



Listening Effort for Speech in Noise Perception Using Pupil Dilation: A Comparison Among Percussionists, Non-Percussionists, and Non-Musicians

Vallampati Lavanya^{1*}, Ramaprasad Rajaram^{2*}, Ramya Vaidyanath¹, and Ajith Kumar Uppunda³

¹Department of Audiology, Sri Ramachandra Faculty of Audiology and Speech Language Pathology, Sri Ramachandra Institute of Higher Education and Research, Chennai, India

²Health Economist and Carnatic Musician (Independent Researcher), Chennai, India

³Department of Audiology, All India Institute of Speech and Hearing, Mysuru, India

Received June 22, 2023
Revised September 22, 2023
Accepted September 27, 2023

Address for correspondence

Ramya Vaidyanath, PhD
Department of Audiology,
Sri Ramachandra Faculty of
Audiology and Speech Language
Pathology, Sri Ramachandra Institute
of Higher Education and Research,
Porur Chennai 600116, India
Tel +91-8105737392
Fax +91-044-24767008
E-mail
ramyavaidyanath@sriramachandra.
edu.in

*These authors contributed equally to this work.

Background and Objectives: Most studies in literature attribute the benefits of musical training on speech in noise (SIN) perception to “experience-based” plasticity, which assists in the activation of speech-processing networks. However, whether musicianship provides an advantage for the listening effort (LE) required to comprehend speech in degraded environments has received less attention. The current study aimed to understand the influence of Indian classical music training on SIN perception and its related LE across percussionists, non-percussionists, and non-musicians. **Subjects and Methods:** A quasi-experiment was conducted on 16 percussionists, 17 non-percussionists, and 26 non-musicians aged 18–35 years with normal hearing. In phase 1, musical abilities were assessed using Mini-Profile of Music Perception Skills (Mini-PROMS). Phase 2 examined SIN using Tamil Phonemically-Balanced Words and Tamil Matrix Sentence Test at +5 dB, 0 dB, and -5 dB SNR and LE using pupillometry, measuring pupil dilations with an eye-tracker. **Results:** Fractional Logit and Linear Regression models demonstrated that percussionists outperformed non-percussionists in Tuning and Speed subsets of Mini-PROMS. Percussionists outperformed non-percussionists and non-musicians in SIN and LE at -5 dB SNR for words and at 0 dB and -5 dB SNR for sentences. **Conclusions:** Percussionists have the greatest advantage in decoding SIN with reduced LE followed by non-percussionists and non-musicians, demonstrating musician-advantage in most challenging listening conditions.

J Audiol Otol

Keywords: Listening effort; Musical training; Speech-perception-in-noise; Pupil dilation.

Introduction

Perceiving and interpreting speech in noise (SIN) is an integral part of everyday communication. Comprehending a new sentence during SIN typically entails both deciphering the individual words as well as the acoustic patterns within the sentence. These acoustic patterns or temporal variations can provide key perceptual cues in difficult listening conditions, such as attending to a person's speech in a busy restaurant. Musicians may have a distinctive advantage in efficient-

ly decoding these acoustic patterns since, by definition, they are trained auditorily. Consequently, many studies in the literature provide evidence that expertise in melodic and rhythmic processing aid with speech processing [1-3], while few others do not offer much empirical support for that claim [4-6].

Musical training involves neural circuitry that encompasses complex motor, auditory, and visual skills, hence musical training and speech processing share structural and functional overlaps [7-10]. Also, rhythm sensitivity or training, in particular, has been shown to influence better speech perception, as temporal cues are critical for both music (rhythm) and speech perception, especially in noise [1,3,11,12]. Although the underlying temporal dynamics of speech and music can be different, nuanced music training (associated to rhythm) helps

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

overcome challenges in comprehending SIN because musicians, through their musical training, have the ability to deconstruct subtle fluctuating and contrasting durations of temporal features [13-15].

In general, while the music-SIN association is well researched, the effort expended to decode SIN for musicians is limited [5,16,17]. The current study tries to investigate this less explored, yet crucial area in hearing science. Listening effort (LE) is defined as “the mental exertion required in attending to and understanding an auditory message” [18^{p.434}]. LE serves as a valuable evaluation metric even in the lack of intelligibility scores because two individuals may exert varying levels of effort despite achieving same level of intelligibility scores [19]. While speech intelligibility scores are generally attributed to LE, the common assumption that individuals with less intelligible speech exert more effort has not been supported by empirical studies [19-21].

In line with the studies above, to quantify LE, in our current work we employed pupillometry measure which is a consistent time-series marker and a robust indicator of processing load [20]. Pupil responses emerge as a natural combination of sympathetic and parasympathetic nervous system activity, while pupil dilation is linked to activation of sympathetic nervous system [20]. As speech requires rapid instantaneous auditory encoding disseminated over time, pupillometric methods offer the advantage of measuring LE by displaying continuous changes in dilation across time. Previous research using pupillary measures revealed that musicians had enhanced pitch discrimination and speech perception in noise with reduced processing effort than non-musicians [16,17].

In the current study, whether musicians who were percussionists had an edge over non-percussionists in SIN perception and the associated LE was examined. This was done as rigorous rhythmic training helps efficiently draw short-term attentional resources associated with retention and recall of a signal [22,23]. Owing to such enhanced auditory perceptual and cognitive skills, it was hypothesized that percussionists may perform better than others on attentional allocation tasks with reduced effort.

Current literature on the nexus between music-SIN (and LE) is divided on which specific components of musical proficiency contribute to improved speech perception for musicians. Certain studies have documented the influence of genre of musical training and the advantages musicians behold [1,24], while others do not find such specific advantages [6,23]. The present study, to the best of our knowledge, is one of the first few to use Indian classical genre to understand the musician-ship-LE nexus. Indian classical music distinguishes from other genres of music wherein, training predominantly relies on

learning to sing or play by the ear, which is pertinent in studying the relationship between LE and musical training [24].

The current study incorporated South Indian “Carnatic” music, which is unique in many ways, particularly in the style of its rendition that involves “gamakas” (oscillation across notes) and microtones. This music form is built upon “Raga” (melodic aspect) and “Tala” (rhythmic components), and is distinctly complex with significant subtle pitch fluctuations compared to the melody and rhythm in Western music [24,25]. While general musical training has been shown to increase SIN scores by 6% to 8%, Carnatic music has demonstrated enhanced SIN scores by up to 16% to 18% [24].

In this background, the current study aims to understand and compare the influence of Indian classical music training on SIN perception and its related LE across percussionists, non-percussionists, and non-musicians, using task elicited pupil dilations, and analyzing the findings by employing regression techniques.

Subjects and Method

Participants

A total of 59 participants (34 females and 25 males) in the age range of 18 to 35 years (mean age: 24.1 years) were recruited in the study. The study population consisted of 17 non-percussionists (vocal and stringed instruments), 16 percussionists (drum-based instruments), and 26 non-musicians (Table 1). In line with earlier research [1,23], musicians were classified as those formally trained in Indian classical music Carnatic with minimum 5 years of continuous training in the last 6 years and 3.5 hours of practice/week in any 2 of the last 3 years. Non-musicians were classified as those having no or less than 3 years of formal music training in their lifetime (and not trained or practiced in the last 4–5 years). However, none of the non-musicians recruited in the study had any musical training in their life time. Only participants who reported no history of any otologic or neurological deficits were included. All individuals underwent hearing screening and had pure-tone air conduction thresholds within 25 dB HL across all the octave frequencies from 250 Hz to 8 kHz (ANSI S3.6, 1996).

A participant recruitment questionnaire was administered to ensure all participants selected satisfy the inclusion criteria. The questionnaire had information on demographics, handedness, musical training, economic status, occupation, and noise exposure. Participation willingness and written consent was obtained in accordance with Institutional Ethics Committee of Sri Ramachandra Institute of Higher Education and Research (IEC-NI/20/SEP/75/82).

Table 1. Descriptive statistics of percussionists, non-percussionists, and non-musicians

Parameters	Percussionists (n=17)	Non-percussionists (n=16)	Non-musicians (n=26)
Age (yr)	23.56 ± 4.22	23.30 ± 3.14	23.52 ± 3.29
Male (%)	100	82	30
Duration of musical training (yr)	19.52 ± 3.29	14.80 ± 3.18	NA
Age at first onset of musical training (yr)	4.03 ± 2.48	4.75 ± 0.68	NA
Listening to classical music (hours/day)	1.50 ± 0.68	2.00 ± 0.58	0.50 ± 1.12
Listening to non-classical music (hours/day)	0.50 ± 0.90	0.50 ± 1.12	1.00 ± 0.87
Languages known			
Bilingual	7 (41)	5 (32)	17 (65)
Multilingual	10 (59)	11 (68)	9 (35)

Data are presented as mean ± standard deviation or n (%) unless otherwise noticed. NA, not applicable

Procedure

The study was carried out in two phases. In phase one, the participants' general musical ability and working memory was tested. Working memory was included as literature has demonstrated that innate musical capabilities and working memory to have an influence on processing SIN and LE [5,25]. In phase two, the LE was measured using pupillometry. The entire test procedure was administered in a quiet room using circum-aural headphones (Sennheiser HDA 280 pro; Sennheiser Electronic GmbH & Co., KG, Wedemark, Germany) calibrated for an output of 70 dB SPL and routed through the laptop (Lenovo X1 Carbon with Intel i7 Processor; Sichuan, China). The average noise levels of the testing room was below 30 dBA and did not cross 40 dBA and was monitored using NIOSH Sound Level Meter app (Centers for Disease Control and Prevention; <https://www.cdc.gov/niosh/topics/noise/app.html>), throughout the testing duration.

Stimulus

Phonemically balanced (PB) wordlist in Tamil [26] and Tamil Matrix Sentence Test (TMST) [27], in the presence of Tamil multi-talker babble [28] were used for measuring word and sentence perception in noise. The root mean square of speech and noise were matched and combined at 3 fixed SNRs of +5 dB, 0 dB, and -5 dB, using MATLAB version 2014b software (MathWorks, Natick, MA, USA). The onset of noise for words was 1.5 seconds before the onset on the speech stimuli, with the offset of speech and noise being the same and the inter-stimulus interval for words was maintained at 3 seconds. A total of 3 lists (one list at each SNR) with a total of 25 words per list were administered.

The 5-word sentences of TMST had a fixed semantic structure with nouns, numbers, adjectives, objects, and verbs. The order of these words in the sentences followed the sentence structure of Tamil language. Each word in the sentence list had

10 alternatives with a total of 50 words per sentence list. A total of 3 lists, with 10 sentences per list for sentence perception were chosen. The onset of noise was 3 seconds before the onset on the speech stimuli, with the offset of speech and noise being the same and the inter-stimulus interval was 6 seconds. Each of the PB words and TMST sentences were binaurally presented at an intensity level of 70 dB SPL at 3 SNRs of +5 dB, 0 dB, and -5 dB.

Phase I: Assessment of musical abilities and working memory

All the participants were administered the online Mini-PROMS test (Profile of Music Perception Skills) to study their individual musical ability, which is standard in the literature [29]. This test comprised of 4 subsets for evaluation—melody, tuning, accent, and tempo. Each stimulus, from all the 4 subsets consisted of a block of 3 musical tones (2 reference and 1 comparison). The participants were instructed to listen to the stimulus and identify if the 3 stimuli were same/different. The participants had to choose the right option from the 5 choices displayed. A score of “2” was given for correct identification while “1” was scored for correct but uncertain response and “0” for incorrect identification making a total maximum possible score of 36. After completion of the test, the total score and the subset total score—melody (10), tuning (8), accent (10), speed/tempo (8)—were recorded.

A Backward Digit Span test was conducted to assess working memory, which is also standard in the literature [5]. The participants were instructed to listen to the sequence of digits presented aurally and repeat the digits in a backward manner. The digits were presented in an increasingly larger digit sequences until the point when the participants were unable to accurately repeat the sequence of digits backward. Digit span was calculated as the longest sequence of accurate digits correctly repeated in two continuous presentations.

Phase II: Assessment of LE for speech perception in noise

Assessment of binaural speech perception in noise

The Tamil PB words and TMST in eight-talker babble (consisting of four female and four male speakers), were presented binaurally at 70 dB SPL at randomized SNRs of +5 dB, 0 dB and -5 dB. The participants were instructed to repeat the words and sentences heard in the presence of multi-talker babble. Each correctly identified word was scored as “1” and “0” for incorrect responses making a total maximum score of 25 per list and a total of 75 words were scored. Similarly, for TMST each correctly identified word in a sentence was given a score of “1” with a total possible score of 50 per list and a total of 30 sentences from 3 lists were scored.

Assessment of LE using pupil dilation for binaural speech perception in noise

LE was measured using pupilometer with an eye tracker by monitoring the changes in task evoked pupil dilation. The listeners were presented with PB words and TMST in noise at SNRs of +5 dB, 0 dB and -5 dB and were instructed to repeat the words and sentences heard. During the entire duration of stimuli presentation (6.56 min, sentences; and 7.58 min, words), listeners were asked to visually fixate on a dot presented on a 14" computer screen with a resolution of 1,920×1,050 pixels. Participants were seated 1 m away from the computer screen. An eye tracker system developed by Balance Eye (Cyclops MedTech Pvt. Ltd, Bengaluru, India) was used with a sampling rate of 40 Hz to monitor the participants' pupil area. The eye tracker sampled only from the right eye. The percentage of correct score for word and sentence perception in noise was also measured for each SNR condition. Each participant underwent a total of 105 trials presented at randomized SNR, for a total duration of the experiment of 20 minutes.

Data analysis of pupil dilation

For each trial, the baseline pupil dilation was measured in the pupil resting state—one second interval preceding each experimental block. The mean baseline was then calculated by averaging the pupil size in trials preceding the beginning of stimulation with words and sentences. Each mean baseline was then subtracted from each trial of task-evoked peak pupil dilation: $\text{mean pupil dilation} = \text{task evoked pupil dilation} - \text{mean baseline pupil dilation}$.

The mean pupil size across 75 trials for words and 30 trials for sentences were calculated across the 3 SNR conditions. For a robust measurement, apart from eye closure, pupil

movements exceeding 450 degrees (very high velocity) were also coded as eye blinks. As is typical in the literature, trials containing more than 15% of samples resulting from eye blinks during the test were excluded from the analysis.

Statistical analysis

Prior studies employ group means (analysis of variance) to study the differences in SIN (and LE) outcomes among musicians and non-musicians [5,23]. Current study aims to advance the literature by employing causal analysis that can control for other variables which can potentially affect the variable of interest. Two distinct regression models were employed to account for the continuous and bounded nature of the dependent outcome variables of interest, respectively. To explain, scores from Mini-PROMS and SIN perception were measured in percentages, which by definition are bounded between 0 and 100. Since normal linear regression models are unsuitable to handle bounded dependent variables, a fractional logit model was employed in this regard. Raw scores from SIN were also converted to percentage of correct responses. Also, for ease of comprehension of the relative benefit of musicians compared to non-musicians, we converted the coefficient values from the model into their corresponding odds ratios (OR).

On the other hand, a linear regression model was used to explore the influence of musical training on LE for words and sentence perception in noise—measured as pupil dilation in millimeter—which is a continuous variable. In this regard, unlike earlier studies in LE, we gathered data with a statistically large sample size. Typically, a sample size of more than 30 is considered large statistically to make generalizable inferences. We had a total of 59 participants, which is essential for conducting causal analysis.

Data were analyzed by using the statistical software Stata/MP version 13.0 (StataCorp, College Station, TX, USA).

Results

Performance of SIN and LE among musicians and non-musicians

Fig. 1 presents the average score (mean±2SD) across the three groups: non-musicians, non-percussionists, and percussionists for SIN and its related LE. The performances across the groups are comparable; no perceptible differences are observed in the raw scores of SIN or its related LE across the musician and non-musician groups.

Table 2 presents the results comparing the performance of all musicians together with that of non-musicians, while controlling for other variables, unlike Table 1. Results of fraction-

al logit and linear regression employed for each SNR conditions are shown in Table 2. The results show that only under more challenging listening conditions, i.e., -5 dB SNR for TMST, did musicians outperform non-musicians by almost twice (OR=1.87). Otherwise, there are no statistical difference in SIN scores or LE between the two groups. In normal or less challenging conditions, we do not find any evidence for specific or systematic advantage for musicians over non-musicians either in SIN or in LE, despite controlling for other variables.

Comparison of musical abilities and working memory among percussionists, non-percussionists, and non-musicians

Subsequently, we analyzed sub-group association on LE and SIN by dividing the musician group into percussionists and non-percussionists, for Mini-PROMS scores, working memory, SIN and LE separately. Interestingly, despite controlling for other variables, there were no statistically significant differences between the three groups in working memory ($p=0.10$) and Mini-PROMS scores (overall and sub-categories), making it difficult to demonstrate a musicianship-advantage

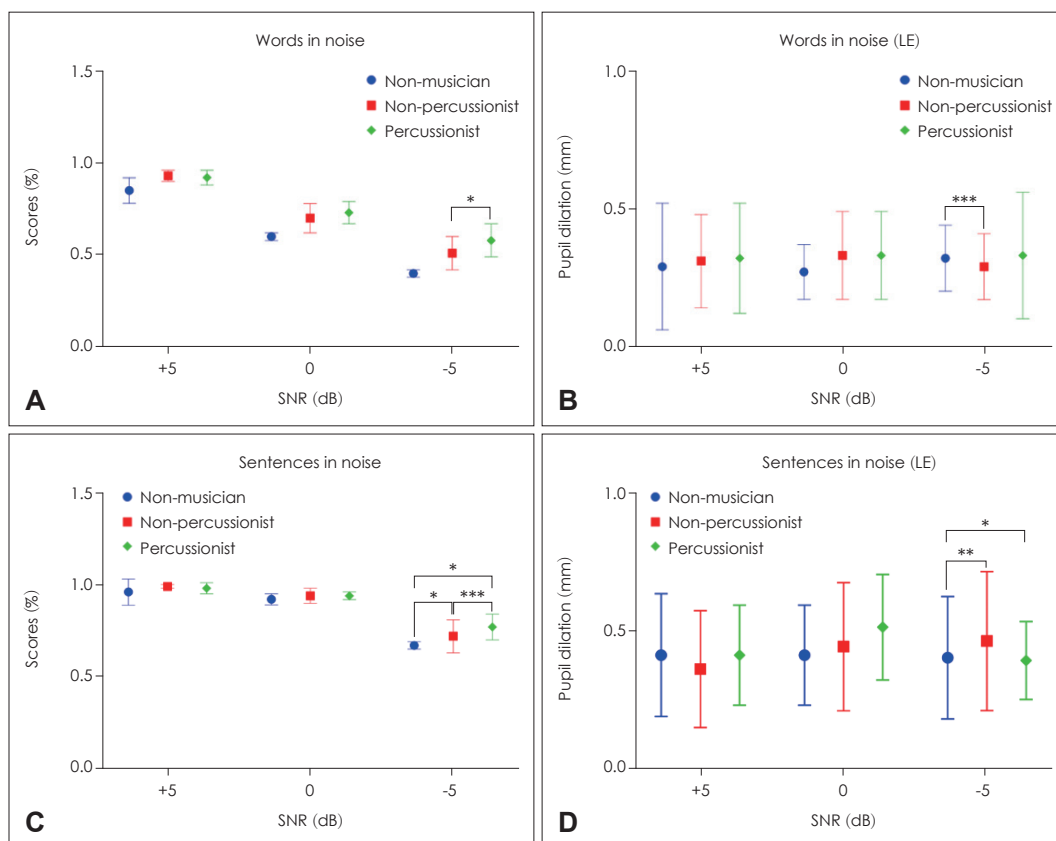


Fig. 1. Comparison of performance of SIN and LE among non-musicians, non-percussionists, and percussionists. A: Perception of PB words in noise. B: Pupil dilation (LE) in perception of PB words in noise. C: Perception of TMST in noise. D: Pupil dilation (LE) in perception of TMST in noise. Error bars represent $\pm 2SD$ of the mean. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; statistical significance. SIN, speech in noise; LE, listening effort; TMST, Tamil Matrix Sentence Test; SD, standard deviation; SNR, signal-to-noise ratio.

Table 2. OR for SIN (fractional logit) and coefficient estimate for LE (linear regression) for PB words and TMST in noise among musicians and non-musicians using controlled models

SNR	PB words in noise				TMST in noise			
	SIN		LE		SIN		LE	
	OR	p	r	p	OR	p	r	p
+5 dB	1.00	0.99	-0.03	0.77	2.27	0.21	-0.13	0.51
0 dB	1.18	0.48	-0.08	0.28	1.40	0.40	-0.15	0.19
-5 dB	1.37	0.19	0.05	0.61	1.87	0.01**	-0.15	0.25

r , coefficient. OR, odds ratio; SIN, speech in noise; LE, listening effort; PB, phonemically-balanced; TMST, Tamil Matrix Sentence Test. ** $p \leq 0.01$

in Mini-PROMS (Table 3). However, a significant difference was observed in percussionists over non-percussionists only for tuning and speed subsets, which is consistent with the literature [1].

Performance of SIN and LE among percussionists, non-percussionists, and non-musicians

We further examined if the SIN perception and LE was different across these sub-groups by employing fractional logit and linear regression models similar to Table 3.

Tables 4 and 5 also presents a very interesting case, in which sub-group differences occur in most difficult listening conditions, i.e., -5 dB SNRs both in SIN and LE, and not otherwise. The left panel in Table 4 for SIN in words shows that across the sub-categories and SNRs, only at -5 dB SNR do percussionists perform much better than non-musicians (OR=5.20). For sentences as well (right panel in Table 4), statistically significant differences were observed across all categories in -5 dB SNR. The order of difference in each of the pairs is note-

worthy; difference in SIN scores was greatest among percussionists compared to non-musicians (OR=6.42), followed by non-percussionists over non-musicians (OR=3.56), and percussionist over non-percussionists (OR=2.45). For LE (Table 5), percussionists exerted less effort (reduced by 0.9 percentage average) over non-musicians and non-percussionists exerted lesser effort (reduced by 0.8 percentage average) than non-musicians, indicating that percussionists and non-percussionists exert less effort for SIN than non-musicians.

Tables 4 and 5 show evidence that in challenging SNRs, musicians with sub-categorical specialization indeed have an advantage over non-musicians for enhanced SIN with lesser effort. In other words, percussionists seem to have the greatest advantage in decoding sentences in noise, over non-percussionists, who in turn, have an advantage over non-musicians. This order seems to be in line with the literature which argues that musicians trained in rhythm have an advantage over others in SIN, particularly in sentence decoding [1,30].

Table 3. OR for Mini-PROMS among non-musicians, non-percussionists, and percussionists using controlled fractional logit model

PROMS	M vs. NM		P vs. NP		P vs. NM		NP vs. NM	
	OR	p	OR	p	OR	p	OR	p
Overall	1.53	0.09	0.55	0.06	0.95	0.94	2.01	0.29
Melody	0.46	0.07	2.31	0.13	0.59	0.65	4.22	0.27
Tuning	0.57	0.20	0.57	0.03*	1.20	0.14	4.22	0.27
Accent	2.15	0.79	0.87	0.72	1.09	0.92	0.61	0.53
Speed	0.47	0.10	0.33	0.03*	0.16	0.27	0.46	0.51

Level of significance *p≤0.05. OR, odds ratio; Mini-PROMS, Mini-Profile of Music Perception Skills; M, musicians; NM, non-musicians; P, percussionists; NP, non-percussionists

Table 4. OR for SIN (PB words and TMST) among non-musicians, non-percussionists, and percussionists using controlled models (fractional logit model)

SNR	Words in noise						Sentence in noise					
	P vs. NP		P vs. NM		NP vs. NM		P vs. NP		P vs. NM		NP vs. NM	
	OR	p	OR	p	OR	p	OR	p	OR	p	OR	p
+5 dB	0.74	0.42	1.69	0.59	1.05	0.94	2.27	0.21	1.01	0.81	0.44	0.21
0 dB	1.46	0.16	1.16	0.85	1.68	0.49	1.78	0.22	0.42	0.51	3.18	0.00***
-5 dB	1.13	0.58	5.20	0.05*	2.41	0.24	2.45	0.00***	6.42	0.02*	3.56	0.05*

*p≤0.05; ***p≤0.001. OR, odds ratio; SIN, speech in noise; PB, phonemically-balanced; TMST, Tamil Matrix Sentence Test; SNR, signal-to-noise ratio; P, percussionists; NP, non-percussionists; NM, non-musicians

Table 5. Coefficient estimate for LE among non-musicians, non-percussionists, and percussionists using controlled models (linear regression model)

SNR	Words in noise						Sentence in noise					
	P vs. NP		P vs. NM		NP vs. NM		P vs. NP		P vs. NM		NP vs. NM	
	r	p	r	p	r	p	r	p	r	p	r	p
+5 dB	0.01	0.91	-0.39	0.32	-0.42	0.10	-0.10	0.44	-0.21	0.80	-0.50	0.45
0 dB	-0.74	0.43	-0.29	0.24	-0.29	0.16	-0.09	0.45	0.19	0.52	0.3	-0.39
-5 dB	0.09	0.41	-0.41	0.22	-0.71	<0.01***	-0.02	0.84	-0.90	0.03*	-0.81	0.01**

r, coefficient. *p≤0.05; **p≤0.01; ***p≤0.001. LE, listening effort; SNR, signal-to-noise ratio; P, percussionists; NP, non-percussionists; NM, non-musicians

Discussion

The purpose of the current study was to understand if musicians trained in Indian classical music performed better on SIN and the associated LE than non-musicians by employing regression models, as against the traditional group mean differences. We subdivided musicians further into percussionists and non-percussionists in this process for few important reasons. For instance, “timing” is an essential component to distinguish speech sounds (especially consonants) that are exclusively differentiated by the voice-onset time [8]. Also, comprehension of a novel speech pattern necessitates complicated temporal processing and may entail processes comparable to those involved in perception and composition of musical phrase [1]. Musical training that focuses on rhythmic skills may thus improve sensitivity to timing patterns that are critical for speech perception, as well as the ability to perceive speech in noisy environments [1,8].

Results from our study, in line with earlier research [1], reveals that percussionists outperformed non-musicians in words-perception and outperformed both non-musicians and non-percussionists in sentence perception, for the most difficult SIN condition (at -5 dB SNR). And, the order of difference in magnitude was maintained across the three groups for LE as well: percussionists outperforming non-musicians, non-percussionists outperforming non-musicians, and percussionists outperforming non-percussionists. Our general finding that musicians outperform non-musicians corroborates with past research [17] suggesting that musicians may be recruiting different strategy than their counterparts, in processing SIN and the associated LE.

Furthermore, it is possible that non-percussionists much like percussionists, also demonstrate similar patterns of enhancement in SIN perception (ascribed to improved rhythm skills), as rhythm plays an integral and fundamental part of overall musical practice regardless of specialization [1,31]—particularly so in Carnatic music. Past studies reported that when listening to SIN, there was increased activation in motor regions implying a greater relevance of temporal cues in suboptimal listening situations [1,22,32,33]. This claim is supported by our study findings which shows non-percussionists also outperformed non-musicians in SIN and LE for sentences in noise, albeit not as much as percussionists did. Hence, it is possible that rhythm has a special role to play in mediating these advantages.

As for the LE associated with SIN, within the musician group, although percussionists outperformed non-percussionists in sentences in noise at -5 dB SNR, the related effort exerted across both the groups were not statistically different.

Since the current study, to the best of our knowledge, is among the first to investigate LE differences across percussionists and non-percussionists, we are unable to draw comparisons to past studies. Therefore, at this point, we only find it reasonable to hypothesize that training in primary instrument may not influence changes in auditory-neural circuitry associated with LE—a claim, which is established for SIN by past studies [1,23].

Although the effort exerted by all the individuals showed an increase in association with increased task complexity, musicianship-advantage on SIN and LE was consistently established only for sentence perception in most challenging conditions i.e., -5 dB SNR, which is again in line with earlier work [17]. This effect may reflect either an improved capacity to distinguish target speech stimuli from ambient noise due to musical training or a greater sensitivity to speech stimuli along the auditory pathway (finer cortical representation) among musicians [1,32].

Relatedly, the lack of difference in less-challenging SNRs in the literature is attributed to the innate variances in timing skills that may alter speech perception, in the absence of musical expertise. In other words, even when particular words are unclear, a listener can determine the rhythm of what is stated when listening to SIN [14]. In fact, studies suggest that limiting the word patterns to those that fit the perceived rhythm, may aid in the process of disambiguating speech [30]. However, cues such as prosody, phonological information, phrase boundaries, and syntactic structure may also be used by the listener to resolve ambiguities.

Overall, the current study does provide evidence to support the claim that musical training improves speech perception in noisy environments, and reduces listening effort especially for sentences, in most difficult masking conditions. The results also show that musicians with rhythmic training have a slightly greater advantage than their non-rhythm trained counterparts in this regard.

Given the specific scope and aim with which the study was conducted, the findings must be contextualized in certain ways. Primarily, our study does not normalize the “easy” or “difficult” task-complexity across the musician and non-musician group. However, individually adjusting the task complexity may aid in understanding the sensitivity of auditory pathway in resolving complex auditory signals (in the current study: speech in noise).

While the current study is a quasi-experimental design comparing group differences across musicians and non-musicians, an intervention-based analysis could better reflect such “within” and “across” group differences providing nuanced insights into understanding effects of standardized musical training.

Future work could also focus on the function of different components of rhythm processing in speech perception, to understand the subtle and complex overlap between music and speech processing.

Conflicts of Interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: Ramaprasad Rajaram, Vallampati Lavanya. Data curation: Ramaprasad Rajaram, Vallampati Lavanya. Formal analysis: Ramaprasad Rajaram, Vallampati Lavanya. Investigation: all authors. Methodology: Ramaprasad Rajaram, Vallampati Lavanya, Ramya Vaidyanath. Project administration: all authors. Software: all authors. Supervision: Ramaprasad Rajaram, Ramya Vaidyanath. Validation: all authors. Writing—original draft: Vallampati Lavanya, Ramaprasad Rajaram. Writing—review & editing: all authors. Approval of final manuscript: all authors.

ORCID iDs

Vallampati Lavanya <https://orcid.org/0000-0002-5332-6472>
 Ramaprasad Rajaram <https://orcid.org/0000-0002-4763-9624>
 Ramya Vaidyanath <https://orcid.org/0000-0003-4093-0424>
 Ajith Kumar Uppunda <https://orcid.org/0000-0002-1368-9834>

Funding Statement

Vallampati Lavanya was funded by the Sri Ramachandra Founder-Chancellor Shri. N. P. V. Ramaswamy Udayar Research Fellowship.

Acknowledgments

The authors would like to acknowledge Mr. Udhaya Kumar R for helping in developing the speech in noise stimulus. The authors would also like to acknowledge Cyclops MedTech Pvt. Ltd, Bengaluru and Mr. Ajith S Rao for developing the stimulus software for Pupillometry, Mr. Narendra Kumar M for significant inputs on software refinement and all the participants for participating in the study. Dr. Heramba Ganapathy is also acknowledged for the initial coordination and support for the experimental set-up. The Department of Audiology, Faculty of Audiology and Speech Language Pathology, SRIHER (DU) is acknowledged for providing the set-up to conduct the study.

REFERENCES

- Slater J, Kraus N. The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cogn Process* 2016;17:79-87.
- Swaminathan J, Mason CR, Streeter TM, Best V, Kidd G Jr, Patel AD. Musical training, individual differences and the cocktail party problem. *Sci Rep* 2015;5:11628.
- Yates KM, Moore DR, Amitay S, Barry JG. Sensitivity to melody, rhythm, and beat in supporting speech-in-noise perception in young adults. *Ear Hear* 2019;40:358-67.
- Boebinger D, Evans S, Rosen S, Lima CF, Manly T, Scott SK. Musicians and non-musicians are equally adept at perceiving masked speech. *J Acoust Soc Am* 2015;137:378-87.
- Escobar J, Mussoi BS, Silberer AB. The effect of musical training and working memory in adverse listening situations. *Ear Hear* 2019;41:278-88.
- Ruggles DR, Freyman RL, Oxenham AJ. Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One* 2014;9:e86980.
- Fauvel B, Groussard M, Eustache F, Desgranges B, Platel H. Neural implementation of musical expertise and cognitive transfers: could they be promising in the framework of normal cognitive aging? *Front Hum Neurosci* 2013;7:693.
- Anderson S, Kraus N. Neural encoding of speech and music: implications for hearing speech in noise. *Semin Hear* 2011;32:129-41.
- Herholz SC, Zatorre RJ. Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron* 2012;76:486-502.
- Asaridou SS, McQueen JM. Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Front Psychol* 2013;4:321.
- Smith MR, Cutler A, Butterfield S, Nimmo-Smith I. The perception of rhythm and word boundaries in noise-masked speech. *J Speech Hear Res* 1989;32:912-20.
- Andreou LV, Kashino M, Chait M. The role of temporal regularity in auditory segregation. *Hear Res* 2011;280:228-35.
- Martin JG. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychol Rev* 1972;79:487-509.
- Patel AD. The OPERA hypothesis: assumptions and clarifications. *Ann N Y Acad Sci* 2012;1252:124-8.
- Cummins F. Joint speech: the missing link between speech and music? *Percepta* 2013;1:17-32.
- Bianchi F, Santurette S, Wendt D, Dau T. Pitch discrimination in musicians and non-musicians: effects of harmonic resolvability and processing effort. *J Assoc Res Otolaryngol* 2016;17:69-79.
- Kaplan EC, Wagner AE, Toffanin P, Başkent D. Do musicians and non-musicians differ in speech-on-speech processing? *Front Psychol* 2021;12:623787.
- McGarrigle R, Munro KJ, Dawes P, Stewart AJ, Moore DR, Barry JG, et al. Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper.' *Int J Audiol* 2014;53:433-45.
- Winn MB, Teece KH. Listening effort is not the same as speech intelligibility score. *Trends Hear* 2021;25:23312165211027688.
- Gagné JP, Besser J, Lemke U. Behavioral assessment of listening effort using a dual-task paradigm. *Trends Hear* 2017;21:2331216516687287.
- Winn MB, Wendt D, Koelewijn T, Kuchinsky SE. Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear* 2018;22:2331216518800869.
- Parbery-Clark A, Skoe E, Lam C, Kraus N. Musician enhancement for speech-in-noise. *Ear Hear* 2009;30:653-61.
- Priyanka VK, Krishna R. Exploring music induced auditory processing differences among vocalists, violinists and non-musicians. *Int J Health Sci Res* 2019;9:13-21.
- Amemane R, Gundmi A, Madikeri Mohan K. Effect of carnatic music listening training on speech in noise performance in adults. *J Audiol Otol* 2021;25:22-6.
- Devi N, Ajith Kumar U. Brainstem encoding of Indian carnatic music in individuals with and without musical aptitude: a frequency following response study. *Int J Health Sci Res* 2015;5:487-95.
- Menon MS, Thangaraj MS. Development of phonemically balanced word list in Tamil for speech audiometry and evaluation of its effectiveness in adults. *Indian J Otolaryngol Head Neck Surg* 2023 Sep 12 [Epub]. <https://doi.org/10.1007/s12070-023-04209-y>.
- Krishnamoorthy T, Vaidyanath R. Development and validation of Tamil matrix sentence test [dissertation]. Chennai: Sri Ramachandra Medical College and Research Institute;2018.
- Gnanasekar S, Vaidyanath R. Perception of Tamil mono-syllabic and bi-syllabic words in multi-talker speech babble by young adults with normal hearing. *J Audiol Otol* 2019;23:181-6.
- Zentner M, Strauss H. Assessing musical ability quickly and objectively: development and validation of the short-PROMS and the mini-PROMS. *Ann N Y Acad Sci* 2017;1400:33-45.
- Slater J, Kraus N, Carr KW, Tierney A, Azem A, Ashley R. Speech-in-noise perception is linked to rhythm production skills in adult per-

- cussionists and non-musicians. *Lang Cogn Neurosci* 2018;33:710-7.
- 30) Bangert M, Peschel T, Schlaug G, Rotte M, Drescher D, Hinrichs H, et al. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage* 2006;30:917-26.
- 31) Alain C, Du Y. Recruitment of the speech motor system in adverse listening conditions. *J Acoust Soc Am* 2015;137:2211.
- 32) Salvi RJ, Lockwood AH, Frisina RD, Coad ML, Wack DS, Frisina DR. PET imaging of the normal human auditory system: responses to speech in quiet and in background noise. *Hear Res* 2002;170:96-106.