Effect of Musical Training on Temporal Resolution Abilities and Speech Perception in Noise

[1]Anoop Oommen Thomas & [2]K. Rajalakshmi

Abstract

Many studies had reported that musical training will improve not only the ability to perceive musical aspects, but also different other factors like language processing, working memory, and also auditory perceptual abilities like temporal resolution and speech perception in noise. The present study aimed to find the effect of musical training on temporal resolution abilities and speech perception in noise. Total 20 trained Carnatic musicians, who were classified into 4 groups based on experience participated. Temporal Modulation Transfer Function (TMTF) and Gap Detection Threshold (GDT) were done for measuring temporal resolution and speech perception in noise (SPIN) were administered. The results revealed that TMTF and GDT showed significant difference across groups. But speech perception in noise was not significantly different across the four groups, though the scores were better as the experience increased.

Key words: *Carnatic music, temporal resolution, musical training*

Introduction

Music perception involves complex brain functions underlying acoustic analysis, auditory memory, and auditory scene analysis and processing of musical syntax. Moreover, music perception potentially affects emotion, influences autonomic nervous system, the hormonal and immune systems and activates (pre)motor representations.

Many studies have reported that musicians have better auditory perception skills when compared to non-musicians. There are many studies in literature which have documented that musical training improves basic auditory perceptual skills resulting in enhanced behavioral (Jeon & Fricke 1997; Koelsch, Schroger & Tervaniemi 1999; Oxenham, Fligor, Maso & Kidd, 2003; Tervaniemi, Just, Koelsch, Widmann, & Schorge, 2005; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Rammsayer & Altenmuller 2006) and neurophysiological responses (Shahin, Roberts & Pantev, 2007; Tervaniemi et al., 2005; Kuriki, Kanda & Hirata, 2006; Kraus, Skoe, Parbery-Clark & Ashley, 2009).

Musicians' life long experience of melodies from background harmonies can be considered as a process analogous to speech perception in noise. Studies report that musicians had a more robust sub- cortical representation of the acoustic stimulus in the presence of noise (Kraus et al., 2009). Musical practice not only enhances the processing of music related sounds but also influences processing of other domains such as language (Marques et al., 2007; Moreno et al., 2009; Parbery-Clark et al., 2009).

Musical training involves discrimination of pitch intonation, onset, offset and duration aspects of sound timing as well as the integration of multisensory cues to perceive and produce notes. Because of their musical training, musicians have learned to pay more attention to the details of the acoustic stimuli than non-musicians. (Musacchia, Sams, Skoe & Kraus, 2004).

The studies have documented better auditory perceptual skills in trained musicians when compared to non-musicians. But there are only very few studies which were done on the temporal resolution and speech perception abilities in trained musicians, as the experience increases in terms of years of training and practice. Recent study has shown that individuals with western instrumental musical training have enhanced speech perception ability in noise and working memory (Kraus et al., 2009). As a combined consequence of their extensive experience with auditory stream analysis within the context of music; more honed auditory perceptual skills and temporal resolution, musicians seem well equipped to cope with the demands of adverse listening situations such as speech in noise.

The aim of the present study was to find the temporal resolution abilities in trained Carnatic Vocal musicians over the years of musical training and practice. The speech perception abilities in the presence of

[1] Email: aot1985@gmail.com; [2]Reader in Audiology, Email: veenasrijaya@gmail.com

background noise at different signal-to-noise ratios (SNR) was also compared for the same group.

Method

Participants

A total of 20 trained Carnatic vocal musicians were included in the study and were classified into 4 groups, each group consisting of 5 members based on their years of experience in terms of training and practice. Group 1 had musicians who received training for 1-5.11 years. Group 2 included musicians who received training for 6-10.11 years. Group 3 with musicians having training received for 11-15.11 years. Finally, Group 4 consisted musicians who received training for 16 years and above.

All subjects should had normal air conduction and bone conduction thresholds (≤15 dBHL) at all octave frequencies from 250 Hz to 8000 Hz, normal middle ear function ('A' type tympanogram at 226 Hz probe tone with normal acoustic reflexes in both ears). Speech Recognition Threshold was within ±12 dB (re. PTA of 0.5, 1 and 2 KHz), speech identification scores of >90% at 40 dBSL (re. SRT) in both ears. Subjects had no indication of Retrocochlear Pathology (RCP), no history of neurological or Otological problems, no illness on the day of testing. All subjects were native Kannada speakers and had training in Carnatic music for duration of minimum 5 years.

Environment

All testing was carried out in a sound treated two room situation as per the standards of ANSI S3.1 (1991).

Instrumentation

A calibrated dual channel clinical audiometer (Orbiter 922) with TDH 39 headphones housed in MX-41/AR ear cushion was used for air conduction testing, Gap Detection Test (GDT) and for Speech perception in noise (SPIN) testing. A Radio ear B-71 bone vibrator was used for bone conduction testing.. A calibrated Immittance Meter (GSI Tympstar) was used to rule out middle ear problems.

Gap Detection Test was done with stimulus developed by Shivaprakash and Manjula (2003). It consists of 3 noise bursts, one of which contains a gap in it.

Recorded phonetically balanced (PB) word list in Kannada developed by Yathiraj and Vijayalakshmi (2005) was used for Speech Perception in Noise (SPIN) Test. It consists of 4 lists, each having 25 monosyllables.

Amplitude modulated white noise was used to find the Temporal Modulation Transfer Function (TMTF).

An output of PC which was routed to the audiometer was used to deliver stimulus for GDT, TMTF and SPIN

Procedure: Pure tone thresholds were obtained using bracketing method for both air conduction and bone conduction for the octave frequencies from 250 Hz to 8000 Hz.

Speech Identification Scores in quiet for both ears were obtained with Kannada monosyllables (Yathiraj & Vijayalakshmi, 2005) for both ears separately at 40 dB SL with reference to SRT. A total of 25 words were presented to each ear separately and each monosyllable was given a score of 4%.

Gap Detection Test (GDT) was done for both ears separately (Shivaprakash & Manjula, 2003) through three interval forced choice method. A total of 56 stimuli are present including 6 catch trials. Each stimulus consists of three noise bursts, one of which contains a gap of variable duration. Subject had to indicate verbally which of the set has a gap. The stimuset is presented monaurally at 40 dBSL (with reference to PTA) or at comfortable level. The minimum gap that the subject can identify was taken as the Gap Detection Threshold.

Temporal Modulation Transfer Function (TMTF) was assessed to determine the sensitivity to sinusoidally amplitude modulated broadband noise, as a function of modulation frequency (TMTF).Two test stimuli, unmodulated white noise of 500 ms duration and sinusoidal amplitude modulated white noise of 500 ms duration with a ramp of 20 ms, was used. The modulated white noise was derived by multiplying the broadband noise by a DC shifted sine wave. The depth of modulation was varied by changing the amplitude of modulating sine wave and modulation depth was varied between 0 to -30 dB (where 0 dB is equal to 100% modulation depth). Six different modulation frequencies were used (4 Hz, 8 Hz, 16 Hz, 32 Hz, 64 Hz & 128 Hz). All the stimuli were generated using a 32 bit digital to analogue converter at a sampling frequency of 44.1 kHz.

The stimuli were played manually using a PC, the output of which was routed to a calibrated audiometer. The stimulus was presented through a headphone. The level of presentation was at 40 dBSL (ref: PTA). The participants were required to discriminate between amplitude modulated noise and unmodulated noise.

On each trial, unmodulated and modulated stimuli was successively presented with an inter-stimulus interval of 500 ms. Modulation depth was converted into decibels [20 log 10 (m), where m refers to the depth of modulation]. A step size of 4 dB was used initially and then reduced to 2 dB after two reversals. This procedure provides an estimate of the value of amplitude modulation necessary for 70.7% estimate correct responses (Levitt, 1971). The mean of eight reversals in a block of 14 was considered as threshold.

Speech Perception in Noise test was done using the phonetically balanced (PB) Kannada word list recorded in female voice of a typical Kannada speaker. The monosyllables and the speech noise was presented monaurally at different SNR (0 dB, 10 dB and 20 dB). 25 monosyllables were presented for each trial and each monosyllable was given a score of 4%. Number of correctly identified monosyllables at different SNRs was noted down.

Results

The present study was aimed to compare temporal resolution abilities and speech perception in noise in Carnatic vocal musicians across their years of experience. The data was appropriately tabulated and statistically analyzed using SPSS (Version 18) software. Statistical analyses were carried-out to infer the findings of the present study.

Descriptive statistics (mean and standard deviation) were obtained for all the parameters for both ears separately. Kruskal-Wallis test was administered to compare the parameters across all the four groups. For the parameters which showed significant results under Kruskal-Wallis test, pair wise groups comparison was done with the help of Mann-Whitney test. Friedman test and Wilcoxon Signed Rank test (for pair-wise comparison) were done to compare the parameters within the group.

Temporal Resolution

Temporal Modulation Transfer Function (TMTF)

Table 1 depicts the mean and standard deviation of TMTF for the four groups at different modulation frequencies.

Temporal modulation transfer function was measured for 6 different modulation frequencies, for both the ears separately, for all the four groups. Table 1 shows the descriptive statistics (Mean & SD) of the TMTF of all the six modulation frequencies across the four groups.

It was observed from the mean data that the group with more than 16 years of musical experience (Group 4) showed better temporal modulation detection thresholds in both ears, when compared with other groups.

Kruskal-Wallis test was done for comparing across the four groups. It revealed no statistically significant difference for 4 Hz for both ears, 8 Hz for right ear, 16 Hz for both ears, 32 Hz for right ear, and 128 Hz for left ear. But statistically significant difference was present for 8 Hz for right ear ($\chi^2_{(3)}$=9.30, p<0.05), 32 Hz for right ear ($\chi^2_{(3)}$=11.18, p<0.05), 64 Hz for right ear ($\chi^2_{(3)}$=9.00, p< 0.05), and for left ear ($\chi^2_{(3)}$=7.94, p<0.05) and 128 Hz for right ear ($\chi^2_{(3)}$=9.73, p< 0.05).

The results of Kruskal-Wallis test revealed that there is statistically significant difference between the scores in at least any of the two groups. In order to find out which all groups are statistically different Mann-Whitney U test was administered. When the groups 1 and 2 & 3 and 4 were compared, no significant difference (p>0.05) was found for any of the modulation frequencies studied. However, when groups 1 and 3 were compared, there was statistically significant difference at 8 Hz in right ear (|Z|=2.12, p<0.05), 32 Hz in right ear (|Z|=2.00, p<0.05), 64 Hz in right ear (|Z|=2.65, p<0.05) and at 128 Hz in right ear (|Z|=1.79, p<0.05). For all other frequencies there was no statistically significant difference (p>0.05).

When groups 1 and 4 were compared, there was statistically significant difference at 32 Hz in right ear (|Z|=2.5, p<0.05), 64 Hz in right ear (|Z|=2.44, p<0.05) and in left ear (|Z|=2.51, p<0.05) and at 128 Hz in right ear (|Z|= 2.62, p< 0.05). Whereas, when Groups 2 and 3 were compared, the results revealed that only at 8 Hz in right ear was significantly different (|Z|=2.38, p<0.05). Statistically significant differences were not found for all the modulation frequencies in both ears, at 5% level of significance.

For the comparison of Groups 2 and 4, there was statistically significant difference at 8 Hz in both right (|Z|=2.02, p<0.05) and left ear (|Z|=2.27, p<0.05), at 32 Hz for right ear (|Z|=2.64, p<0.05) and at 128 Hz for right ear (|Z|=2.015, p<0.05). For all other frequencies there were no significant differences at 5% level of significance.

Within group comparison was done using Friedman test. Temporal modulation transfer function was compared across different frequencies. The results revealed that in Group 1 statistically significant difference at 5% level of significance was obtained for

Table 1: Mean and Standard Deviation (SD) of TMTF for the four groups at different modulation frequencies.

Modulation frequencies	Groups							
	1		2		3		4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
4 Hz (R)	-21.07	0.90	-21.34	2.50	-23.73	2.85	-22.27	3.64
4 Hz (L)	-22.93	3.58	-21.07	1.21	-22.00	2.87	-22.13	3.45
8 Hz (R)	-17.86	3.57	-17.73	3.48	-21.87	0.73	-21.33	1.63
8 Hz (L)	-20.00	3.09	-17.60	3.32	-12.00	18.74	-22.00	1.88
16 Hz (R)	-14.47	3.25	-14.33	3.00	-15.80	1.61	-16.40	3.82
16 Hz (L)	-14.32	4.43	-14.07	2.23	-15.87	2.56	-17.13	3.41
32 Hz (R)	-10.93	1.92	-11.34	1.25	-14.93	3.70	-16.40	2.43
32 Hz (L)	-13.47	2.96	-13.87	2.28	-14.33	3.74	-15.86	3.28
64 Hz (R)	-6.73	1.69	-10.53	2.72	-9.67	0.71	-11.20	1.88
64 Hz (L)	-7.67	2.17	-9.13	2.11	-9.93	2.24	-12.00	1.33
128 Hz (R)	-5.47	1.92	-7.27	1.57	-7.87	1.61	-9.33	1.03
128 Hz (L)	-7.53	1.30	-7.53	0.93	-7.73	1.28	-8.67	0.34

Table 2: Mean and SD of Gap Detection Threshold (GDT) for both ears.

Ear	Groups							
	1		2		3		4	
	Mean (ms)	SD	Mean (ms)	SD	Mean (ms)	SD	Mean (ms)	SD
Right	3.80	0.84	3.60	0.55	2.60	0.55	2.80	0.45
Left	3.60	0.90	3.60	0.55	3.00	0.00	2.60	0.55

8 Hz (|Z|=2.03, p<0.05) and 128 Hz (|Z|=2.03, p<0.05) only. For all other frequencies there were no statistically significant differences at 5 % level of significance. Groups 2, 3 and 4 showed no statistically significant differences at 5% level of significance, when frequencies were compared using Friedman test.

Gap Detection Threshold (GDT) test: was administered for both ears separately to find the minimum temporal gap, the subject could identify. GDT test was done for all the four groups. Mean and standard deviation (SD) of gap detection threshold for both the ears are shown in Table 4.2.

Descriptive statistical analysis showed that the gap detection threshold reduced as the musical experience increases. Group 1 was having a gap detection threshold of 3.8±0.87 for right ear and 3.6±0.89 for left ear, where as for group 4, the threshold was 2.8±0.45 and 2.6±0.55 respectively.

Kruskal-Wallis test was done to compare the thresholds across the four groups. For both right ear ($\chi^2_{(3)}$=9.27, p<0.05) and left ear ($\chi^2_{(3)}$=8.20, p<0.05), the results were statistically significant.

Mann-Whitney test was done to compare the GDT results across two groups. The results were statistically not significant in both ears (p>0.05) when Groups 1 and 2, and 3 and 4 were compared.

The thresholds were statistically significant only for right ear, when Groups 1 & 3, (/Z/=2.13, p<0.05); groups 1 & 4, (/Z/=2.00, p<0.05) and Groups 2 & 3, (/Z/=2.15, p<0.05). Whereas, when groups 2 and 4 were compared, the thresholds were statistically significant for both right ear (/Z/=2.03, p<0.05) and left ear (/Z/=2.15, p<0.05).

Within group comparison of gap detection thresholds were done using Friedman test and pair wise comparison was done using Wilcoxon signed rank test. The results revealed that there was no statistically significant difference at 5% level of significance in any of the groups. When gap detection thresholds were compared across right and left ears for all the groups using Wilcoxon signed rank test, there was no statistically significant difference at 5% level of significance.

Speech Perception in Noise

The speech perception in noise was assessed for all the 20 subjects for both the ears. The test was done at three signal-to noise ratios (SNRs): 0 dB SNR, -5 dB SNR and -10 dB SNR. The descriptive statistics (Mean & SD) of the speech perception in noise (SPIN) test for the three SNRs (0 dB, -5 dB & -10 dB) for both ears are shown in Table 4.3. The mean values showed that ability to perceive speech in the presence of the noise in all the three SNRs is better as the experience of the

Table 3: Mean and SD of speech perception in noise test scores at different SNRs for both ears.

Signal to Noise ratio (Ear)	Groups							
	1		2		3		4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0 dB (R)	92.80	3.35	93.60	2.19	95.20	1.79	93.60	2.19
0 dB (L)	90.40	4.56	94.40	2.19	93.60	2.19	94.40	2.19
-5 dB (R)	76.80	15.34	83.20	3.35	80.80	3.35	82.40	4.56
-5 dB (L)	76.80	12.46	81.60	2.19	80.80	1.79	82.40	2.19
-10 dB (R)	64.00	18.76	72.80	1.79	69.60	5.37	71.20	4.38
-10 dB (L)	64.80	17.75	72.80	3.35	70.40	5.37	69.60	4.56

musicians increased. It was found that as the experience of musician increased, the ability to perceive speech in the presence of background noise also increased, especially at lower SNRs.

The results across the four groups for three different SNRs were compared using Kruskal-Wallis test. The results revealed that there is no significant difference across the four groups at three different SNRs at 5% level of significance.

Within group comparison for three different SNRs (0 dB, -5 dB & -10 dB) were done using Friedman test. Pair wise comparisons were done using Wilcoxon signed rank test. The comparison of SNRs in the right ear showed statistically significant difference, ($\chi^2_{(2)}$=10.00, p<0.05). Wilcoxon signed rank test revealed statistically significant difference for all the three SNRs, at 5% level of significance. For left ear also, the three different SNRs were compared using Wilcoxon signed rank test. The results revealed statistically significant difference for the three SNRs, ($\chi^2_{(2)}$=10.00, p<0.05). From Wilcoxon Signed Rank test all the three SNRs were significantly different at 5% level of significance.

Discussion

The results, of TMTF across the four groups revealed statistically significant difference for the modulation frequencies like 8 Hz, 32 Hz, 64 Hz and 128 Hz, across different groups except for groups 1 & 2, and 3 & 4. The reason for no significant difference in these groups might be the closeness of these groups in terms of their experience. The literature which specifically explains about TMTF in musicians is limited. But in general, when random gap detection test was administered on musicians and nonmusicians, the gap detection thresholds were better in trained musicians when compared to non-musicians. This concludes that temporal resolution abilities are better in musicians when compared to non-musicians.

In the present study, for GDT there was no statistically significant difference when the groups compared were closer in terms of experience or practice (i.e., Groups 1 & 2; 3 & 4). But for other group comparison there was statistically significant difference in the gap detection thresholds at 5% level of significance. These results are in agreement with the study by Moreno et al., (2009), where it was concluded that musicians had better temporal resolution abilities when compared to non-musicians and the years of experience was a factor in deciding about the temporal resolution ability. As the experience in music increased, better temporal resolution ability was observed. Studies also reported that initiation of musical training also matters for the better abilities. According to Ohnishi et al., (2001), music training can induce functional reorganization of the cerebral cortex. Therefore, the contact with music before the age of seven could contribute to the development of primary auditory cortex and more precisely the planum temporale. When the GDT was compared between the two ears within the group, there was no statistical significant difference at 5% level of significance.

When the SPIN results were compared across the groups, there was no statistically significant difference at *p*=>0.05, for all the three SNRs (0 dB, -5 dB, & -10 dB). But this is in contrast to the previous research done in speech perception abilities in musicians. According to a study done by Parbery-Clark et al (2009b), musical experience enhances the ability to hear speech in challenging listening environments. In another study Parbery-Clark et al., (2009) found that musical experience resulted in more robust subcortical representation of speech in the presence of background noise. The difference in the results of the present study with the earlier studies reported in the literature can be accounted for a few reasons. First, the noise used in the previous studies were speech shaped noise or multi-talker babble. But in the present study, speech noise was used to study the speech perception in noise. It is evident that the speech shaped noise or multi-talker

babble will give better results for speech perception in noise when compared to speech noise. Second, the previous studies were conducted on instrumental musicians, whereas the present study was carried out in vocal musicians. Moreover, the subjects taken in Parbery-Clark et al., (2009b) study were having more experience than the subjects for the present study.

When within group comparison was done for each ear at three different SNRs there was a reduction in the speech identification scores for all the subjects as the SNRs decreased which was statistically significant at 5% level of significance. This means that when the noise level increased there was difficulty in the perception of speech.

Conclusions

The results from the present study showed that the temporal resolution abilities and the ability to perceive speech in the presence of noise were better in musicians than in non-musicians. The results of temporal modulation transfer function and gap detection threshold values showed that the temporal resolution abilities becomes better as the years of musical experience of the musicians increased. The results were statistically significant. But the results of the speech perception in noise were not statistically significant when the musicians were compared across their experience, though the scores were better in experienced musicians when compared to the musicians with less experience.

References

American National Standards Institute (1991). *Maximum Ambient Noise Levels for Audiometric Test Rooms.* (ANSI S3. 1-1991). New York: American National Standards Institute.

Jeon, J. Y., & Fricke, F. R. (1997). Duration of perceived and performed sounds. *Psychology of Music, 25*, 70–83.

Koelsch, S., Schroger, E., & Tervaniemi, M. (1999). Superior attentive and pre-attentive auditory processing in musicians. *Neuroreport, 10*, 1309– 1313.

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Training-induced malleability in neural encoding of pitch, timbre, and timing: implications for language and music. *Annals of NY Academy of Science. Neuroscience. Music III* 1169, 543-557.

Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *Journal of Neuroscience, 26*, 4046–4053.

Levitt, H., (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467-477.

Marques, C., Moreno, S., & Casstro, S. L. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioural and electrophysiological evidence. *Journal of Cognition and Neuroscience, 19*, 1453–1463

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psycho acoustical training on pitch discrimination. *Hearing Research, 219*, 36–47.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S.L., & Besson, M., (2009). Musical training influences linguistic abilities in 8-year old children: more evidence for brain plasticity. *Cerebral Cortex, 19*, 712-723.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N., (2004) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of National Academy of Sciences 104*, 15894-15898.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., & Nishikawa, M. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex, 11(8)*, 754-760

Oxenham, A. J., Fligor, B. J., Mason, C. R.., & Kidd, G. (2003). Informational masking and musical training. *Journal of the Acoustical Society of America, 114*, 1543–1549.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., Almut., Ross.,& Bernhard (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport, 12*, 169–174.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009a). Musician enhancement for speech in noise. *Ear and Hearing. 30*, 653-661.

Parbery-Clark, A., Skoe, E., Lam, C.,& Kraus, N, (2009b). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience, 29*, 14100-14107.

Schon, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*, 341–349.

Shahin, A. J., Roberts, L. E., & Pantev, C. (2007). Enhanced anterior-temporal processing for complex tones in musicians. *Clinical Neurophysiology, 118*, 209–220.

Shivaprakash. S & Manjula. P. (2003). *Gap Detection Test-Development of Norms*, Unpublished Master's Independent Project, University of Mysore.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schorger, E., (2005). Pitch discrimination accuracy in musicians vs. non- musicians: an event-related potential and behavioural study. *Experiments in Brain Research,. 161*, 1- 10.

Yathiraj, A., & Vijayalakshmi, C. S. (2005). *Phonemically Balanced word list in Kannada*. Developed in Department of Audiology, AIISH, Mysore.