Auditory Plasticity in Musicians: A Comparative Study

[1]Deepika Verma & [2]Rajalakshmi K.

Abstract

Musical training is a rigorous routine involving the segregation of vocal and instrumental sounds presented concurrently. To investigate the effect of musical experience on the neural representation of speech-in-noise, we compared behavioral scores of speech perception in noise and sub cortical neurophysiological responses to speech in a group of highly trained musicians and non-musician controls. A total of 50 participants comprised musicians and non musicians were taken, aged between 7-18 years. Speech evoked ABR and LLR were done along with SPIN test. Musicians were found to have a more robust sub cortical representation of the speech stimulus. Musicians had earlier response onset timing, than non-musicians Fo encoding is better in musicians than in non musicians. Speech perception ability in musicians becomes better with increased years of musical exposure and experience. The increase in Fo amplitude is positively correlated with the speech perception in noise.Musicians had better behavioral performance on the Speech in Noise Test (SPIN) outperforming the non-musician controls. These findings suggest that musical experience limits the negative effects of competing background noise.

Keywords: *Plasticity, Musicians, Speech perception in noise, Speech evoked ABR*

Introduction

From the cochlea to the auditory cortex, sound is encoded at multiple locations along the ascending auditory pathway, eventually leading to conscious perception. Speech is a stream of acoustic elements produced at an astounding average rate of three to six syllables per second (Laver, 1994).The ability to decode these elements in a meaningful manner is a complex task that involves multiple stages of neural processing.

Neural plasticity is a term used to describe alterations in the physiological and anatomical properties of neurons in the brain in association with auditory stimulation and deprivation. Depending on the experience, mechanism of plasticity can involve synaptic changes that occur rapidly or slowly over a period of time (Tremblay & Kraus, 2002). Everyday learning and training involves of continuous improvement of our abilities the sensory, cognitive & behavioural levels (Menning, Roberts & Pantev, 2000). Peripheral and central structures along the auditory pathway contribute to speech processing and learning. However, because speech requires the use of functionally and acoustically complex sounds which necessitates high sensory and cognitive demands, long-term exposure and experience using these sounds is often attributed to the neocortex with little emphasis placed on subcortical structures (Song, Skoe, Wong & Kraus, 2008).

Music is a complex auditory task and musicians spend years fine-tuning their skills. It is no wonder that previous research has documented neuroplasticity to musical sounds as a function of experience (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005; Koelsch, Schroger & Tervaniemi, 1999; Musacchia, Sams, Skoe & Kraus, 2007; Pantev et al., 1998; Pantev, Roberts, Schulz, Engelien, & Ross, 2001; Tervaniemi, Rytkonen, Schroger, Ilmoniemi & Naatanen, 2001). The domains of music and language share many features, the most direct being that both exploit changes in pitch patterns to convey information. Music uses pitch contours and intervals to communicate melodies and tone centers. Pitch patterns in speech convey prosodic information; listeners use prosodic cues to identify indexical information, i.e., information about the speaker's intention as well as emotion and other social factors.

Musicians have a variety of perceptual and cortical specializations compared to non-musicians. Recent studies have shown that potentials evoked from primarily brainstem structures are enhanced in musicians, compared to non-musicians. Specifically, musicians have more robust representations of pitch periodicity and faster neural timing to sound onset when listening to sounds or both listening to and viewing a speaker. However, it is not known whether musician-related enhancements at the subcortical level are correlated with specializations in the cortex (Musacchia, Strait & Kraus, 2008). The effects of musical experience on the nervous system include relationships between brainstem and cortical Evoked Potentials recorded simultaneously in the same subject to seen and heard speech. Moreover, these relationships were related to behavioural measures of auditory perception and were stronger in the audiovisual condition. This implies that musical training promotes plasticity throughout the auditory and multisensory pathways. This includes encoding mechanisms that are relevant for musical sounds as well as for the processing of linguistic cues and multisensory information (Musacchia et al., 2008).

Hearing speech in noise is a difficult task for everyone, but young children and older adults are particularly vulnerable to the deleterious effects of background noise. Children with learning disorders can ex-

[1]Email: verma.deepika28@gmail.com,
[2]Professor of Audiology, Email: veenasrijaya@gmail.com

hibit noise exclusion as a primary symptom (Sperling, Lu, Manis & Seidenberg, 2005). Musicians, in contrast, demonstrate enhanced noise-exclusion abilities (Parbery-Clark, Skoe & Kraus, 2009a; Parbery-Clark, Skoe, Lam & Kraus, 2009b). Musical experience enhances the ability to hear speech in challenging listening environments. Speech in Noise performance is a complex task requiring perceptual cue detection, stream segmentation, and working memory. Musicians performed better than nonmusicians in conditions where the target and the background noise were presented from the same source, meaning parsing was more reliant on the acoustic cues present in the stream (Parbery & Kraus, 2009).

There is evidence of musical expertise contributing to an enhanced subcortical representation of speech sounds in noise. Musicians had more robust temporal and spectral encoding of the eliciting speech stimulus, thus offsetting the deleterious effects of background noise. Faster neural timing and enhanced harmonic encoding in musicians suggests that musical experience confers an advantage resulting in more precise neural synchrony in the auditory system. These findings provide a biological explanation for musicians' perceptual enhancement for speech-in-noise (Anderson & Kraus, 2010).

Thus, this study was taken up to study the brainstem correlation of speech in noise perception in musicians and to document the auditory plasticity induced by music in musicians on the basis of experience in Carnatic music and also to compare the brainstem and cortical plasticity in musicians and non musicians.

Method

The present study aimed to find out the effect of musical training on auditory plasticity and speech perception in noise in musicians with various years of Carnatic vocal musical training or practice, using Speech evoked Auditory Brainstem Response, Speech evoked Late Latency Response, and Speech Perception in Noise (SPIN) tests.

Participants

A total of 50 participants aged between 7 to 18 years. Twenty children enrolled for Carnatic music learning and 25 untrained children were included in this study. The musicians were classified in to 3 groups.Twenty five trained musicians were classified into 4 groups.

Group 1: Music learning age ranging from 7 to 10 yrs with minimum experience of 2 to 3 years (5 participants).

Group 2: Music learning age ranging from 10 to 13 yrs with minimum experience of 4 to 5 years (10 participants).

Group 3: Music learning age ranging from 13 to 18 yrs with experience of greater than 6 years (10 participants).

Test Stimulus

The /da/ stimulus is a 40 ms synthesized speech syllable produced using KLATT synthesizer (Klatt, 1980) which is available in the Biologic Navigator Pro EP system in the BIOMARK protocol. This stimulus simultaneously contains broad spectral and fast temporal information characteristic of stop consonants, and spectrally rich formant transitions between the consonant and the steady-state vowel. The fundamental frequency (F0) of the /da/ stimulus linearly rises from 103 to 125 Hz with voicing beginning at 5 ms and an onset noise burst during the first 10 ms. The first formant (F1) rises from 220 to 720 Hz, while the second formant (F2) decreases from 1700 to 1240 Hz over the duration of the stimulus. The third formant (F3) falls slightly from 2580 to 2500 Hz, while the fourth (F4) and fifth formants (F5) remain constant at 3600 and 4500 Hz, respectively.

The stimulus used was the default BIOMARK synthetic /da/ syllable of 40 ms duration produced using KLATT synthesizer (Klatt, 1980). The stimuli were presented at a rate of 10.9/s through ER-3 insert earphones in alternating polarities at an intensity of 80 dB SPL with a filter setting of 100 to 3000 Hz. The responses were collected using three AgCl scalp electrodes. Responses were differentially recorded over a time window of 64 ms (including pre-stimulus period of 11ms) with a vertical montage (Test Ear Mastoid-Active, Forehead-Ground, and Non-Test Ear Mastoid- Reference). The waveforms of the participants, acquired after 1500 artifact free sweeps were weighted added for each ear and then analyzed using the BIOMARK module .Waveforms were collected for rarefaction & condensation polarities and weighted addition was done to obtain calculated waveforms. The speech evoked ABR and FFR waveform, were converted into ASCII format using the software called 'AEP TO ASCII'.

Test Procedure

Routine audiological evaluation was carried out in an acoustically treated room. Air conduction and bone conduction thresholds were established using modified Hughsen and Westlake (Carhart & Jerger, 1959) procedure. Speech audiometry was also carried out on all the participants. A calibrated two-channel Madsen (Orbiter-922) audiometer with TDH-39 headphones was used to establish air conduction pure tone thresholds and speech audiometry. B-71 bone vibrator was used to establish bone conduction thresholds. Hearing was considered normal if puretone thresholds were within 15 dB HL bilaterally at all octave frequencies from 250 Hz to 8 kHz. Tympanometry and Acoustic Reflex Thresholds were established using a calibrated Grason Stadler-Tympstar middle ear analyzer. Presence of 'A' or 'As' type of tympanogram with reflexes present in both the ears below 100 dB HL at 500 Hz, 1 kHz and 2 kHz was considered as normal.

Speech Perception in Noise (SPIN): SPIN test was done using the phonemically balanced (PB) Kannada word list (Yathiraj & Vijayalakshmi, 2005), recorded in the voice of a typical female Kannada speaker. The stimuli were played in a laptop and were routed through the audiometer. The presentation level was 40 dB SL (with reference to SRT) or at most comfortable level. The monosyllables and the speech noise were presented monaurally at two different SNRs (0dB & -5 dB). Twenty five monosyllables were presented for each trial. The participants' task was to perceive the monosyllables presented in the presence of noise and repeat them back. Each word was given a score of 4 %. Number of correctly identified word at different SNRs was noted down to find the SPIN score.

Speech Evoked LLR: Biologic Navigator Pro EP System version 7.0 was used for recording speech evoked late latency responses. LLR was assessed for P1, N1, and N2 in terms of latency.

Speech Evoked Auditory Brainstem Response: Biologic Navigator Pro EP System version 7.0 was used for recording speech evoked auditory brainstem response.

Data Analysis

Speech evoked ABR is composed of the transient and the sustained responses (also known as frequency following responses). Transient response consists of peak V and peak A whereas the sustained responses consist of peaks D, E, F, and O. In the present study latency of both the transient as well as sustained responses were analyzed.

The transient response was analyzed in terms of latency and amplitude of V and A peak. The FFR response was analyzed in terms of latency and amplitude of D, E, F, O peaks for the earlier mentioned three repetition rates (the distance between the peak D, E, F, and O is approximately 10 ms which gives the information regarding the encoding of fundamental frequency). The sustained portion was analyzed using Fast Fourier Transformation (FFT) for the latency range of 11.4 ms to 40.6 ms for speech evoked ABR to extract the information regarding the coding of fundamental frequency, first formant frequency and second formant frequency at different repetition rates using the MATLAB software.

Procedure for FFT analysis: To know the coding of fundamental frequency, first formant frequency and higher harmonics, FFT analysis of the sustained response of the speech evoked ABR was done. This was executed using the MATLAB version 7.0 software (Brainstem toolbox) developed by Skoe and Kraus (2004), at North western university. For measuring the fundamental frequency and higher harmonics, Fourier analysis was performed on the 11.4 to 40.6 ms epoch of the FFR in order to assess the amount of activity occurring over three frequency ranges. Activity occurring in the frequency range of the response corresponding to the fundamental frequency of the speech stimulus (103 to 121 Hz), first formant frequencies of the stimulus (454 to 719 Hz) and for the higher harmonics (721 to 1155 Hz) were measured for all the participants.

Results and Discussion

The data was appropriately tabulated and statistically analyzed using SPSS (version 18) software. Descriptive statistics (mean and standard deviation) were obtained for all the parameters for both ears separately. Separate 2-WAY MANOVA was done to see the significant difference between musician and non-musicians for all the parameter for speech ABR (Latency of wave V, A, C, D, E, F, O), LLR (Latency of P1, N1, P2) and for amplitudes of Fo and F1. Two-way ANOVA was done to see the significant difference between musicians & non-musicians for SPIN scores. Pearson correlation was calculated to see the correlation between SPIN scores and the amplitudes of Fo & F1 in musicians & non-musicians.

Speech Evoked ABR

Latency of Waves V, A, C, D, E, F and O was measured for non-musicians & musicians, for both ears separately, for all three groups. Descriptive statistics (mean & SD) was done for latencies of speech evoked ABR waves for musicians & non-musicians across three groups.

Two-way MANOVA was done to see the differences between musicians and non-musicians for latencies of speech evoked ABR. There was a significant difference between non-musicians & musicians ($p<0.05$) for latency of wave V & O for all three groups. There was no significant difference present for latencies of other waves between musicians and non-musicians.

In the present study, the latency of wave V and O responses were significantly different between musicians and non-musicians. There was no significant difference present in transition latencies between musicians and non-musicians. These results are in agreement with the study by Parbery-Clark et al. (2009a), where it was concluded that musicians had earlier response onset timing, than non-musicians. Musacchia et al. (2008) reported that latency and amplitude of wave V differed between musicians and non-musicians. The results of the study by Parbery-Clark et al. (2009a) also suggested that there was no significant difference in transition latencies between musician and non musician in quiet conditions.

Table 1: Mean and SD for FFT- Fast Fourier Transform; Fo, F1 & F2- Fundamental frequency, first and second Formants for non-musicians

Wave	Age Group (in years)	Mean (dB)	Standard Deviation
	7 to10	2.076	0.645
Fo	10 to13	3.379	1.308
	13 to18	4.062	1.204
	7 to10	1.412	0.368
F1	10 to13	1.286	0.343
	13 to18	1.252	0.434
	7 to10	.5392	0.095
F2	10 to13	0.447	0.115
	13 to18	0.460	0.162

Musicians and non-musicians had equivalent stimulus response correlation in quiet. Their results are in agreement with the results of the present study.

Speech Evoked Late Latency Response

Latency of Waves P1, N1& P2 was measured for non-musicians & musicians, for both ears separately. Descriptive statistics (mean & SD) was done for latencies of speech evoked LLR waves across all age groups.

Two-way MANOVA was done to see the differences between musicians and non-musicians for latencies of speech evoked LLR waves across years of musical experience There is no significant difference between musicians & non-musicians in terms of latencies of wave P1, N1 and P2 of speech evoked LLR ($p< 0.05$).The results of the present study are in consonance with the results by Strait et al. (2011). There was no response variability among musicians and non-musicians at any electrode site.

Shahin et al. (2003) reported enhanced P2 and N1c responses in musicians compared to non-musicians. Krista et al. (2009) reported that long term music training offers structural plasticity in developing correlation with behavioural changes. T1 weighted MRI was used in their study for assessment.

The difference in the results of the present study with the earlier studies reported in the literature can be accounted on the following reason: First, the assessing tool used in previous studies is magneto encephalography (MEG), electroencephalography (EEG) and MRI. These radiological tests are different from far field electrophysiological responses. Second, the previous studies were conducted on instrumental musicians, whereas the present study was carried out on Carnatic vocal musicians. Moreover, the participants taken in Shahin et al. (2003) study were having more years of musical experience (greater than 11 years) than the participants of the present study.

Table 2: Mean and SD for FFT- Fast Fourier Transform; F0, F1 & F2- Fundamental frequency, first and second formants for musicians

Wave	Age Group (in years)	Mean (dB)	Standard Deviation
	7 to10	2.303	0.669
Fo	10 to 13	4.975	0.971
	13 to 18	5.554	1.196
	7 to 10	1.326	0.556
F1	10 to 13	1.518	0.475
	13 to 18	1.389	0.381
	7 to 10	0.482	0.153
F2	10 to 13	0.523	0.136
	13 to 18	0.527	0.149

FFT- Fast Fourier Transform: F0, F1 and F2

Fo, F1 and F2- Fundamental frequency, first and second formants were measured for non-musicians & musicians, for both ears separately, for all three groups. Table 1 and Table 2 show descriptive statistics (mean & SD) for FFT- Fast Fourier Transform; Fo, F1 and F2-Fundamental frequency, first and second formants for non-musicians and musicians.

Two-way MANOVA was done to see the differences between musicians and non-musicians for FFT- Fast Fourier Transform; F0, F1 & F2- Fundamental frequency, first and second Formants (Table 3).

There was a significant difference present ($p< 0.05$) for Fo Formant between musicians and non-musicians for group 3. The amplitude of energy concentration in Fo formant is significantly larger in Group 3 compared to non-musician. The results of the present study are in agreement with results of Mussachia et al. (2007). They reported that musicians have larger response amplitudes for encoding of speech and music stimuli compared to non-musicians. Mussachia et al. (2008) reported experienced musicians had larger Fo peak amplitudes. In the present study, musicians in Group 3, with higher years of musical experience had better mean SPIN scores than non-musicians. This result draws support study by Anderson et al (2010), which suggests that good SPIN perceivers had greater spectral magnitudes for Fo and H2.

Table 3: Statistical Results for Two way Manova for FFT

Source	Variable	df	F	p
	Fo	2	2.910	0.000*
grpNM*M	F1	2	0.934	0.397
	F2	2	1.689	0.190

Table 4: Mean and SD for SPIN for all groups

Group	NM/M	Mean (%)	Standard Deviation
1	Non musician	72.80	3.676
	Musician	73.20	4.237
2	Non musician	77.00	4.657
	Musician	80.00	4.768
3	Non musician	80.80	3.488
	Musician	86.00	2.865

The difference in the results for groups 1 and 2 of the present study can be explained on the following reasons; first, the musical year of experience in group 1 and 2 were less than reported in the previous studies. Second, the previous studies were conducted on instrumental musicians, whereas the present study was carried out on vocal musicians. Moreover, the participants taken in Mussachia et al. (2008) study were having more experience than the participants in the present study. Most of the study reports experience of 10 for their participants.

Speech Perception in Noise

The speech perception in noise was assessed for all the 50 participants for both the ears. The test was carried out at 0 dB SNR.

2-WAY ANOVA was employed to see the significant difference between musicians & non-musicians. There was a significant difference across the three groups *($p< 0.05$)*. However, there was no significant difference between musicians & non-musicians across the different groups for SPIN.

On comparing mean score (Table 4) the mean scores are quite better for musicians in group 3, but not significantly better. But this is in contrast to the previous research done on speech perception abilities in musicians. According to a study done by Parbery-Clark et al. (2009a), musical experience enhances the ability to hear speech in challenging listening environments. In another study Parbery-Clark et al. (2009b) 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 on the following 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 on vocal musicians. Moreover, the participants taken in Parbery-Clark et al. (2009a) study were having more experience than the participants in the present study. Third, the speech material used in previous studies was sentences (Quick SIN, HINT). The sentences are more redundant than words. Fourth, the present study was conducted at 0 dB SNR, the studies reported in literature suggested that SPIN is better in adverse listening conditions. Parbery-Clark et al. (2009b) reported musicians were able to repeat sentences presented at a lower, more challenging SNR than non-musicians.

Correlation between SPIN & FFT

The results revealed that there is a positive and highly significant correlation between SPIN and Fo. With the increase in Fo amplitude, there is an increase in the scores for the perception of speech in noise. This finding can be supported by study of Anderson et al. (2010), good SIN perceivers have greater spectral magnitudes for Fo and H2.

Table 5: Correlation between FFT and Speech Perception in Noise

Formants	SPIN
Fo	0.000**
F1	0.826
F2	0.914

There was no significant correlation between other formants & SPIN scores. As the Formants increases, the amplitude of the harmonics decreases. Thus the results of the present study reveal that latencies of wave V and O were significantly different between musicians and non musicians.

The latencies of speech evoked LLR was did not reveal statistically significant difference for musician and non musician. A significant difference was noticed between musician and non musician for Fo formant. There is a significant difference across the three groups of musicians on SPIN score. A highly significant positive correlation is present between Fo and SPIN scores.

Conclusions

Fo encoding is better in musicians than in non musicians. Speech perception ability in musicians becomes better with increased years of musical exposure and experience. The increase in Fo amplitude is positively correlated with the speech perception in noise. There was no significant difference in latencies of P1, N1 and N2 between musicians and non musicians for speech evoked LLR.

It can be implemented in hearing aid technology for musicians with hearing loss to improve their speech perception. Music training can be used as a potential remediation strategy for children requiring language training and auditory processing disorders with noise exclusion deficits.

Future research on clinical population who may exhibit neural encoding deficits such as autism. Brainstem maturation as an indication in infants & preschool children at risk.

The present study can b replicated across vocal musicians and instrumental musicians. It can be compared between Hindustani and Carnatic musicians. Musicians and dancers can be compared to find whether there are differences in Fo encoding and ability to perceive speech in the presence of noise.

References

Anderson, S., & Kraus. N. (2010). Sensory-cognitive interaction in the neural encoding of speech in noise: A review (2010). *Journal of American Academy of Audiology, 21*, 575-585.

Carhart, R., & Jerger, J. (1959). Preferred method for clinical determination of pure-tone thresholds.*Journal of Speech and Hearing Disorders, 24,* 330-345.

Fujioka, T., Trainor, L., Ross, B., & Kakigi, R. (2004). Musical Training Enhances Automatic encoding of Melodic Contour and Interval Structure. *Journal of cognitive Neuroscience, 16*, 1010-1021.

Klatt, D. (1980). Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America, 67,* 13-33.

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

Laver, J. (1994). *Principles of phonetics.* Cambridge, UK: Cambridge University Press.

Menning, H., Roberts, L., & Pantev, C., (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Auditory and Vestibular Systems, 11*, 817- 822.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of United States of America, 104*, 15894-15898. Musacchia, G., Strait, D., & Kraus, N. (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research, 241,* 34-42.

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). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience, 29,* 14100-14107.

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

Shahin , A., Bosnyak , D., Trainor, L., Roberts, & Larrey, R., (2003). Enhancement of neurolplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience, 12*, 5545-5552.

Skoe, E., & Kraus, N. (2010). Auditory brain stem response to complex sounds: A tutorial. *Ear and Hearing, 31*(3), 302-324.

Song, J, H., Skoe, E, Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognition Neuroscience, 20,* 1892-1902.

Sperling, A. J., Lu, Z., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience, 8*, 862-863.

Tervaniemi, M., Rytkonen, M., Schroger, E., Ilmoniemi, R. J, & Naatanen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning and Memory, 8*, 295-300.

Tremblay, K. L., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortica; neural activity. *Journal of Speech-Language-Hearing Research, 43*(3), 564-72.

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