**PROJECT PROPOSAL FORMAT**

FORM – A

|  |
| --- |
| **Title of the Project : ‘***Effect of musical experience on formant frequency discrimination’* |
|  |
| **Area of Research :** Hearing/Audiology |
| **Principal Investigator and** |
| **Principal Co-Investigator(s)**   |  |  |  |  | | --- | --- | --- | --- | | **Sl. no** | **Name of the investigators** | **Qualification** | **Designation** | |  | Dr. K. Rajalakshmi | M.Sc. (Speech & Hearing)  Ph.D. (Speech & Hearing) | Professor of Audiology, AIISH | |  | Ms. Spoorthi T | M.Sc. (Audiology) | Research Consultant, Research Project funded by Lamar University, Texas, USA | |

**EFFECT OF MUSICAL EXPERIENCE ON FORMANT FREQUENCY DISCRIMINATION**

Music is present in every culture and common to all people; because certain emotions are communicated directly by music. Music has always been an important part of Indian life. Indian classical music has also been significantly influenced by Indian folk music. Indian Music is probably the most complex musical system in the world with a very highly developed melodic and rhythmic structure. Carnatic music or Carnatic sangeet is the south Indian classical music. Carnatic music has a rich history and tradition and is one of the gems of world music.

Musicians develop a number of auditory skills, including the ability to track a single instrument embedded within a multitude of sounds. This skill relies on the perceptual separation of concurrent sounds that can overlap in pitch but differ in timbre. One contributor to timbre is a sound’s spectral fine structure: the amplitudes of different harmonics and how they change over time (Caclin et al., 2005, 2006; Kong et al., 2011). Consequently, it is not surprising that musicians, compared to non-musicians, demonstrate enhanced perception of rapid spectro-temporal changes (Gaab et al., 2005) and harmonic differences (Musacchia et al., 2008; Zendel& Alain, 2009), as well as having a greater neural representation of harmonics (Shahin et al., 2005; Lee et al., 2009; Parbery-Clark et al., 2009; Strait et al., 2009; Zendel and Alain, 2009). Harmonic amplitudes are one of the most important sources of information for distinguishing speech sounds. During speech production, individuals modify the filter characteristics of the vocal apparatus to enhance or attenuate specific frequencies, creating spectral peaks (i.e., formants) that distinguish vowels and consonants. Given the similarity of the acoustic cues that characterize speech and music—that is, differences in the distribution of energy across the harmonic spectrum—musicians’ extensive experience distinguishing musical sounds may provide advantages for processing speech (Tallal & Gaab, 2006).

The musicians have heightened sensitivity to small harmonic differences (Musacchia et al., 2008; Zendel & Alain, 2009) as well as a greater neural representation of harmonics than non-musicians (Shahin et al., 2005; Musacchia et al., 2008; Lee et al., 2009; Strait et al., 2009; Zendel and Alain, 2009; Parbery-Clark et al., 2009a).

A variety of studies have shown that the benefits of the musical training is not only limited to musical domain, but it also helps in processing of speech stimuli (Musacchia et al., 2007; Schon et al., 2004; Wong et al., 2007), working memory (Chan et al., 1998; Forgeard et al., 2008; Parbery-Clark et al., 2009), processing of prosody and linguistic features in speech (Bidelman, Gandour, & Krishnan, in press; Chandrasekaran, Krishnan, & Gandour, 2009; Wong etal., 2007), phonological skills (Forgeard et al.,2008), processing emotion in speech (Strait, Kraus,Skoe,& Ashley, 2009a), auditory attention (Strait, Kraus, Parbery-Clark, & Ashley, 2010), and auditory stream segregation (Beauvois & Meddis, 1997).

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. Further, in tone languages, changes in pitch are used lexically; that is, in differentiating between words. A significant body of research has focused on the extent to which musical experience provides benefits in language abilities; the results unambiguously suggest that musicians show enhanced processing of prosodic and linguistic pitch. Musicians show an enhanced ability to detect subtle incongruity in prosodic pitch as well as consistent neural differences relative to non-musicians (Besson, Schon, Moreno, Santos, & Magne, 2007; Magne, Schon, & Besson, 2006). Differences between musicians and non-musicians show up even during preattentive stages of auditory processing (Chandrasekaran, et al., 2009; Musacchia et al., 2007; Wong & Perrachione, 2007).

Perception of speech in the presence of background noise is the most difficult task faced by everyone. This devastating effect of background noise is more problematic especially for younger children and older adults. In contrast, musicians demonstrate an enhanced ability to exclude the noise (Parbery-Clark et al., 2009). Musical training requires consistent practice, online manipulation and monitoring of their instruments and as a consequence of these musicians become experts in extracting the relevant information from the complex soundscape (e.g., sound of their own instrument in an orchestra). The musicians have the ability to parse melodies from background harmonics. This ability can be considered to be similar to speech perception in noise. Perception of speech in the noisy situation requires the separation of the signal from competing signal.

The musicians have such robust neural representation of sound that sub-cortical differentiation of speech sounds can be clearly seen in an average comprised of only 700 trials. It is expected that non-musicians would also demonstrate differentiated neural responses but only with a greater number of trials, as reported in non-musician children (Skoe et al., 2011). Musical training results in auditory perceptual advantages in musicians and these advantages are supported by functional and structural changes seen both cortically and sub cortically for the processing of sound and specifically for processing speech in noise.

Parbery-Clark et al. (2009) measured the speech perception in noise using 2 standardized tests (QuickSIN and HINT, Hearing In Noise Test) and found that musicians exhibit more robust speech-evoked auditory brainstem responses in background noise and thus have distinct advantage for speech perception in the presence of noise. Addition of background noise delays the timing of auditory brainstem response in both musicians and non-musicians, but musician’s exhibit smaller delays in timing than non-musicians in noise. In the presence of the noise, amplitude of onset response is greatly reduced or eliminated in both the groups (Russo et al., 2004). The transition response was more robust to the effect of background noise than the onset peak amplitude. Neural representation of the stimulus harmonics is greater for musician than non-musicians in presence of noise. That is sub-cortical representation of stimulus temporal feature was equivalent for musicians and non-musicians in quiet, but musician’s responses were less degraded by the presence of noise.

Study done by Parbery-Clark et al. (2009) found that musical training modulates the effect of the noise on sub-cortical auditory representation as studied using speech evoked ABR. Speech evoked ABR is being used as a tool which provides reliable information about the speech coding at the brainstem level. Frequency-following responses (FFRs), which ensemble neural responses originating at the auditory brainstem that reflect phase-locking to stimulus features, were recorded from musicians and non-musicians who were listening to the speech syllable /da/ (Musacchia etal., 2007). Relative to non-musicians, musicians showed more robust encoding of timing and pitch features in the speech signal at the level of the brainstem. Using FFR as an index, musicians showed a superior representation of dynamic pitch contours, as reflected by improved pitch tracking accuracy at the level of the brainstem (Wong et al., 2007).

Speech evoked ABR is used to characterize neural encoding of speech sounds for clinical and research application.

**NEED FOR THE STUDY:**

Learning music is an extremely complex task that involves the interaction of several modalities and higher-order cognitive functions and that results in behavioural, structural and functional changes on time scales ranging from days to years (Herholtz & Zatorre, 2012).

Speech perception is highly dependent on the perception of the slight changes present in the formant frequencies of consonants and vowels. Daily listening conditions are rarely devoid of background noise. The identification of the formant frequencies becomes more strenuous when the speech is in presence of background noise. Musicians are found to be more expert in differentiating the minute changes in the pitch, rhythm, timbre. The musicians have an enhanced response specifically to the timbre of their own instrument (Pantev et al., 1998).

A little has been studied regarding whether this enhanced ability of the musicians is helpful in identification of formant frequencies of speech in quiet and in presence of background noise. There are only few studies done in musicians experts in western music. Thus there is a need of more research on carnatic musicians in this area. To shed light in this aspect, this study is considered.

**OBJECTIVES OF THE STUDY**

1. To explore behavioral discrimination ability of formant frequencies in musicians (vocal vs. instrumental) and non-musicians
2. To study electrophysiological measures of pre-attentive discrimination (MMN) of formant frequencies in musicians and non musicians.
3. To study behavioral SPIN and Speech ABR in musicians
4. To correlate behavioral discrimination and speech in noise perception measures with electrophysiological Speech ABR and MMN measures

**METHOD**

**Participants**

A total of 40 individuals will be considered for this study. This group will include 10 Vocalists, 10 Mridangam players and 10 Violinists. Remaining 10 participants will be age matched peers with no considerable music training (Non Musicians).

**Participant selection criteria**

Firstly, participants will be oriented about the aim and the procedure involved in the study. Also written consent for the willingness to participate in the study will be taken.

Other criteria followed were as below:

1. No history of any neurological, psychological, cognitive or otological problems.
2. Normal speech and language development as per reports.
3. Native Kannada speaker.
4. Air conduction thresholds (at all octave frequencies from 250 Hz to 8000 Hz) and bone conduction thresholds (at all octave frequencies from 250 Hz to 4000 Hz) obtained using Modified Hughson-Westlake procedure (Carhart & Jerger, 1959) ≤ 15 dB HL.
5. Bilateral normal middle ear function (‘A’ type tympanogram at 226 Hz probe tone and normal ipsilateral& contralateral reflexes at 500 Hz, 1000 Hz, 2000 Hz & 4000 Hz).
6. Speech recognition thresholds, SRT ± 12 dB to Pure Tone Average, PTA (PTA – average of air conduction thresholds at 0.5 kHz, 1 kHz and 2 kHz).
7. Speech Identification Scores in Quiet > 90% at 40 dB SL (re. SRT).
8. Transient Evoked Otoacoustic Emissions (TEOAEs) present (6 dB SNR & 90% reproducibility) for non-linear clicks of 260 sweeps at 80 dB pe SPL.
9. Auditory Brainstem Response (ABR) for click – normal wave V latency (Repetition rates = 11.1/sec & 90.1/sec, Intensity = 90 dB nHL)
10. No illness on the day of testing.

**Test Environment**

All testing will be carried out in an electrically shielded and sound treated room where noise levels were maintained within permissible limits - ANSI S3.1 (1999).

**Instrumentation**

1. A calibrated (ANSI S3.6-1996) dual channel clinical audiometer, Madsen OB-922 (Version 2) with TDH 39 supra-aural headphones housed in MX-41/AR ear cushion will be used for air conduction testing and Radio ear B-71 bone vibrator will be used for bone conduction testing.
2. A calibrated immittance meter, GSI Tympstar (version 2) will be used for immittance evaluation.
3. ILO System (Version 6) will be used for recording TEOAEs.
4. Evoked potential system - Biologic Navigator Pro EP (version 7) will be used for recording ABR and MMN recording.

**Procedure**

*A. Preliminary evaluations:*

Detailed history regarding otological, neurological, psychological, and cognitive problems will be taken along with the details of speech and language development. Once the possible deficits were ruled out in all these areas, pure tone audiometry, speech audiometry, immittance evaluation and TEOAE measurements will be carried out. Further, only those participants who will pass the criteria in all the above evaluations will be subjected to SPIN testing, behavioural discrimination task and followed by Speech ABR and MMN.

*B. Speech in Noise (SPIN) testing:*

Phonemically Balanced Kannada word lists developed by Yathiraj and Vijayalakshmi, (2005) will be used. Words will be presented through monitored live voice at 40 dB SL (re. SRT) and 0 dB SNR. Speech noise will be used. 25 bi-syllables in every list will be presented for each trial and every word will be given a score of 4 %. Participant will have to repeat the words heard. Number of correctly identified words will be noted down to find the SIN score.

1. *Behavioral Discrimination test:*

Synthetic continuum /da/ - /ba/ will be used for discrimination testing. The continuum will be developed such that it comprises four stimuli which are identical except onset frequencies of second and third formant. Stimuli will be presented through laptop routed through calibrated headphones of a diagnostic audiometer. 16 possible combinations of these stimuli will be presented trials at 40dB SL. Participants will be asked to report if the pair of stimulus are same or different. Final total discrimination score will be obtained.

1. *Speech Auditory Brainstem Responses (Speech ABR):*

Stimulus that will be used to record speech ABR will be synthesized /da/ syllable. This syllable is available under BioMARK protocol in the evoked potential system. It is a synthesized speech syllable produced using KLATT synthesizer (Klatt, 1980). It has broad spectral, fast temporal information and has spectrally rich formant transitions between the consonant and steady state vowel. Though the steady state vowel is not present, the stimulus is still perceived as being a consonant-vowel syllable. Here the fundamental frequency (F0) linearly rises from 103 to 125 Hz with voicing beginning at 5 ms and an onset noise burst at first 10 msec. 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.

Participant will be made to sit comfortably on a reclining chair. The skin surface on the two mastoids and forehead will be cleaned with skin abrasive. Gold cup electrodes will be used to record responses. The electrodes will be placed with the help of skin conduction paste and surgical plaster will be used to hold the electrodes tightly on the respective places. Absolute electrode impedance will be maintained below 5k Ω with inter electrode impedance below 2k Ω. Before starting the recording, participants will be instructed to relax and refrain from extraneous body movements to minimize artifacts. Speech ABR will be recorded and replicated. Weighted average of the two waves will be obtained. Table 1 shows the protocol that will be used to record speech ABR.

*Table 1:Protocol for Speech ABR recording*

|  |  |  |
| --- | --- | --- |
| **Stimulus**  **Parameters** | Transducer | ER-3A insert earphones |
| Stimulus type | Synthesized CV Syllable- /da/ |
| Stimulus duration | 40 ms |
| Stimulus intensity | 80 dB SPL |
| Repetition rate | 7.1/sec |
| Sweeps | 2000 |
| Polarity | Rarefraction |
| Noise type | White noise |
| Noise presentation | Ipsilateral |
| **Acquisition parameters** | Electrode montage | Vertical |
| Electrode sites | Inverting = M1 (Ipsilateral mastoid)  Non inverting = Cz (Vertex)  Ground = M2 (Contralateral mastoid) |
| Amplification | 1 lakh times |
| Analysis time | 64 ms |
| Pre-stimulus time | 10 ms |
| Filters | 100-3000 Hz |
| Notch filter | Off |
| Number of channels | 1 |

Analysis of Speech ABR response:

Response will usually contain onset peaks and sustained components. The following analysis will be carried out for each ABR recording:

1. Subjective analysis:

Latencies of the onset peaks V, A, and transition peaks C, D, E and F will be marked and noted. Also, amplitude will be measured from voltage difference between wave V and wave A.

1. Objective analysis:

It will be done by evaluating the spectral composition of sustained portion of the response using Fast Fourier Transform (FFT). Fourier analysis will be performed on 11.4 – 40.6 ms epoch of the FFR in order to assess the amount of activity occurring over three frequency ranges; (103-121 Hz), (454-719 Hz) and (721-1155 Hz). These frequency ranges will be chosen because the neural responses at these frequencies correspond to fundamental frequency, formant formant and higher harmonics of the stimulus /da/ respectively (Johnson et al. 2008). On and off Hanning ramp of 2 m sec each will be applied to the waveform to avoid spectral splatter. The raw amplitude of F0, F1 and higher harmonic components will be measured and noted**.**

1. **Mismatch negativity (MMN)**

Four stimuli obtained from synthetic /da/-/pa/ continuum will be used as stimulus even for obtaining MMN response. Response will be obtained for each stimulus contrast in odd ball paradigm which will contain waveforms for frequent and infrequent stimulus. Probability used will be standard 80% and deviant 20%. Protocol that will be used for MMN recording is shown in table 2.

Table 2: *Protocol for MMN recording*

|  |  |  |
| --- | --- | --- |
| Stimulus  Parameters | Transducer | ER-3A insert earphones |
| Stimulus type | 4 stimuli obtained from synthesized CV Syllable continuum /da/-/pa/ |
| Stimulus intensity | 80 dB SPL |
| Repetition rate | 1.1/sec |
|  | 200 |
| Polarity | Rarefraction |
| Acquisition parameters | Electrode montage | Vertical |
| Electrode sites | Inverting = M1 (Ipsilateral mastoid)  Non inverting = Cz (Vertex)  Ground = M2 (Contralateral mastoid) |
| Amplification | 50000 times |
| Analysis time | 400 ms |
| Pre-stimulus time | 10 ms |
| Filters | 1-30Hz |
| Number of channels | 1 |

**Statistical analysis:**

Statistical analysis will be carried out using Statistical Package for Social Sciences (SPSS) software.

**Implications of the study:**

The study is expected to provide evidence for effect of musical training on formant frequency discrimination through behavioral and electrophysiological tests. Further, study explores if correlations exists between formant frequency discrimination and speech in noise perception skills.