Vocal Emotion Recognition in Group of Individuals with Normal Hearing, Sensorineural Hearing Loss and Auditory Dys-synchrony

[1]Priyanjali Harit & [2]Devi N.

Abstract

The study aimed to construct and compile the vocal emotion recognition battery in Kannada language and also to assess vocal emotion perception in group of individuals with normal hearing, sensori-neural hearing loss and auditory dys-synchrony. Using vocal emotion recognition battery in Kannada language 20 normal hearing ears (Group-A) and 10 ears of individuals with sensori-neural hearing loss(Group-B) and auditory dys-synchrony (Group-C) each were assessed. Comparison of performance across the three groups on vocal emotion perception recognition revealed that there was a significant difference across all the five emotions. Between Group-A and Group-B, there was significantly different seen across all the five emotions. Similar findings were also obtained between Group-A and Group-C. Whereas when the performance of Group-B and Group-C were compared, result revealed significant difference across all the emotions except the emotion 'fear'. For the emotion 'fear' the perception was not significantly different between the Group-B and Group-C, which indicates that perception ability of the participants in both the groups were similar in perceiving 'fear', which could be because of the underlying acoustic parameters of the stimuli. The vocal emotion perception battery developed as a part of this study is a useful is assessing the vocal emotion perception abilities of individuals with normal hearing, sensori-neural hearing loss and auditory dys-synchrony/neuropathy.

Key words: *Sensorineural hearing loss, auditory dys-synchrony, vocal emotion recognition.*

Introduction

Speech is a major form of communication that conveys both linguistic and non-linguistic cues. The linguistic component of the speech includes the properties of the speech signal and word sequence and it deals more with what is being said, not how it is said. The non-linguistic component of speech deals more with talker attributes such as age, gender, dialect, and emotion. Cues to non-linguistic properties can also be provided in non-speech vocalizations, such as laughter or crying (Peters, 2006). According to Luo, Fu and Galvin (2007) imperative element of speech communication involves identification of a talker's emotional state, using only acoustic cues. Although facial expressions may be strong indicators of a talker's emotional state, vocal emotion recognition is an important component of auditory-only communication such as telephone conversation or listening to the radio.

Ear is the most essential part for the communication. Perception of any acoustic stimuli involves various steps, which includes conversion of an auditory stimulus into electrical signal at the receptor level, transmission of the electrical signal through the peripheral nerve, and processing and interpretation of the electrical signal in the central nervous system. Any breakdown in the process results in the major consequences in the perception. There is abundant literature on the perceptual consequences of both peripheral and central auditory disorders. For example, peripheral damage in the inner ear and auditory nerve leads to threshold elevation, abnormal loudness, pitch and temporal processing (Ryan & Dallos, 1975; Nienhuys & Clark, 1978; Prosen, Moody, Stebbins, & Hawkins, 1981; Formby, 1986; Moore, 1996; Moore & Oxenham, 1998; Buss, Labadie, Brown, Gross, Grose, & Pillsbury, 2002; Oxenham & Bacon, 2003) whereas, central auditory disorders produces complex processing deficits in speech and sound object recognition (Levine et al., 1993; Wright, Lombardino, King, Puranik, Leonard & Merzenich, 1997; Cacace & McFarland, 1998; Gordon-Salant & Fitzgibbons 1999).

The most noticeable indication of cochlear hearing loss is a reduced ability to detect weak sounds. However cochlear hearing loss is also accompanied by a variety of other changes in the way that sound is perceived. According to Moore (1996) severe to profound cochlear hearing loss not only reduces the functional bandwidth of hearing but also reduces frequency selectivity, impaired intensity and temporal resolution, and also results in making detection of the subtle acoustic features contained in vocal emotion difficult to detect.

Auditory dys-synchrony/neuropathy (AD) is a clinical syndrome which has disrupted auditory nerve activity with concurrently normal or near normal cochlear amplification function (Starr, Picton, Sininger, Hood &

[1]E-mail: priyannjali@gmail.com; [2]Lecturer in Audiology, E-mail: deviaiish@gmail.com

Berlin, 1996). Clinically, the disrupted auditory nerve activity is reflected by highly distorted or absent auditory brainstem responses, whereas the normal cochlear amplification function is reflected by the presence of oto-acoustic emission and/or cochlear microphonics (Rance et al., 1999; Starr et al., 1996; Hood, Berlin, Bordelon & Rose, 2003). The other main characteristic of AD is a significantly impaired ability for temporal processing and difficulty in speech understanding, especially in noise, that is disproportionate to the degree of hearing loss measured by puretone thresholds (Zeng, Oba, Garde, Sininger & Starr, 1999; Kraus, Bradlow, Cheatham, Cunningham, King & Koch, 2000; Rance, Cone-Wesson, Wunderlich & Dowell, 2002; Rance, McKay & Grayden, 2004; Zeng, Kong, Michalewski & Starr, 2005). The above mentioned difficulties in temporal processing and speech understanding may also result in difficulty in recognizing vocal emotions.

In Indian scenario, incorporating assessment tool for the evaluation of speech perception abilities exists, findings of which can be utilised in the management of individuals with cochlear hearing loss as well as for those with auditory dys-synchrony/neuropathy. It is comparatively easy to measure speech intelligibility through metrics like the word recognition scores. Listeners are asked to identify spoken words under various conditions and their recognition score is taken as a measure of intelligibility. But we don't have such straightforward method to assess vocal emotion perception in general especially in Indian scenario. Instead we can assess its constituent elements such as pitch, duration, and intensity.

Disruption in the perception of temporal cues has been demonstrated in both children and adults with auditory dys-synchrony/neuropathy (Starr et al., 1996; Zeng et al., 1999; Kraus et al., 2000; Rance, et al., 2004; Michalewski, Starr, Nguyen, Kong & Zeng, 2005). In addition to distortion of the spectral information that is seen in cochlear hearing impaired individuals, individuals with AD also possess distortion in temporal information (Zeng et al., 1999; Kraus et al., 2000; Rance et al., 2004). Hence the input signal in the auditory system is lot more distorted in individuals with AD compared to those with cochlear pathologies.

Reviewing the literature, the perception of segmental aspects of speech in AD had been proposed but researchers are lacking in the knowledge of vocal emotion perception in the hearing impaired group and AD population. As the features of suprasegmentals in terms of rhythm, stress, intonation, prosody, are included in acoustic feature of vocal emotion, an attempt is required to understand the recognition of different vocal emotion by normal hearing adults, hearing impaired individuals and individuals with AD.

The aims of the present study were to compile/construct vocal emotion perception battery in the context of Kannada language. To investigate the recognition of vocal emotions perception abilities of individuals with normal hearing. minimal to mild sensori-neural hearing loss and auditory dys-synchrony (AD)/neuropathy. To compare the performance across the groups on the vocal emotion perception task.

Method

The study was carried out in four phases.

Phase I: Generation of stimulus (vocal emotion perception battery)

The test stimulus consisted of 5 sentences each in 5 target emotions: happy, sad, angry, fear and neutral. Initially fifty 3-4 words sentences in Kannada language which were more appropriate to fit into all the five emotions were selected. Out of those fifty sentences; five 3-word sentences were selected for the study, which were further evaluated by the five Audiologists and Speech Language Pathologists on the basis of their simplicity and appropriateness. Each target stimuli sentences (Kannada) were recorded in five different emotions. Hence total of 25 stimuli sentences was developed. Overall amplitude differences between the stimuli was either preserved or normalized.

The recording was done in the sound treated room where the noise level were as per the ANSI guidelines (1991) on a data acquisition system with a 16 bit analog to digital convertor at a sampling frequency of 44.1 kHz. The material was edited and scaling was done using audition version 3 software with a calibrated microphone. Interval of 10 seconds is added between the sentences to function as a response time. Professional stage actors of native Kannada language were taken for the recording of the stimulus. They were instructed to say the sentences with different emotions. Recording was done for 2 male and 2 female actors and then the 'goodness test' was performed by the 2 Audiologists and 2 Speech Language Pathologists to choose the best clarity voice for the final recording. A female voice was taken for the final recording of the stimulus. Randomly 2 sentences from the 25 sentences were taken as the 'trial' sentences. Hence the final battery consisted of 2 trial stimuli with 25 test stimuli. Recorded stimulus was then stored into a compact disk (CD).

Phase II: Selection of participants under different groups

Participants

Three groups of participants were included in the study; the groups were named as Group-A, Group-B and Group-C. Participants in all the groups were native speaker of Kannada language within an age range of 15 to 40 years. 20 ears were included in Group-A whereas Group-B & Group-C consisted each of 10 ears.

Participants with hearing sensitivity within normal limits formed the group-A. Their hearing threshold were <15 dBHL at octave intervals between 250 Hz and 8000 Hz for air conduction and between 250 Hz to 4000 Hz for bone conduction. Speech identification scores of all the 20 ears were above 90% in quite. All the ears had normal middle ear functioning with type 'A' tympanogram with normal ipsilateral as well as contralateral reflex threshold. All the ears had bilateral normal cochlear outer hair cell functioning with presence of transient evoked oto-acoustic emissions (TEOAEs). Normal click evoked auditory brainstem responses (ABR) were obtained, with the occurrence of wave 'V' within the stimulus level of 30 dBnHL for all the ears. The participants did not have any past or present history of otological abnormalities and/or neurological deficit. No physical illness was seen on the day of testing.

Participants with minimal to mild sensori-neural hearing loss were included in Group-B with the hearing thresholds ranged from 15 dBHL to 40 dBHL at octave intervals between 250 Hz and 8000 Hz for air conduction and between 250 Hz to 4000 Hz for bone conduction. Speech identification scores of all the 10 ears were proportionate to their pure tone average in quite. All the ears had normal middle ear functioning with type 'A' tympanogram with normal ipsilateral as well as contralateral reflex threshold. All the ears had normal cochlear outer hair cell functioning with presence of transient evoked otoacoustic emissions (TEOAEs). Wave 'V' of click evoked auditory brainstem responses was obtained till 50 dBnHL for all the ears. They did not have any past or present history of otological abnormalities and/or neurological deficit. No physical illness was seen on the day of testing.

Ears with minimal to mild sensorineural hearing loss were included in Group-B with the hearing thresholds ranged from 15 dBHL to 40 dBHL at octave intervals between 250 Hz and 8000 Hz for air conduction and between 250 Hz to 4000 Hz for bone conduction. Speech identification scores of all the 10 ears were not proportionate to their pure tone average scores in quite (poorer as compared to their pure tone average scores).

All the ears had normal middle ear functioning with type 'A' tympanogram with absence of ipsilateral as well as contralateral reflexes. All the ears had normal cochlear outer hair cell functioning with presence of transient evoked oto-acoustic emissions (TEOAEs). Abnormal/Absent click evoked auditory brainstem responses were obtained for all the ears. No physical illness was seen on the day of testing.

Test environment

Audiometric testing and administration of the test battery were carried out in sound treated room with the ambient noise levels within permissible limits (ANSI S 3.1-1991).

Instrumentation and test protocol

A calibrated diagnostic audiometer, GSI-61 with TDH-39 earphones was used for estimating the air conduction thresholds. Radio ear B-71 bone vibrator was used for bone conduction testing. A calibrated diagnostic audiometer MAICO MA 53 with TDH 39 earphones was used to present the developed stimulus to administer vocal emotion perception battery. A calibrated middle ear analyzer GSI tympstar was used to record tympanogram with a probe tone frequency of 226 Hz and the acoustic reflexes thresholds were measured for 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz. Brainstem responses to click stimuli were recorded using Biologic Navigator Pro evoked potential systems. The site of electrode placement was prepared with skin preparation gel. Silver chloride electrodes were used with a conducting pate. Responses were differentially recorded from Ag-AgCl electrodes with each electrode impedance of < 5 kΩ. The following test protocol (Table 1) was used for the recording of ABR;

ILO version 6.0 was used to record the transient evoked oto-acoustic emissions (TEOAEs). The test stimulus was recorded on Adobe audition version 3 installed in a personal computer via a microphone (Ahuja, AUD-101XLR) placed at a distance of 10 cm from the lips of the speaker while recording.

Phase III: Assessment of vocal emotion perception battery

Participants were made to be seated in a sound treated room (two room situation) and were made to listen to the stimuli. The vocal emotion perception battery from the laptop computer was routed via auxiliary input to the audiometer. The prepared stimulus was administered through headphone at most comfortable

Table 1: Parameters used to acquire ABR

Stimulus parameters		*Acquisition parameters*	
Type of stimulus	Click	Low pass filter	100 Hz
Polarity	Rarefaction	High pass filter	3000 Hz
Intensity	Variable	Notch filter	On
Number of stimuli	1500	Artifact rejection	On
Repetition rate	30.1/s	Time window	15ms
		Electrode montage	A1-Fz-A2

level (40dBSL) of the participants in each of the three groups. Before administering the vocal emotion perception battery, a response sheet was given to each of the participants along with an appropriate instruction.

A closed set, 5 alternative identification tasks was used to measure vocal emotion recognition. In each trail, a sentence was randomly selected (without replacement) from the stimulus set and presented to the participants. The participants were instructed torespond by ticking (√) on the 5 response choices (picture along with label- happy, sad, angry, fear and neutral). In the commencement of test, two sentences were given as a trial. No feedback was provided. Response were collected and scored in terms of raw correct scores.

A scoring sheet was also developed which was used to score the responses given by the participants. Scoring was done separately for each of the emotion. A score of '1' was given for every correct response and '0' for every incorrect response. The maximum score for each emotion was 5 and the maximum overall score was 25.

Phase IV: Test-Retest reliability

Test to retest reliability of vocal emotion perception battery was checked by administering it to randomly selected one-half of the participants from all the three groups for a second time after duration of two weeks. The procedure for the carrying out test-retest reliability was same as it was done for the first time.

Statistical analysis

Descriptive statistical analysis of the scores in terms of mean, standard deviation, and other non-parametric tests such as Kruskal-Wallis test, Mann-Whitney test, Friedman Test, Wilcoxon Signed Rank test were performed using the SPSS (10.0 & 17.0) software. The results obtained are presented and discussed in the subsequent section.

Results and Discussion

Audiological findings of the participants Audiological assessment was carried out to divide the participants into Group-A, Group-B and Group-C. Different audiological tests such as pure tone audiometry, Speech audiometry (speech identification scores), immittance evaluation (tympanogram, ipsilateral and contralateral acoustic reflexes), TEOAEs, ABR were administered on each of the ears participated in the study. Before performing the audiological testing, all the participants went underwent Otoscopic examination which revealed no otological problem.

Audiological findings of participants in Group-A: Group-A included 20 ears with hearing sensitivity within normal limits. The audiological findings revealed hearing sensitivity within normal limits in all the 20 ears, and their speech scores were proportionate to the pure tone audiometric thresholds with >90% scores in quite. On immittance evaluation, all the ears showed type 'A' tympanogram with presence of ipsilateral as well as and contralateral reflexes. Normal outer hair cell functioning was established in all the ears with presence TEOAEs. All the ears had normal click evoked ABR with presence of wave 'V' till the intensity level of 30 dBnHL.

The audiological findings of Group-A is in congruence with the results obtained by Clark (1981) that pure tone average of three frequencies (500 Hz, 1 kHz and 2 kHz) ranged from -10 to 15 dBHL that indicated normal hearing sensitivity. According to Harris (1991) and Glattke (2002) oto-acoustic emissions (OAEs) are the sounds of cochlear origin and Transient evoked OAEs are highly sensitive to cochlear pathology and in a frequency-specific way. TEOAE responses are typically absent at frequencies at which hearing thresholds exceed 20 to 30dBHL, therefore TEOAEs of all the ears included in the Group-A were present with thresholds' not exceeding beyond 20 to 30 dBHL. In normal hearing listeners wave 'V' of ABR is the most visible peak at the lower intensity level of 25 to 30 dBnHL, which is an indication of normal auditory brainstem functioning. Wave 'V' was present within the intensity level of 30dBnHL in all the ears in Group-A. Jerger (1970), Jerger, Jerger and Mauldin (1972) and Liden, Harford and Hallen (1974) classified various tympanogram according to their height and location of peaks, according to these authors type ‘A’ tympanogram shows normal peak height and is characteristic of normal middle ear functioning. Similar findings of immittance evaluation was seen in Group-A. Jerger et al., (1972) stated that ARTs at the sensation level less than 60 dBSL are consistent with cochlear hearing impairment where as ARTs more than 60 dBSL are consistent with retrochochlear hearing impairment in sensori-neural hearing impaired ears.

Audiological findings of participants in Group-B: To include the participants into Group-B, various audiological tests were performed such as pure tone audiometry, immittance evaluation, OAEs and ABR to confirm the sensori-neural hearing impairment with a degree of minimal to mild. All the ears included in Group-B were diagnosed as minimal/mild sensori-neural hearing loss depending upon their hearing thresholds. Pure tone thresholds for all the ears ranged between 15 dBHL and 40 dBHL with speech identification score proportionate to their pure tone average scores with not less than 90% in quite. Immittance evaluation revealed type ‘A’ tympanogram with present ipsilateral as well as contralateral reflexes for all the ears included in this group. OAEs revealed normal cochlear outer hair cell functioning with the presence of TEOAEs in all the ears except the ear 5 and 6 where TEOAEs were absent indicating the cochlear outer hair cell damage. Wave ‘V’ of ABR was obtained till 50dBnHL for all the ears who were included in the Group-B.

The audiological findings of Group-B is in congruence with the results obtained by Clark (1981) that pure tone average of minimal and mild sensori-neural hearing loss ranges from 15 to 25 dBHL and 25 to 40 dBHL respectively. According to Harris (1991) and Glattke (2002), OAEs are the sounds of cochlear origin and TEOAEs are highly sensitive to cochlear pathology and in a frequency-specific way. TEOAE responses are typically absent at frequencies at which hearing thresholds exceed 20-30dBHL, this results supports the findings of the present study with the absence of TEOAEs in the ear 5 and 6. In minimal to mild sensori-neural hearing loss wave ‘V’ of ABR usually appears till the intensity level of 50 dBnHL. Jerger (1970), Jerger et al., (1972) and Liden et al. (1974) have described various tympanogram according to their height and location of peaks, according to these authors middle ear function is indicated by type ‘A’ tympanogram.

Audiological findings of the participants in Group-C: Ears which were considered in the Group-C were diagnosed with minimal or mild sensori-neural hearing loss with auditory dys-synchrony(AD)/neuropathy based on their audiological findings of behavioural and electrophysiological tests. Results of behavioural and electrophysiological test performed on the participants showed ears which were selected under Group-C were diagnosed as minimal or mild sensori-neural hearing loss with AD/ neuropathy. Speech identification score were not proportionate to their pure tone average with scores poorer than the expected score in quite condition, except for the ear 5 and 6 where the speech identification scores were proportionate to their pure tone average scores. Immittance evaluation results showed type ‘A’ tympanogram with absent ipsilateral as well as contralateral reflexes for all the ears. No click evoked ABR were recorded for any of the ear included in Group-C.

The Audiological findings of Group-C is in congruence with the results obtained by Starr et al. (1991) that auditory dys-synchrony/neuropathy is a clinical syndrome which has disrupted auditory nerve activity with concurrently normal or near normal cochlear amplification function. Clinically, the disrupted auditory nerve activity is reflected by highly distorted or absent auditory brainstem responses, whereas the normal cochlear amplification function is reflected by the presence of oto-acoustic emission and/or cochlear microphonics (Starr et al., 1996; Rance et al., 1999; Hood et al., 2003). The other main characteristic of AN is a significantly impaired ability for temporal processing and difficulty in speech understanding, mainly in noise, that is disproportionate to the degree of hearing loss measured by pure-tone thresholds (Zeng et al., 1999; Kraus et al., 2000; Rance wt al., 2002; Rance et al., 2004; Zeng et al., 2005). Absent or severely abnormal ABR which does not correlate with audiometric thresholds and preserved otoacoustic emissions (Starr et al., 1991, Berlin et al., 1993, Starr et al., 1996).

Mean and standard deviation (SD) of different emotions among different groups

Vocal emotion perception battery was administered on all the ears included under three groups mentioned above to assess their vocal emotion recognition abilities. Assessment of the vocal emotion perception battery was done monaurally under headphone condition at the most comfortable level of the listeners.

As it can be seen in the Graph 1 the scores obtained by the participants of Group-A were the highest as compared to the Group-B and Group-C. In Group-A, 100% scores were obtained for the emotion ‘sad’, ‘anger’ and ‘fear’ with a mean value of 5.00 each. For the emotion ‘happy’ and ‘neutral’ the mean scores were 4.75 and 4.90 respectively.

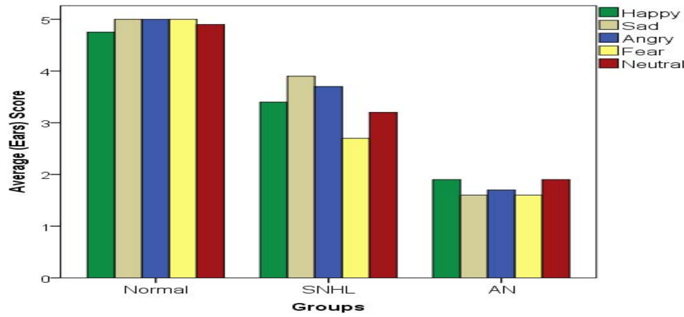

Figure 1: Comparison of different emotions among different groups.

In Group-B, the highest mean score was seen for the emotion ‘sad’ with the value of 3.90 followed by the emotion ‘anger’ with the mean value of 3.70. The least mean value of 2.70 was seen for the emotion ‘fear’, whereas the emotions ‘happy’ and ‘neutral’ are intermediate with a mean score of 3.40 and 3.20 respectively.

In Group-C, the overall scores for all the different emotions were comparatively less from the Group-A and Group-B. The emotion ‘happy’ and ‘neutral’ shared same mean scores of 1.90, for ‘anger’ the mean score was 1.70, whereas for ‘sad’ and ‘fear’ the mean score was least of 1.60 each.

Comparison of emotions among different groups

Comparison across different emotions was carried out using Kruskal-Wallis test (non-parametric tests). Non-parametric test (Table 2) was applied due to large variability of the sample size among the groups. Results of Kruskal-Wallis test showed significant difference ($p<0.05$) across the five emotions.

Comparison of emotions between Group-A and Group-B

Mann-Whitney Test (non-parametric test) was applied for the comparison of different vocal emotions among Group-A and Group-B. Non-parametric test was applied due to the large variability in the sample size between the Group-A and Group-B.

Table 3 summarises the statistical results of comparison of different emotions between Group-A and Group-B. Findings reveal significant difference ($p<0.05$) among Group-A and Group-B across all the five emotions, which indicates that the vocal emotion perception by individuals with normal hearing is different from those individuals with minimal to mild sensori-neural hearing loss.

Comparison of different emotions between Group-A and Group-C

Mann-Whitney Test (Non-parametric test) was applied for the comparison of emotions between Group-A and Group-C.

Table 2: Result of Kruskal-Wallis test comparing across different emotions

	Happy	Sad	Anger	Fear	Neutral
Chi-Square	23.927	32.013	29.391	25.242	26.804
df	2	2	2	2	2
Asymp. Sig.	0.000	0.000	0.000	0.000	0.000

Table 3: Results of Mann- Whitney test comparing different emotions between Group-A and Group-B.

		Group-B				
		Happy	Sad	Anger	Fear	Neutral
Group-A	Happy	-	SD	SD	SD	SD
	Sad	SD	-	SD	SD	SD
	Anger	SD	SD	-	SD	SD
	Fear	SD	SD	SD	-	SD
	Neutral	SD	SD	SD	SD	-

Note: ‘SD’ indicates ‘significant difference’.

Table 4: Results of Mann- Whitney test comparing different emotions between Group-A and Group-C

		Group-C				
		Happy	Sad	Anger	Fear	Neutral
Group-A	Happy	-	SD	SD	SD	SD
	Sad	SD	-	SD	SD	SD
	Anger	SD	SD	-	SD	SD
	Fear	SD	SD	SD	-	SD
	Neutral	SD	SD	SD	SD	-

Note: 'SD' indicates 'significant difference'.

Table 4 revealed a significant difference ($p<0.05$) between both the groups across all the five emotions, which indicates that the vocal emotion perception by normal hearing listeners is different from those individuals with minimal or mild sensori-neural hearing loss with auditory dys-synchrony(AD)/neuropathy.

Comparison of emotions among Group-B and Group-C

Group-B and Group-C were also compared across different emotions using Mann-Whitney Test. Table 5 shows the findings of statistical analysis applied to Group-B and Group-C for the comparison of different vocal emotions. Results showed a significant difference of $p<0.05$ among both the groups across all the five emotions except the emotion 'fear'. The emotions 'happy', 'sad', 'anger' and 'neutral' were significantly different between the two groups indicating that the perception of these four emotions were different in individuals with minimal to mild sensori-neural hearing loss from those with auditory dys-synchrony (AD)/neuropathy. For the emotion 'fear' the perception was not significantly different between the Group-B and Group-C, which indicates that perception ability of participants in both the group in perceiving the emotion 'fear' is similar. The acoustic parameters of the 'fear' stimuli are, pitch of 300.61 Hz, F1 of 885.8 Hz, intensity of 76.8 dB with duration of 2.1 seconds, which infers that in both groups it is affected in a similar way.

Within group comparison across different emotions.

Friedman test (non-parametric test) was applied as a statistical tool for the analysis of within group comparisons across different emotions. Table 6 summarises the different emotions across groups separately. In Group-A some difference is seen across the emotions whereas for Group-B and Group-C, no difference is seen across different emotions.

Comparison among different emotions

Comparison of different emotions was done with respect to the responses obtained by the participants in all the three groups as well as based on the acoustic analysis of the stimuli used in the vocal emotion perception battery. Wilcoxon Signed Rank test was applied for the comparison of different emotions. Table 7 reveals that the results of various pair wise comparison across different emotions such as sad with happy, anger with happy, fear with happy, neutral with happy, anger with sad, fear with sad, neutral with sad, fear with anger, neutral with anger and neutral with fear. Difference is seen among all the pairs listed above. These findings were supported by the findings of the acoustical analysis of the stimuli which is discussed further in the following subsection.

Table 5: Results of Mann- Whitney test comparing different emotions between Group-B and Group-C

		Group-C				
		Happy	Sad	Anger	Fear	Neutral
Group-B	Happy	-	SD	SD	SD	SD
	Sad	SD	-	SD	SD	SD
	Anger	SD	SD	-	SD	SD
	Fear	SD	SD	SD	-	SD
	Neutral	SD	SD	SD	SD	-

Note: 'SD' indicates 'significant difference'.

Table 6: Results of Friedman test comparing different emotions within each group

Statistical parameters	Group-A	Group-B	Group-C
Chi-Square	13.714	1.246	4.946
df	4	4	4
p	0.008	0.871	0.293

Table 7 : Results of Wilcoxon signed rank test comparing different emotions based on the responses

	Z	Asymp. Sig. (2-tailed)
S-H	-2.236	0.025
A- H	-2.236	0.025
F-H	-2.236	0.025
N-H	-1.134	0.257
A-S	0.000	1.000
F-S	0.000	1.000
N-S	-1.414	0.157
F-A	0.000	1.000
N-A	-1.414	0.157
N-F	-1.414	0.157

Table 8: Results of Signed rank test comparing different emotion based on their acoustic characteristics

	Happy	*Sad*	*Anger*	*Fear*	*Neutral*
Pitch (Hz)	*395.65*	*257.81*	*324.01*	*300.61*	*238.73*
F1	*856.89*	*776.80*	*861.91*	*885.80*	*781.83*
F2	*1927.46*	*1869.65*	*1994.45*	*1963.6*	*1949.07*
F3	*2985.17*	*3033.95*	*3181.91*	*3099.15*	*3106.95*
F4	*3852.63*	*4012.27*	*3974.20*	*4022.41*	*4054.35*
Intensity (dB)	*78.66*	*77.6*	*79*	*76.80*	*73.92*
Duration (sec)	*1.7*	*1.8*	*1.55*	*2.1*	*1.53*

Table 8 shows the findings of the acoustic analysis of the stimuli used in the vocal emotion perception battery. Difference in pitch values (Hz) was seen in all the emotions. The target emotions were ordered in terms of their pitch values (from high to low) as 'happy', 'anger', 'fear', 'sad' and 'neutral' with a pitch values of 395.65 Hz, 324.01 Hz, 300.61 Hz, 257.81 Hz and 238.73 Hz respectively. Intensity difference was also seen across the emotions, the target emotions were ordered in terms of their intensity values (from high to low) as 'anger', 'happy', 'sad', 'fear' and 'neutral' with the a values of 79 dB, 78.66 dB, 77.6 dB, 76.80 dB, and 73.92 dB respectively. These findings are in congruence with the findings of Petrushin (2000), Scherer (2003) and Yildirim et al., (2004) that relative to neutral speech, anger and happy speech exhibits higher mean pitch, wider pitch range, greater intensity, faster speaking rate. On the other hand, sad speech exhibits lower mean pitch, narrower pitch range, lower intensity and slow speaking rate.

The target emotions were ordered in terms of their F1 values (from high to low) as 'fear', 'anger', 'happy', 'neutral' and 'sad' with a values of F1 as 885.80 Hz, 861.91 Hz, 856.89 Hz, 781.83 Hz and 776.80 Hz respectively. Luo et al., (2007) found similar findings for male talker with mean F1 values (from high to low) as 'anger', 'anxious', 'happy', 'neutral' and 'sad'.

According to Williams and Stevens (1972), when the speaker is angry the basic opening and closing articulatory gestures characteristic of the vowel-consonant alteration in speech appeared to be more extreme. The vowels tended to be produced with more open vocal tract and hence have higher first-formant frequencies, and the consonants were generated with a more clearly defined closure. In the present study also, higher first formant frequency was found for the emotion 'anger' as compared to other emotions. The vowel and consonants produced in a fear situation were often more precisely articulated than they were in neutral situation and the average Fo for fear was found to be lower than that observed for anger and for some voices it was found to be close to that of the utterance spoken in the neutral situation. The finding of Fo of fear in the present study correlates with the findings of Williams and Stevens (1972). These authors also concluded that the average fundamental frequency for the utterances spoken in the sorrow situation was considerably lower than that for neutral situations and the range of Fo was usually found to be quite narrower. In the present study, Fo for the emotion 'sad' was found to be higher than the emotion 'neutral'.

In the present study, acoustic characteristic in terms of the duration of the stimuli was also taken into consideration. The target emotions were ordered in terms of their duration values (from high to low) as 'fear', 'sad', 'happy', 'anger', and 'neutral'. Williams and Stevens (1972) also summarised the duration parameters associated with various parameters. They concluded that the duration of utterances spoken in 'anger' were usually longer, but this effect was not great and was not always consistent in all voices. Whereas the duration of an utterance in fear situation tended to be longer than in the case of 'anger' or 'neutral' situation. Increased duration of the stimulus is also seen for the emotion 'sad' which resulted from longer vowels and consonants and from pauses that were often inserted in a sentence.

Therefore it can be concluded from the findings given in Table 7 and Table 8 that all the emotions differ from each other in terms of their acoustic properties as well as the responses given by the participants of the present study.

Results of reliability assessment of the test

Test to retest reliability of the vocal emotion perception battery was checked by administering it to randomly selected one-half of the participants from all the three groups for a second time after duration of two weeks. The number of correct responses given by each participant in each of the group for the test was noted. The results were compared with the respective participant's results obtained during the first test. Alpha coefficient was used for the reliability assessment. For Group-A the alpha value was 0.7, for Group-B, it was 0.98, where as for Group-C it was 0.90. The results showed a good test-retest reliability in all the three groups.

Conclusions

Vocal emotion perception battery developed as a part of this study was useful is assessing the vocal emotion perception abilities of individuals with normal hearing, sensori-neural hearing loss and auditory dys-synchrony/neuropathy. Mean and standard deviation of scores obtained for different emotions showed significant difference among all the three groups. Comparison of performance across the three groups on vocal emotion perception recognition revealed that there was a significant difference across all the five emotions. It was observed for within Group comparison, Group-A showed some difference among the emotions whereas for Group-B and Group-C, no difference was seen among the emotions. Results obtained from comparison among different emotions with respect to both responses obtained from the participants and the acoustic characteristic of the stimulus reveals that the difference is seen among all the various pair wise comparison across different emotions such as sad with happy, anger with happy, fear with happy, neutral with happy, anger with sad, fear with sad, neutral with sad, fear with anger, neutral with anger and neutral with fear. Comparison of the acoustic characteristics of the stimulus reveals the difference in the pitch, intensity and duration values of all the five emotions

References

American National Standard Institute (1991). *Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms*. American National Standard Institute S 3.1 – (1991), New York.

Buss, E., Labadie, R.F., Brown, C., Gross, A. J., Grose, J. H., & Pillsbury, H. C. (2002). Outcome of cochlear implantation in pediatric auditory neuropathy. *Otology and Neurotology, 23*, 328-332.

Cacace, A. T., & McFarland, D. J. (1998). Central auditory processing disorder in school-aged children: a critical review. *Journal of Speech Language and Hearing Research, 41*, 355-373.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *American Speech and Hearing Association, 23*, 493–500

Formby, C. (1986). Modulation detection by patients with eighth-nerve tumors. *Journal of Speech and Hearing Research, 29*, 413-419.

Glattke, T. J., & Robinette, M. S. (2002). Transient evoked otoacoustic emissions. In: Robinette, R.M., & Glattke, T. (Eds). *Otoacoustic Emissions – Clinical Applications* (2nd edn.) Thieme publications, New York.

Gordon-Salant, S., & Fitzgibbons, P. J. (1999). Profile of auditory temporal processing in older listeners. *Journal of Speech Language and Hearing Research, 42*, 300-311.

Harris, F. P., & Probst, R. (1991). Reporting click-evoked and distortion-product otoacoustic emission results with respect to the pure tone audiogram. *Ear and Hearing, 12*, 399–405.

Hood, L. J., Berlin, C. I., Bordelon, J., & Rose, K. (2003). Patients with auditory neuropathy/dys-synchrony lack efferent suppression of transient evoked oto-acoustic emissions. *Journal of the American Academy of Audiology*, *14*, 302–313.

Jerger, J. (1970). Clinical experience with impedance audiometry. *Archives of Otolaryngology, 92,* 311-324.

Jerger, J., Jerger, S., & Mauldin, L. (1972). Studies in impedance audiometry: In Normal and sensorineural ears. *Archives of Otolaryngology, 96,* 523-531.

Kraus, N., Bradlow, A. R., Cheatham, J., Cunningham, C. D., King, D. B., Kouch, T. G., Nicol, T. J., McGee, L. K., Stein, L. K., & Wright, B. A (2000). Consequences of neural asynchrony: a case study of auditory neuropathy. *Journal of the Association for Research in Otolaryngology, 1,* 33-45.

Levine, R. A., Gardner, J. C., Fullerton, B. C., Stufflebeam, S. M., Carlisle, E. W., Furst, M., Rosen, B. R., & Kiang, N. Y. (1993). Effect of multiple sclerosis brainstem lesions on sound lateralization and brainstem auditory evoked potentials. *Hearing Research*, *68*, 73-88.

Liden, G., Harford, E., & Hallen, O. (1974). Tympanometry for the diagnosis of ossicular disruption. *Archives of Otolaryngology*, *99*, 23-29

Luo, X., Fu, Q. J., & Galvin, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, *11,* 301-315.

Michalewski, H. J., Starr, A., Nguyen, T. T., Kong, Y. Y., & Zeng, F. G. (2005). Auditory temporal processing in normal – hearing individuals and in patients with auditory neuropathy. *Clinical Neurophysiology, 116,* 669-680.

Moore. B. C. J. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear and Hearing*. *17,* 133-161.

Moore, B.C. J. & Oxenham, A. J. (1998). Psychoacoustic consequences of compression in the peripheral auditory system. *Psychology Review, 105,* 108-124.

Nienhuys, T. G. & Clark, G. M. (1978). Frequency discrimination following the selective destruction of cochlear inner and outer hair cells. *Science*, *199*, 1356-1357.

Oxenham, A. J. & Bacon, S. P. (2003). Cochlear compression: perceptual measures and implication for normal and impaired hearing. *Ear and Hearing, 24*, 352-366.

Peters, K.P. (2006). "Emotion perception in speech: Discrimination, identification, and the effects of talker and sentence variability". *Independent Studies and Capstones. Paper 121.* Program in Audiology and Communication Sciences, Washington University School of Medicine.

Petrushin, V. A. (2000). Emotion recognition in speech signal: experimental study, development, and application. In: *Proceedings of the International Conference on Spoken Language Processing, 2,* 222-225.

Prosen, C. A., Moody, D. B., Stebbins, W. C., & Hawkins, J. E. (1981). Auditory intensity discrimination after selective loss of cochlear outer hair cells. *Science*, *212*, 1286-1288.

Rance, G., Beer, D. E., Cone-Wesson, B., Shepherd, R. K., Dowell, R. C., King, A. M., Rickards, F.W., & Clark, G. M. (1999). Clinical findings for a group of infants and young children with auditory neuropathy. *Ear and Hearing*, *20*, 238–252.

Rance, G., Cone-Wesson, B., Wunderlich, J., & Dowell, R. (2002). Speech perception and cortical event related potentials in children with auditory neuropathy. *Ear and Hearing*, *23*, 239–253.

Rance, G., McKay, C., & Grayden, D. (2004). Perceptual characteristic of children with auditory neuropathy. *Ear and Hearing, 25,* 34-46*.*

Ryan, A. & Dallos, P. (1975). Effect of absence of cochlear outer hair cells on behavioural auditory threshold. *Nature*, *253*, 44-46.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227-256.

Starr, A., McPherson, D., Patterson, J., Don, M., Luxford, W., Shannon, R., Sininger, Y., Tonakawa, L., & Waring, M. (1991). Absence of both auditory evoked potentials and auditory percepts dependent on timing cues. *Brain, 114,* 1157-1180.

Starr, A., Picton, T. W., Sininger, Y., Hood, L.J., & Berlin, C.I. (1996). Auditory neuropathy. *Brain*, *119* (3), 741–753.

Williams, C. E., & Stevens, K. N. (1972). Emotions and Speech: Some Acoustical Correlates. *Journal of the Acoustical Society of America, 52,* 1238-1250.

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997). Deficit in auditory temporal and spectral resolution in language-impaired children. *Nature*, *387*, 176-178.

Yildirim, S., Bulut, M., Lee, C., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., & Narayanan, S. (2004). An acoustic study of emotions expressed in speech. In: Proceedings of the International Conference on Spoken Language Processing (pp. 2193-2196). Korea.

Zeng, F. G., Kong, Y. Y., Michalewski, H. J., & Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *Journal of Neurophysiology*, *93*, 3050–3063.

Zeng, F. G., Oba, S., Garde, S., Sininger, Y., & Starr, A. (1999). Temporal and speech processing deficits in auditory neuropathy. *Neuroreport*, *10*, 3429–3435.

Zeng, F. G., Oba, S., & Starr, A. (2001). Suprathreshold processing deficit due to desynchronous neural activities in auditory neuropathy. In Breebaart, D. J., Houstma, A. J. M., Kohlrausch, A., Prijis, V. F., & Schoonhoven, R. (Eds.): *Physiological and Psychological Bases of Auditory Function* (pp. 365-372). The Netherlands: Shaker Publishing.