short-term memory capacity. Somewhat surprisingly, the present study, even though methodological very similar (using also a longitudinal design, a comparable cognitive test, similarly aged participants, and a musical-training regimen of similar duration and frequency) failed to find evidence for a musical training-based cognitive improvement.

Correlations Between Speech-in-Noise Perception and Phonological Short-Term Memory

The strength of the relationships between phonological short-term memory and SiN perception was assessed using Pearson correlations between FDS scores and SRTs in the different masker and spatial conditions. Across groups, there was a significant moderate inverse correlation at T1 and T2 when the masker was collocated SSN. Similarly, Strait et al. (2012) found that auditory working memory correlated significantly with SiN perception in spatially separate SSN. Although spatial conditions differed, both studies found that the energetic masker used (i.e., SSN) covaried significantly with memory processes. This suggests that these cognitive skills are most useful when dealing with speech-perception challenges to the auditory periphery. However, it would be more intuitive to expect that cognitive skills should be useful when dealing with the more cognitively demanding maskers (i.e., informational maskers) and spatial conditions (i.e., collocated). Although, less obviously cognitively taxing conditions (e.g., spatially separated SSN maskers) could have a cognitive component for which stronger cognitive abilities could potentially provide benefits.

Limitations

Most prior studies investigating the musicians’ advantage used a cross-sectional design, probably due to logistical and practical difficulties associated with the implementation of an actual musical-training intervention. For the present study, a longitudinal design was adopted so as to investigate possible causal relationships between the studied variables. To mimic a realistic context for a training program targeting typically developing young children, and also for logistic reasons, the musical training was delivered as part of the school curriculum. These choices imposed certain limits on the experimental design of the current study. First, the children were not randomly assigned to one of the two groups, limiting the causal claims that could be made by the present study. Their choice to attend the music-focused or sports-focused school determined their group membership. Hence, a bias in terms of participant characteristics (e.g., motivation, cognitive abilities) cannot be ruled out, even though all participants were normally performing pupils and the two groups did not differ in age or maternal socio-economic status. Second, the nature, amount and frequency of musical training was fixed by the curriculum in the music-focused school. It could be argued that other forms of or more musical training could have produced improvements in SiN perception and/or in phonological short-term memory capacity. However, it should be noted that studies using even less musical training have reported significant effects of musical training on cognitive abilities, such as improvements in phonological short-term memory after 45-min-long weekly training over 1 year (Roden et al., 2014) or in reading ability after 30-min-long weekly training for 8 months (Myant et al., 2008). Finally, although the present study considered some potential confounds (i.e., socio-economic status, hearing and language ability) that might have motivated children to take up musical training and might have led to pre-existing between-group inequalities, personality is an additional factor which has shown to be predictive of involvement in musical activities in adults and children (Corrigall et al., 2013; Swaminathan and Schellenberg, 2018). As personality was not measured, it was beyond the scope of this study to evaluate the extent to which this factor contributed to children’s motivations to attend the respective schools, and thus represents a potential confound that should be controlled for in future studies.

CONCLUSION

This study assessed the impact of 1 year of musical instrument training on phonological short-term memory and SiN perception in children aged 5–7 years. Musical training improved neither phonological short-term memory, nor SiN perception in any of the listening conditions combining different maskers and spatial target-masker configurations that aimed to simulate realistic classroom conditions. This contrasts with previous studies in similarly aged children reporting evidence of musical-training benefits for SiN perception (Slater et al., 2015) and phonological short-term memory (Roden et al., 2014).

While our study adds to the list of investigations failing to find evidence for a musicians' advantage, more (especially longitudinal) research is warranted to investigate the nature, amount and frequency of musical training required for potential benefits in SiN perception and its underlying cognitive processes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Faculty of Humanities Research Ethics Committee, University of Pretoria (GW20171130HS). Written informed

consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DM designed the study, collected and analyzed the data, wrote the manuscript, and prepared the tables. CF assisted with revising the manuscript and responding to reviewers comments. RE collected the data and provided comments. JL and RL supervised the project and provided comments.

FUNDING

The authors report grant funding from the Swedish Foundation for International Cooperation in Research and Higher Education (IB2017-7004) awarded to RL.

REFERENCES

- Başkent, D., and Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *J. Acoust. Soc. Am.* 139, EL51–EL56. doi: 10.1121/1.4942628
- Benz, S., Sellaro, R., Hommel, B., and Colzato, L. S. (2016). Music makes the world go round: the impact of musical training on non-musical cognitive functions – a review. *Front. Psychol.* 6:2023. doi: 10.3389/fpsyg.2015.02023
- Bergman Nutley, S., Darki, F., and Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Front. Hum. Neurosci.* 7:926. doi: 10.3389/fnhum.2013.00926
- Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627
- Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., and Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *J. Acoust. Soc. Am.* 137, 378–387. doi: 10.1121/1.4904537
- Bradley, J. S., and Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *J. Acoust. Soc. Am.* 123, 2078–2086. doi: 10.1121/1.2839285
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696
- Chan, A. S., Ho, Y. C., and Cheung, M. C. (1998). Music training improves verbal memory. *Nature* 396:128. doi: 10.1038/24075
- Corbin, N. E., Bonino, A. Y., Buss, E., and Leibold, L. J. (2016). Development of open-set word recognition in children: speech-shaped noise and two-talker speech maskers. *Ear Hear.* 37, 55–63. doi: 10.1097/AUD.0000000000000201
- Corrigall, K. A., Schellenberg, E. G., and Misura, N. M. (2013). Music training, cognition, and personality. *Front. Psychol.* 4:222. doi: 10.3389/fpsyg.2013.00222
- Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., and Sauls, J. S. (1999). The role of attention in the development of short-term memory: age differences in the verbal span of apprehension. *Child Dev.* 70, 1082–1097. doi: 10.1111/1467-8624.00080
- DiSarno, N. J., Schowalter, M., and Grassa, P. (2002). Classroom amplification to enhance student performance. *Teach. Except. Child.* 34, 20–25. doi: 10.1177/004005990203400603
- Dockrell, J. E., and Shield, B. (2012). The impact of sound-field systems on learning and attention in elementary school classrooms. *J. Speech Lang. Hear. Res.* 55, 1163–1176. doi: 10.1044/1092-4388(2011/11-0026)
- Dole, M., Hoen, M., and Meunier, F. (2012). Speech-in-noise perception deficit in adults with dyslexia: effects of background type and listening configuration. *Neuropsychologia* 50, 1543–1552. doi: 10.1016/j.neuropsychologia.2012.03.007
- Fels, J., Buthmann, P., and Vorländer, M. (2004). Head-related transfer functions of children. *Acta Acust. United Acust.* 90, 918–927.
- Fleming, D., Belleville, S., Peretz, I., West, G., and Zendel, B. R. (2019). The effects of short-term musical training on the neural processing of speech-in-noise in older adults. *Brain Cogn.* 136:103592. doi: 10.1016/j.bandc.2019.103592
- Franklin, M. S., Sledge Moore, K., Yip, C. Y., Jonides, J., Rattray, K., and Moher, J. (2008). The effects of musical training on verbal memory. *Psychol. Music* 36, 353–365. doi: 10.1177/0305735607086044
- Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.* 106, 3578–3588. doi: 10.1121/1.428211
- Fuller, C. D., Galvin, J. J. III, Maat, B., Free, R. H., and Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Front. Neurosci.* 8:179. doi: 10.3389/fnins.2014.00179
- Hall, J. W. III, Grose, J. H., Buss, E., and Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear Hear.* 23, 159–165. doi: 10.1097/00003446-200204000-00008
- Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J. Acoust. Soc. Am.* 115, 833–843. doi: 10.1121/1.1639908
- Johnstone, P. M., and Litovsky, R. Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *J. Acoust. Soc. Am.* 120, 2177–2189. doi: 10.1121/1.2225416
- Klatte, M., Bergström, K., and Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Front. Psychol.* 4:578. doi: 10.3389/fpsyg.2013.00578
- Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., et al. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *J. Neurosci.* 34, 11913–11918. doi: 10.1523/JNEUROSCI.1881-14.2014
- Lee, Y. S., Lu, M. J., and Ko, H. P. (2007). Effects of skill training on working memory capacity. *Learn. Instr.* 17, 336–344. doi: 10.1016/j.learninstruc.2007.02.010
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375
- Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *J. Acoust. Soc. Am.* 117, 3091–3099. doi: 10.1121/1.1873913
- Lorenzi, C., Dumont, A., and Füllgrabe, C. (2000). Use of temporal envelope cues by children with developmental dyslexia. *J. Speech Lang. Hear. Res.* 43, 1367–1379. doi: 10.1044/jslhr.4306.1367
- MacCutcheon, D., Pausch, F., Fels, J., and Ljung, R. (2018). The effect of language, spatial factors, masker type and memory span on speech-in-noise thresholds in sequential bilingual children. *Scand. J. Psychol.* 59, 567–577. doi: 10.1111/sjop.12466
- MacCutcheon, D., Pausch, F., Füllgrabe, C., Eccles, R., van der Linde, J., Panebianco, C., et al. (2019). The contribution of individual differences in memory span and language ability to spatial release from masking in young

- children. *J. Speech Lang. Hear. Res.* 62, 3741–3751. doi: 10.1044/2019_JSLHR-S-19-0012
- McCreery, R. W., Walker, E., Spratford, M., Lewis, D., and Brennan, M. (2019). Auditory, cognitive, and linguistic factors predict speech recognition in adverse listening conditions for children with hearing loss. *Front. Neurosci.* 13:1093. doi: 10.3389/fnins.2019.01093
- Myant, M., Armstrong, W., and Healy, N. (2008). Can music make a difference? A small scale longitudinal study into the effects of music instruction in nursery on later reading ability. *Educ. Child Psychol.* 25, 83–100.
- Nelson, P., Kohnert, K., Sabur, S., and Shaw, D. (2005). Classroom noise and children learning through a second language. *Lang. Speech Hear. Serv. Sch.* 36, 219–229. doi: 10.1044/0161-1461(2005/022)
- Neuman, A. C., Wroblewski, M., Hajicek, J., and Rubinstein, A. (2010). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear Hear.* 31, 336–344. doi: 10.1097/AUD.0b013e3181d3d514
- Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661. doi: 10.1097/AUD.0b013e3181b412e9
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142
- Prodi, N., Visentin, C., Borella, E., Mammarella, I. C., and Di Domenico, A. (2019). Noise, age and gender effects on speech intelligibility and sentence comprehension for 11 to 13 years old children in real classrooms. *Front. Psychol.* 10:2166. doi: 10.3389/fpsyg.2019.02166
- Purdy, S. C., Smart, J. L., Baily, M., and Sharma, M. (2009). Do children with reading delay benefit from the use of personal FM systems in the classroom? *Int. J. Audiol.* 48, 843–852. doi: 10.3109/14992020903140910
- Renfrew, C. E. (1980). *Action Picture Test*. Headington: C. E. Renfrew.
- Rhyner, P. M. (ed.) (2009). *Emergent Literacy and Language Development: Promoting Learning in Early Childhood*. New York, NY: Guilford Press.
- Roden, I., Grube, D., Bongard, S., and Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychol. Music* 42, 284–298. doi: 10.1177/0305735612471239
- Rönnerberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 47(Suppl. 2), S171–S177. doi: 10.1080/14992020802301167
- Ruggles, D. R., Freyman, R. L., and Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One* 9:e86980. doi: 10.1371/journal.pone.0086980
- Schröder, D. (2011). *Physically Based Real-time Auralization of Interactive Virtual Environments*, Vol. 11. Berlin: Logos Verlag.
- Semel, E., Wiig, E. H., and Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals*, 4th Edn, San Antonio, TX: The Psychological Corporation.
- Shield, B. M., and Dockrell, J. E. (2003). The effects of noise on children at school: a review. *Buuld. Acoust.* 10, 97–116. doi: 10.1260/135101003768965960
- Shield, B. M., and Dockrell, J. E. (2008). The effects of environmental and classroom noise on the academic attainments of primary school children. *J. Acoust. Soc. Am.* 123, 133–144. doi: 10.1121/1.2812596
- Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., and 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
- Stone, M. A., Füllgrabe, C., and Moore, B. C. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *J. Acoust. Soc. Am.* 132, 317–326. doi: 10.1121/1.4725766
- Strait, D. L., O'Connell, S., Parbery-Clark, A., and Kraus, N. (2013). Musicians' enhanced neural differentiation of speech sounds arises early in life: developmental evidence from ages 3 to 30. *Cereb. Cortex* 24, 2512–2521. doi: 10.1093/cercor/bht103
- Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang.* 123, 191–201. doi: 10.1016/j.bandl.2012.09.001
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G. Jr., and Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Sci. Rep.* 5:11628. doi: 10.1038/srep11628
- Swaminathan, S., and Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Sci. Rep.* 8:9223. doi: 10.1038/s41598-018-27571-2
- Swanepoel, D. W., Myburgh, H. C., Howe, D. M., Mahomed, F., and Eikelboom, R. H. (2014). Smartphone hearing screening with integrated quality control and data management. *Int. J. Audiol.* 53, 841–849. doi: 10.3109/14992027.2014.920965
- Talamini, F., Altoè, G., Carretti, B., and Grassi, M. (2017). Musicians have better memory than nonmusicians: a meta-analysis. *PLoS One* 12:e0186773. doi: 10.1371/journal.pone.0186773
- Tallal, P. (2014). "Experimental studies of language learning impairments: from research to remediation," in *Speech and Language Impairments*, eds D. V. M. Bishop, and L. Leonard, (Hove: Psychology Press).
- Vickers, D., Degun, A., Canas, A., Stainsby, T., and Vanpoucke, F. (2016). Deactivating cochlear implant electrodes based on pitch information for users of the ACE strategy. *Adv. Exp. Med. Biol.* 894, 115–123. doi: 10.1007/978-3-319-25474-6_13
- Wightman, F., Kistler, D., and Brungart, D. (2006). Informational masking of speech in children: auditory-visual integration. *J. Acoust. Soc. Am.* 119, 3940–3949. doi: 10.1121/1.2195121
- Wightman, F. L., and Kistler, D. J. (2005). Informational masking of speech in children: effects of ipsilateral and contralateral distracters. *J. Acoust. Soc. Am.* 118, 3164–3176. doi: 10.1121/1.2082567
- Zendel, B. R., West, G., Belleville, S., and Peretz, I. (2019). Music 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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 MacCutcheon, Füllgrabe, Eccles, van der Linde, Panebianco and Ljung. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.