Construction of Hindi Speech Stimuli for Eliciting Auditory Brainstem Responses

Mohammad Shamim Ansari¹ · R. Rangasayee²

Received: 20 June 2015 / Accepted: 20 June 2016 / Published online: 8 July 2016
© Association of Otolaryngologists of India 2016

Abstract Speech-evoked auditory brainstem responses (spABRs) provide considerable information of clinical relevance to describe auditory processing of complex stimuli at the sub cortical level. The substantial research data have suggested faithful representation of temporal and spectral characteristics of speech sounds. However, the spABR are known to be affected by acoustic properties of speech, language experiences and training. Hence, there exists indecisive literature with regards to brainstem speech processing. This warrants establishment of language specific speech stimulus to describe the brainstem processing in specific oral language user. The objective of current study is to develop Hindi speech stimuli for recording auditory brainstem responses. The Hindi stop speech of 40 ms containing five formants was constructed. Brainstem evoked responses to speech sound /da/ were gained from 25 normal hearing (NH) adults having mean age of 20.9 years (SD = 2.7) in the age range of 18–25 years and ten subjects (HI) with mild SNHL of mean 21.3 years (SD = 3.2) in the age range of 18–25 years. The statistically significant differences in the mean identification scores of synthesized for speech stimuli /da/ and /ga/ between NH and HI were obtained. The mean, median, standard deviation, minimum, maximum and 95 % confidence interval for the discrete peaks and V–A complex values of electrophysiological responses to speech

stimulus were measured and compared between NH and HI population. This paper delineates a comprehensive methodological approach for development of Hindi speech stimuli and recording of ABR to speech. The acoustic characteristic of stimulus /da/ was faithfully represented at brainstem level in normal hearing adults. There was statistically significance difference between NH and HI individuals. This suggests that spABR offers an opportunity to segregate normal speech encoding from abnormal speech processing at sub cortical level, which implies that alterations in brainstem responses have clinical significance to identify the subjects with possible processing disorders.

Keywords Speech stimulus · Speech perception · Speech discrimination · Speech encoding · Auditory brainstem responses

Introduction

In human communication, oral mode is the principal avenue and speech is the most important signal to hear and use for interface, interaction and understanding. That is the reason that individuals with hearing loss and other auditory processing disorders have chief complaints of difficulty in understanding oral communication. However, the primary audio logical test procedure employed, i.e. pure-tone audiometry provides only a partial picture of the patient's communicative difficulties and does not offer any direct information about their ability to hear and understand speech [7].

Hence, to approximate the difficulty to understand speech, results obtain on pure tone audiometry are supplemented with speech audiometry. The speech test procedures and speech materials supply more relevant information and

✉ Mohammad Shamim Ansari
msansari5000@yahoo.com

¹ Ali Yavar Jung National Institute for the Hearing Handicapped, K.C. Marg, Bandra (W), Mumbai, Maharashtra 400050, India

² Dr. S.R. Chandrasekhar Institute of Speech and Hearing, Hennur Main Road, Lingarajapuram, Bangalore, Karnataka 560084, India

knowledge about listening and communicative abilities of individuals. The speech audiometry is generally regarded as clinically more acceptable and superior than pure-tone audiometry for identifying patients with poor auditory analytical capability and helpful in accurate categorization of auditory lesions and deficits [36].

Therefore, pure-tone and speech audiometry forms the essential repertoire of the audio logical test protocols for identification of hearing impairment, differentiations of auditory disorders, selection, fitting and verification of hearing devices. However, these tests require a definite age for administration, attention or cognitive skill, active participation from subjects, familiarity of the stimulus and knowledge of language. Further, nature and redundancy of stimulus, administration protocol and recording of response method in these tests have some inherent factors that affect the accuracy of results.

Therefore, these tests are not considered as a reliable and valid, especially in very young children and difficult to test population. To overcome the limitations of behavioral test procedures in assessment of auditory functions, objective instruments and test procedures like otoacoustic emissions (OAEs) and auditory brainstem responses (ABRs) were developed [8].

Measurement of Neurophysiologic Responses of Auditory System

Auditory brain stem responses (ABRs) are measured using electrodes at scalp that picks up electrical potentials generated by the synchronous activity of populations of neurons in the brain stem in response to click or tones [14]. These aggregate neural responses are recorded inattentively and non-invasively. ABR measurement has been considered as an decisive answer to all clinical challenges imposed by the behavioral audiometry and evolved as a excellent tool to use it as a metric for determining auditory thresholds and detecting auditory pathologies in children and difficult to test population [9, 31, 34].

Limitations of Neurophysiologic Response with Click Stimulus

Though, the ABRs to click or tone stimuli have been instrumental in defining the basic response patterns, with five to seven peaks for the stimulus onset in the initial 10 ms of stimulus presentation and proven to be a clinically useful tool to assess auditory function [9, 10]. But, the use of click stimuli to elicit ABR has limitation of being electrically generated and behaviorally irrelevant sound that brain encounters outside the clinical setup (e.g., speech and music, non speech vocal sounds, and environmental sounds, etc.).

Moreover, the complex sounds include both sustained and transient features, thus response to click and tones do not predict the sub cortical response to a complex sound [17, 29, 33]. Hence, it provides inadequate estimation of auditory function and communicative difficulty. Because of these shortcomings, auditory neuroscience has transitioned to use sounds that are more complex and considered to be more advantageous and superior for clinical assessment of auditory system.

Necessity for Neurophysiologic Responses to Speech Stimulus

ABR measurements to click and tones evolved as a standard procedure to replace the behavioral pure tone audiometry in very young children and difficult to test population to estimate the hearing thresholds. However, the click and tone burst ABR provide limited information about the individual's speech processing abilities. Additionally, these individuals cannot perform on speech test procedure due to very young age, limited vocabulary and deficient language or unable to understand speech and give opinion about benefit with the management strategies.

Therefore, speech processing ability of auditory system cannot be estimated and predicted pragmatically in them. Hence, there is need to acquire speech processing information at brainstem. This may augment our knowledge to understand speech processing in brain and help us to establish a complete objective test battery to supplement the behavioral test procedures especially for very young children and difficult to test population.

This necessitate search for a objective mean to measure the auditory system's ability to process speech to determine the locus of the auditory processing deficit resulting from hearing impairment, language processing disorders and to evaluate and monitor the perception of phonetic information with amplification devices and intervention strategies.

Significance of Neurophysiologic Responses to Speech Stimulus

Recent researches have established that complex stimuli like music, complex tones and speech stimuli (Table 1) have been used to elicit ABRs [3, 6, 30]. But, speech stimulus is particularly useful, as it can provide clues as to how temporal and spectral features are transcribed into neural code in the brainstem [32]. The spABRs have shown considerable utility in the study of populations where auditory function is of interest (e.g., auditory experts such as musicians, persons with hearing loss, auditory processing and language disorders).

The assessment of auditory encoding of speech has elucidated the relationship between speech perception and

Table 1 Showing speech stimuli used literature for eliciting ABRs Adopted from Skoe and Kraus [32]

Study material	Speech syllables used	Researchers
Vowels		
Synthetic	/a/,/u/	Krishnan (2002)
Natural	/ε/,/ʌ/,/i/,/a/,/æ/,/ʌ/,/u/	Greenburg et al. (1980), Dajani et al. (2005) and Aiken and Picton (2006, 2008)
Consonant–vowel syllables		
Synthetic	/da/	Cunningham et al. (2001), Plyler and Ananthanarayan (2001), King et al. (2002), Wible et al. (2004, 2005), Russo et al. (2004, 2005), Kraus and Nicol (2005), Johnson et al. (2007, 2008), Banai et al. (2005, 2009), Burns et al. (2009), Chandrasekaran et al. (2009) and Parbery-Clark et al. (2009a)
	/ba/	Akhoun et al. (2008a, b)
	ba-da-ga continuum	Plyer and Ananthanarayan (2001), Johnson et al. (2008) and Hornickel et al. (2009b)
Natural	Mandarin pitch contours	
	/yi/	Krishnan et al. (2005) and Xu et al. (2006)
	/mi/	Wong et al. (2007) and Song et al. (2008)
Hybrid	/ya/with linearly rising and falling pitch contours	Russo et al. (2008)

specific auditory dysfunctions, including sensory loss, learning problem and auditory processing disorder (APD). The relationship between brainstem encoding of click and speech signals in normal children and those with learning impairment have reflected separate neural processes and processes involved in encoding speech are found to be impaired in children with learning problems [33].

The spABR have evidenced a poor physiologic representation of speech encoding in children with language, literacy, reading, and learning deficits [1, 16, 19]. Children with known language-based learning problems exhibited delayed latencies compared to their normal learning peers [16, 19]. In addition, children who were poorer readers had prolonged latencies, poorer waveform morphology and weaker spectral encoding compared to children who were better readers [1, 11].

Perhaps most imperatively, The brainstem responses to speech also offers contributing information in monitoring auditory training progress and effects of rehabilitation strategies like validating the fitting of amplification devices. A computer-aided training program for the children with learning problems (LP) to improve their speech processing abilities was implemented. The speech stimuli were presented to the children with some exaggeration of critical elements of the speech sounds. For example, cue enhancements (such as longer stop gaps and more emphasis on syllables) helped the LP children to process the speech stimuli. Then, as the child showed progress in perceiving the speech stimulus accurately, the exaggerated elements of the speech sound were gradually decreased until the speech sound could be understood at a normal rate of speech. The changes in the speech processing abilities of the children were apparent in the electro physiologic responses [4].

However, the physiological consequence of such kind of training needs to be thoroughly understood. Thus, testing children before and after undergoing such training offers an ideal opportunity to examine neural plasticity at the level of the auditory brainstem with speech. Further, showed that the measurement of the frequency following responses to speech provide additional information of human auditory system's ability in speech processing due to aging [38] and assessment of auditory processing in hearing aid and cochlear implant users [15].

These studies show a trend that difficulties in language, literacy, reading, improvement in learning with training and amplification affects the sub cortical representation of speech and changes in the response latencies recordable in the difference of few milliseconds tend to be associated with physiological condition of normal and clinical population. Thus, measures of the brainstem response to speech are clinically significant because it allows professionals to understand how the auditory system transmits sound from cochlea to the brainstem.

The most common speech sound /da/ has been investigated for brainstem encoding and reported to reflect some of the acoustic–phonetic characteristics of speech with remarkable precision at sub cortical level in English speakers [3, 15, 18, 30, 39]. The presence of speech sound /da/ in multiple languages of the world [27] makes it a nearly universal option to use and record sub-cortical encoding of speech sound /da/.

Therefore, it has been suggested that the evoked ABR with the syllable/da/ can be used universally for the assessment of auditory function. The BioMARK (biological marker of auditory processing) procedure has been developed based on the above mentioned studies and introduced in the United States (US) as a tool for the assessment of speech processing in the brainstem in

clinical population such as the diagnosis of auditory processing disorders and learning problems.

Considerations of Acoustic Parameters of Hindi Speech Sound

Though, the constituents of the syllable *ldal* are characteristic of multiple languages including the Hindi [27]. However, under the theoretical knowledge of acoustic phonetic differences in English and Hindi languages such as voice onset time (VOT). When stop consonants such as *ldal* are produced by English speakers, the vibration of the vocal folds starts at a short lag after the release of the consonant. In contrast, in Hindi sounds, voicing starts earlier, and prior to the release of the consonants (for example, in naturally produced Hindi sound *ldal*, VOT is -140 to -60 ms [26]. Further, it has been reported that infants of 10–12 month old can discriminate the sounds that are linguistically relevant in their native language, suggesting that psycho physical discrimination of consonant is affected by language input [37].

Additionally, the recent studies have shown that experience with language have effects in sub cortical encoding of specific elements of speech sounds. Mandarin speakers encode pitch patterns of mandarin sound more robustly than English speakers that convey linguistic meaning in Mandarin, but not in English [22, 23]. Furthermore, long-term music experiences selectively enhance specific stimulus features in brainstem activity [40].

Hence, we assume that the existence of acoustic differences in Hindi sound may exhibit dissimilar sub cortical processing than reported in the literature. Furthermore, it is always desirable that culturally and linguistically sensitive speech material should be used in any given population. But, there exist no Hindi speech stimulus or in other Indian language to record spABR. Hence, there is a need to develop speech stimuli of Hindi language to record neurophysiologic responses.

Objective of the Study

The purpose of this study was to develop the corpus of speech stimuli for recoding electrophysiological responses in Indian population.

Subjects and Methods

Research Design

This prospective exploratory survey was conducted to develop the Hindi speech stimulus corpus at

electrophysiological lab of Department of Audiology to study the neuronal encoding of speech at brainstem. The study was accorded necessary ethical clearance from Institutional ethical board. Informed consent was obtained from the individual subjects.

Selection Criterion of Speech Stimulus

In order to choose most appropriate Hindi speech syllables to record auditory electrophysiological processing the current study theorize the following criterion.

- (a) Speech syllables should be language free, yet can represent the linguistic component of spoken language to use it as a language specific stimulus.
- (b) They should contain sufficient distinctive feature (to be categorically perceived) and can be presented at supra threshold intensities.
- (c) They can be constructed of equal duration and intensity to prevent the potential cues perceptually.
- (d) The frequency spectrum of speech sound can be analyzed and compared with electrophysiological results.
- (e) It should be of brief duration and contain adequate categorical boundaries for children as well adults to perceive. It should be attractive to infants.

In essence it was aimed to develop the stimulus that can be used as a performance indicator of auditory function behaviorally as well as electro physiologically.

Selection of Speech Stimulus

Speech is characterized by particular combinations of acoustic properties occurring in specific frequency ranges that set it apart from other acoustic stimuli [35]. Firstly, speech shows an alternation between relatively intense segments corresponding to vowels and weaker segments corresponding to consonants. This more or less regular amplitude modulation occurring at a 3–4 Hz rate is essential for speech understanding. Secondly, the spectral envelope of speech shows pronounced maxima (corresponding to formants) interleaved with minima at intervals of 1000–1200 Hz. Thirdly, speech shows both nearly periodic segments corresponding to sonorant such as vowels and irregular (noise-like) segments corresponding to obstruent such as fricative consonants and bursts of stop consonants [12]. Thus, speech is characterized by a triple alternations, amplitude envelope, spectral envelope and fine spectral structure.

Hence, in order to obtain the realistic picture of the neural encoding of speech, it is necessary to use stimuli possessing all of these characteristics. Thus, we decided to select Hindi plosives CV syllables that possess these

characteristics abundantly needed for the neural encoding. The stop voiceless CV syllable *lpal*, *ltal* and *lkal* and voiced syllables *lbal*, *ldal*, and *lgal* were chosen as various previous studies have also indicated that stop consonants have considerable phonetic information and provide consistent robust and reliable traces as compared to other syllables for electrophysiological measurements.

Synthesis of the Stop Consonants–Vowel syllable

The corpus consisting of a set of nonsense stops consonant–vowel (CV) voiceless syllable *lpal*, *ltal* and *lkal* and voiced syllables *lbal*, *ldal*, and *lgal* were created according to methods described in previous publications [2, 15]. A native male and a female Hindi speaker served as subjects and narrated these stimuli. Each CV utterance was written on cards. The cards were then shuffled and presented to the subjects who were asked to read what was written on each card keeping his intonation same on each syllable.

The total 12 utterances were recorded in a sound-treated room using a Newman 189 microphone, MACKIE SR32-4 sound table, M-AUDIO 101LT card sound without the use of Dolby or DBX. Recording was done on Sony H Sound Forge 6.0 recording software. The stimuli were edited using Sony H Vegas 4.0 editing software. All the speech sounds were recorded for same time length of 100 ms, RMS-balanced and digitized at 10 kHz using a low-pass filter with a cut-off frequency of 4.8 kHz and then saved as wave file.

The stimuli were digitally manipulated in software, to construct each stimulus of equal duration of 40 ms containing five-formant in speech syllable, at a sampling rate of 10 kHz. The stimuli were created to include an onset burst frication at F_3 , F_4 , and F_5 during the first 10 ms followed by 30 ms F_1 and F_2 transitions ceasing immediately before the steady-state portion of the vowel. Rise and

fall times were 5 ms. The vowel *lal* was shortened in order to increase the stimulation as higher rate activate the system better and elicits robust responses to acoustic onset. Although, the stimulus does not contain a steady-state portion but contained key acoustic phonetic features to be psycho acoustically perceived as a CV syllable. In this manner, from each original CV syllable, only 40 ms duration containing first five formants of the transient portion of the syllable were generated and transported on a compact disc.

Acoustic Properties of Synthesized Speech Stimuli

The acoustic analyses of speech stimuli were included: F_0 and F_1 frequency at the onset of the vowels immediately following the stop consonants; relative duration of the stop consonants and their adjacent segments (see Figs. 1, 2). Table 2 depicting acoustic properties F_0 , F_1 , F_2 , F_3 , F_4 and F_5 of stop voiceless CV syllable *lpal*, *ltal* and *lkal* and voiced syllables *lbal*, *ldal* and *lgal*.

Assessment of Categorical Perception

Human hearing mechanism identify over a dozen of speech phonemes per second, recognize the words they constitute almost immediately and understand the message generated by the sentences they form. One of the mechanisms thought to be responsible to accomplish speech sound perception is categorical perception. Categorical perception is a sensory phenomenon whereby a physically continuous dimension is perceived as discrete categories, with abrupt perceptual boundaries between categories and poor discrimination within categories.

Hence, once the waveform and spectrum data for speech stimuli were available it was desired to assess the categorical perception of these sounds. The categorical

Fig. 1 Fundamental frequency values of voiceless and voiced plosive CV *lpal*, *lbal*, *ltal*, *ldal*, *lkal* and *lgal* of the normal-hearing Hindi speaker

Fig. 2 Formant frequency values of F_1 , F_2 and F_3 transition of voiceless and voiced plosive CV pal, lbal, lta, ldal, lkal and lgal of the normal-hearing Hindi speaker

Table 2 Depicting acoustic properties F_0 , F_1 , F_2 , F_3 , F_4 and F_5 of stop voiceless CV syllable lpal, lta and lkal and voiced syllables lbal, ldal, and lgal

	F_0 Hz	F_1 Hz	F_2 Hz	F_3 Hz	F_4 Hz	F_5 Hz
lpal	167	690	1350	2470	3800	4400
lbal	171	740	1540	2780	3700	4300
ltal	179	810	1400	2800	3770	4500
ldal	113	720	1400	2450	3590	4600
lkal	137	800	1600	2400	3780	4300
lgal	117	716	1280	2280	3650	4250

perception of speech sounds were studied in two steps. The first step involved estimation and normalization the amplitudes of the burst and formants of stimuli to eliminate the amplitude cue in perception. In the second step these synthesized stimuli were subjected to identification task to assess the phonetic perception in the normal hearing and hearing impaired listeners.

Estimation and Normalization of Amplitude of Speech Stimuli

Since the synthesized speech stimuli can still have amplitude cues, therefore, to eliminate the amplitude differences as a cue for phonetic discrimination, the stimuli were loudness balanced. For this purpose, the each stimulus was presented in trains of four stimuli, separated by 21 ms inter stimulus intervals (ISI: time within a train between stimulus offset and subsequent stimulus onset). A JWin All Terrain JX-CD588 disc man was used to control the intensity of stimulus delivery. Then each trimmed speech sound was loudness-balanced with reference to the loudness of Synthetic vowel/a/to six bilaterally normal hearing children and adults to eliminate the loudness differences as cue for phonetic discrimination. Synthetic vowel lal which

is used in Dr. Speech software from Tiger Electronics Inc USA was presented at constant intensity level of 70 dB SPL through the GSI 61 audiometer (PC based delivery system) for loudness-balancing.

The each sound was presented in the sound field at random intensities and alternated with Synthetic vowel lal which was presented at 70 dB SPL. The subjects were asked to compare the test phoneme intensity whether, they sounded louder or softer than lal or equally as loud as the lal. If the given intensity was scored “louder than lal” for three consecutive presentations, this intensity was considered to be loud and defined the upper fence for the time being. The same was true for intensities that were “softer than lal” for three consecutive presentations and that defined the temporary lower fence.

The step size decreased from 5 dB in the beginning to 1 dB at the end of the test. In this way the intensity range between the upper and lower fence for each speech sound was narrowed until all the remaining intensities resulted in ambiguous comparisons with the lal. Then the each test speech sound was presented six times in random order alternating with the same lal at 70 dB SPL. For each presentation the scores was recorded, and at the end, the intensity with most frequent score “sound equally loud as lal” was saved as loudness balanced intensity for that particular speech sound. The intensity of all speech sounds was modified according to this algorithm.

Thus, all the speech sound was loudness balanced with reference to lal at 70 dB SPL and with precision of 1 dB. Finally the lal was loudness balanced according to the same algorithm with reference to a calibrated 1 kHz narrow band noise at 70 dB HL, and all speech sound were adjusted accordingly. In consequence, the intensity of speech sound can be expressed in dBHL (reference 1 kHz narrow band noise).

Though, the precision of this loudness balancing was 1 dB, the temporal profile of the speech sound still may

contain intensity cues that would help discrimination between two speech sounds. In order to eliminate possible further intensity cues, a gain was added to intensity of all speech sounds. This gain varied depending upon upper and lower limit which was defined by tester. Limits +3 and –3 dB respectively were set as default values. This introduced the random variability in the intensity of the speech sound that overrules any possible intensity differences between speech sounds.

Assessment of Identification of Synthetic Speech

Then, an identification task for these synthesized and normalized speech sounds was designed in which participants heard the target stimuli one at a time and indicated whether each stimulus sounded like lpal, ltal, lkall, lbal, ldal, and lgal. The six synthesized speech syllables were arranged in six randomly-ordered blocks yielding a total of 36 trial stimulus. These were presented to 40 subjects at 80 dB HL to obtain the identification scores.

The subjects were divided into two groups, 20 children with mean age of 9.3 years (SD = 2.2) in the age range of 6–13 years and 20 adults with mean of 21.3 years (SD = 3.2) in the age range of 18–25 years. The each group was further divided into two subgroups of ten subjects with normal hearing and ten subjects with mild (30–40 dB HL) sensorineural hearing loss of almost flat audiometric configuration. The recruited subjects in the study did not had any reported cognitive or language deficit. All testing was performed in quiet room where the subject listened to the stimuli over headphones.

It can be noted from the Table 3, that voiced ldal and lgal have highest identification scores in both children and

adults groups. The statistical analysis has revealed no statistically significant difference in identification of voiced and voiceless sounds between the normal hearing children and adults at value 0.842 ($p \leq 0.05$). However, there was a statistically significant difference in speech identification score between hearing impaired children and adult at p value 0.032 ($p \leq 0.05$). The identification scores of voiced sounds ldal and lgal have the statistically significant difference between normal hearing and hearing impaired in both children and adults. This implies that, in presence of hearing loss the perception of these sounds are more affected in children and adults.

Further, the stimuli ldal and lgal were again examined in pairs of same or different perception in 20 atypical populations (10 children and 10 adults). Each stimulus presentation trial block consisted of two stimulus pairs. One pair contained two of the same stimuli lda–dal and the other pair contained the same stimulus and a different stimulus lda–gal. The purpose of this study was to establish that ldal and lgal sound are perceptually distinguishable.

The above Table 4 indicates that the voiced stop consonants ldal and lgal are perceptually distinguishable and categorical boundaries exist to be used as stimulus for assessing cognitive skills in children and adults. Further, these syllables were selected because they are the earliest speech sounds produced by all developing children during cooing and canonical babbling stage across all linguistic, cultural and socioeconomic boundaries universally. Moreover, the speech stimulus ldal and lgal can be distinguished based on their differing second (F2) formant trajectories [24, 25]. Additionally, Infants of 2 months old have been shown to be able to discriminate stops/glides consonants [5].

And, most importantly, stop consonants reported to have considerable phonetic information that provides robust and reliable traces and their perception is particularly vulnerable to background noise in both normal and clinical populations. Further, these specific stimuli have shown that children with learning problems have difficulty in perceiving stop consonants in general [12] and lgal stimulus specifically [21].

Table 3 Showing the average identification score in percentage of each 40 ms synthesized voiced and voiceless speech stimuli by children 6–13 years and adults 18–25 year with normal hearing and hearing impairment

Speech sound category	Speech sound	Identification score			
		Normal hearing		Hearing impaired	
		Children (n=10) (%)	Adults (n=10) (%)	Children (n=10) (%)	Adults (n=10) (%)
Voiced stop	lbal	58	67	36	41
	ldal*	79	89	48	57
	lgal*	84	90	61	65
Voiceless stop	lpal	71	59	31	53
	ltal	59	64	23	36
	lkall	64	67	19	33
ANOVA at p value 0.05		0.0864		0.032	

* Statistically significant at p value 0.012

Table 4 Showing discrimination responses of (20 subjects) to ldal and lgal as pair of same and different stimulus

Response	Speech sound presentation in pair				Total
	lda-dal	lda-gal	lga-dal	lga-gal	
lda-dal	19	2	4	1	26
lda-gal	1	13	0	0	14
lga-dal	0	1	13	1	15
lga-gal	0	5	3	18	25
Total	20	20	20	20	80

Calibration of Synthesized Speech Stimulus

Constructed stimuli were uploaded using the custom stimulus option in the Intelligent Hearing Systems (IHS) Smart EP software (version 2.39). Stimulus levels were calibrated in dBnHL using a 2 cm³ DB-0138 coupler, Bruel & Kjaer Type 2260 Investigator sound level meter, and ½ in. microphone. The two speech stimulus, referred as l_{dal} and l_{gal} were considered to elicit the auditory brainstem responses.

Validation of Synthesized Speech Stimulus

The synthesized Hindi phoneme l_{dal} was assessed for its efficacy to elicit neurophysiologic responses from 35 subjects divided into two groups. Group A (NH) consisting of 25 (14 male and 11 female) normal hearing (NH) adults having mean age of 20.9 years (SD = 2.7) in the age range of 18–25 years. Group B comprised of ten subjects (6 male and 4 female) with mean of 21.3 years (SD = 3.2) in the age range of 18–25 years, had mild SNHL (HI) of almost flat pure tone audiometric configuration. None of the subjects reported any other neurological, otological and medical problems.

The subjects were sited comfortably on the reclining chair. The Ag–AgCl electrodes with surface contact impedance of <5 kΩ, positioned centrally on the scalp, at Cz, behind the right mastoid (reference) and on the forehead (ground). Stimuli were presented through ER-3 insert earphones into the right ear at a rate of 11.1 per sec at comfortable listening level of 65 dB SL relative to the threshold at 1000 Hz. The sampling rate was 20,000 Hz and responses were online band passed filtered from 100 to 3000 Hz at 12 dB/octave. Trials with eye-blinks or other motion artifacts greater than ±35 μV were rejected online. Two traces of 2000 sweeps were collected at alternating polarity. The recording window was 50 ms starting 10 ms prior to stimulus onset. Waveforms were averaged online.

The obtained waveform peaks (see Fig. 3) were labeled as V, A, C, D, E, F, and O. The discrete waves were analyzed to their absolute latencies and amplitude and V–A complex was analyzed for latency, amplitude and slope (VA amplitude/VA duration) are shown in (Table 4) for both the groups. All the subjects in group A exhibited auditory brainstem responses to Hindi stimulus l_{dal} as shown in Fig. 3.

The results indicated that the discrete wave V, A, C, F and O were observed in 100, 90, 80, 70 and 10 % respectively in HI, whereas NH individuals exhibited 100 % for wave V, A and C, and 98 and 94 % for F and O correspondingly. Table 5 showing mean, standard deviation (in parenthesis), median, minimum, maximum and 95 % confidence interval (CI) values of sub cortical

Fig. 3 Waveform of auditory brainstem responses to Hindi stimulus l_{dal} in one of the subject

responses of individual peak latencies and amplitudes in adults with normal hearing and hearing impairment.

Discussion

The purpose of the study was to construct the culturally and linguistically sensitive Hindi speech stimulus to record the brainstem encoding.

Selection and Construction of Hindi Speech Stimuli

The corpus consisted of a set of synthesized nonsense stop CV voiceless syllables l_{pal}, l_{tal} and l_{kal} and voiced syllables l_{bal}, l_{dal}, and l_{gal} of 40 ms duration containing five formants. The stimuli included an onset burst frication at F₃, F₄, and F₅ during the first 10 ms followed by 30 ms F₁ and F₂ transitions ceasing immediately before the steady-state portion of the vowel. Rise and fall times were 5 ms. The acoustic parameters of these sounds are depicted in Table 2.

Then, an identification task for these synthesized speech sounds was performed on children and adults with normal hearing and hearing impairment. The voiced l_{dal} and l_{gal} were found to have highest phonetic identification scores in both children and adults (Table 3). However, the identification scores of voiced sounds l_{dal} and l_{gal} have the statistically significant difference between normal hearing and hearing impaired in both children and adults at *p* value of 0.012 (≤0.05).

The stimuli l_{dal} and l_{gal} were further examined for discrimination in 20 normal hearing populations (10 children and 10 adults). The voiced stop CV syllable l_{dal} and l_{gal} were found to be perceptually distinguishable (Table 3). Thus, these stimuli were considered as most ideal input signals to record the neurophysiologic responses. Further, these sounds provide consistent robust and reliable traces during electrophysiological recording and can also be used as a performance indicator of auditory

Table 5 Showing mean, standard deviation (in parenthesis), median, minimum and maximum and 95 % confidence values of sub cortical responses peaks latencies and amplitudes in adults with normal hearing and hearing impairment

	Peaks	Group	N (%) of peak detection	Minimum (ms)	Maximum (ms)	Median (ms)	Mean (SD) (ms)	95 % CI levels		<i>p</i> value	
								Min	Max	<i>t</i> test	ANOVA
<i>Hindi speech stimulus—evoked ABRs</i>											
Latency (ms)	V	NH	50 (100)	5.77	10.96	7.03	6.78 (0.43)	6.18	7.73	0.0347*	0.0386*
		HI	10 (100)	4.22	12.67	7.31	7.55 (1.76)	6.72	8.37		
	A	NH	50 (100)	6.79	12.09	7.95	8.04 (0.55)	7.24	8.19	0.0425*	0.0357*
		HI	9 (90)	5.46	15.49	8.75	9.10 (2.08)	8.12	10.07		
	C	NH	50 (100)	14.01	20.23	17.01	18.81 (0.97)	16.01	19.21	0.0758	0.0619
		HI	8 (80)	16.19	24.26	21.13	22.78 (0.85)	20.10	22.16		
	F	NH	49 (98)	34.30	43.81	37.06	38.82 (0.20)	29.16	32.21	0.0763	0.0654
		HI	7 (70)	36.37	45.73	38.67	39.48 (0.51)	36.35	40.02		
	O	NH	47 (94)	43.17	49.23	47.31	48.03 (0.96)	45.27	49.01	–	–
		HI	1 (10)	46.25	49.89	48.11	49.17 (0.82)	47.16	49.79		
Slope	V–A	NH	10 (100)	1.01	1.98	1.23	1.46 (0.19)	1.31	1.69	0.6447	0.6472
		HI	10 (100)	0.71	3.17	1.55	1.40 (0.65)	1.25	1.85		
	V/A	NH	10 (100)	0.87	1.34	0.91	1.21 (0.37)	0.87	1.27	0.0174*	0.0214*
		HI	10 (100)	0.28	1.18	0.43	0.71 (0.26)	0.59	0.83		
Amplitude (μV)	V	NH	50 (100)	0.16	0.63	0.28	0.29 (0.15)	0.27	0.48	0.5967	0.9359*
		HI	10 (100)	0.06	0.70	0.20	0.23 (0.21)	0.20	0.39		
	A	NH	50 (100)	–0.93	–0.13	–0.47	–0.37 (0.31)	–0.93	–0.13	0.0062*	0.0061*
		HI	9 (90)	–0.80	–0.09	–0.43	–0.43 (0.19)	–0.51	–0.33		
	C	NH	50 (100)	–0.36	–0.21	–0.36	–0.37 (0.26)	–0.36	–0.21	0.0043*	0.0042*
		HI	8 (80)	–0.89	–0.16	–0.50	–0.31 (0.17)	–0.60	–0.39		
	F	NH	49 (98)	–1.06	–0.01	–0.37	–0.34 (0.23)	–0.48	–0.24	0.0592	0.0606
		HI	7 (70)	–0.39	–0.07	–0.31	–0.30 (0.23)	–0.39	–0.07		
	O	NH	47 (94)	–1.06	–0.01	–0.31	–0.33 (0.18)	–0.01	–0.47	–	–
		HI	1 (10)	–	–	–	–	–	–		
Area	V–A	NH	10 (100)	0.50	1.44	0.68	0.72 (0.29)	0.80	1.06	0.0314*	0.0241*
		HI	10 (100)	0.25	1.38	0.51	0.38 (0.25)	0.58	0.86		
	VA	NH	10 (100)	1.47	3.56	4.54	5.86 (1.93)	1.98	6.25	0.0467*	0.0389*
		HI	10 (100)	1.32	3.01	3.21	3.67 (1.67)	1.49	3.45		

Wave “O” was observed only in one HI subject hence not considered for analysis

* $p \leq 0.05$

function behaviorally. Therefore, these sounds were selected and constructed.

The sounds were uploaded in the Intelligent Hearing Systems (IHS) Smart EP software (version 2.39). Stimulus output levels were calibrated in dBnHL. Given that, the perception of stimulus lgal is affected in presence of noise even in the normal hearing persons. Since, the perception of lgal in presence of noise was not studied in the subjects. Therefore, the current study used only lgal stimulus for eliciting sub cortical responses.

Validation of Synthesized Speech Stimulus

It has been reported that the brainstem structures are organized to encode rapid temporal and spectral

changes in auditory signal with extreme accuracy [28]. The latency measures provide information about the precision with which the brainstem nuclei synchronously respond to acoustic stimulus while amplitude measures furnish information about robustness of the response of the brainstem nuclei for the acoustic stimulus [15].

The results indicated that the discrete wave V, A, C, F and O were observed in 100, 90, 80, 70 and 10 % respectively in HI, whereas NH individuals exhibited 100 % for wave V, A and C, and 98 and 94 % for F and O correspondingly. The mean values of responses across normal listeners and individuals with SNHL were analyzed and compared to provide insight into speech processing abilities in both the groups.

Table 6 Showing mean, standard deviation (in parenthesis), median, minimum and maximum and 95 % confidence values of sub cortical responses peaks latencies in Indian normal hearing adults

	Peaks	Mean (SD) (ms)	Present study		Russo et al. (2004)	
			95 % confidence level		95 % confidence level	
			Minimum	Maximum	Minimum	Maximum
Latency (ms)	V	6.78 (0.24)	6.18	7.83	6.63	6.74
	A	7.55 (0.35)	7.24	8.19	7.51	7.68
	C	18.81 (0.97)	16.01	19.21	18.35	18.67
	F	39.48 (0.51)	36.35	40.02	39.45	39.69
	O	49.03 (0.68)	45.27	49.01	48.14	48.36

Analysis and Comparisons of spABR Mean Values in NH and HI

The transient portion of spABR exhibited presence of an initial wave V at a mean latency of 6.78 ms in NH and 7.55 ms in HI subjects analogous to the wave V elicited by a click stimulus. However, this is more than the documented mean latency between 5 and 6 ms of peak V for click stimulus [21]. The subsequent peak latencies were calculated with reference to wave V. In NH subjects, peak A appeared at negative trough immediately after wave V at a 8.04 ms, wave C emerged at 18.81 ms, wave F crop up at 38.82 ms and wave O was visible at a mean latency of 48.03 ms. While, in HI individuals only peak A, C and F were observed at 9.10, 22.78 and 39.48 ms respectively. The individual with HI had delayed latencies compared to NH. Thus, there exists statistically significant difference (p value <0.05) between the NH and HI for the discrete peaks latency value.

The onset response waves V and A had largest magnitude, followed by FFR peak F and then peak C in adults with NH. On the other hand the reduced amplitudes of the discrete peaks were observed in HI population. The analysis of V–A complex with respect (Table 4) to the mean values of latency, amplitude and slope (VA amplitude/VA duration) showed a significant difference (p value <0.05) between the groups in regards of the VA complex amplitude values and, consequently, of slope.

The delayed latencies and reduced amplitude in HI suggest changes in the nervous signaling transmission velocity (latency differences) and/or response generator synchronization (amplitude differences) suggesting processing abnormality at sub cortical level [38] may be due to hearing impairment. The previous studies also reported significantly delayed onset response latencies (waves V, A and C) in children with learning disorders [13, 19, 20] attributable to abnormalities in the acoustic representation of speech sounds at the brainstem.

The spABR Values in Normal Hearing Group

The goal of the study was to record the spABR to Hindi stimulus in normal hearing individuals. The obtained values for speech stimulus were compared with the parameters obtained by Russo et al. [30]. The similar values were found, without significant differences at p value 0.05 between the studies in Table 6. Although, the latency values of the present study are slightly delayed. However, this can be justified by existence of mild differences in acoustic parameter of employed stimulus, response collection and analysis.

Hence, it can be concluded that the rapid timing changes of consonant and vowel portion of the Hindi stimulus *ldal* are faithfully and accurately represented and preserved in normal auditory system. The differences and alterations in these measures might indicate neural encoding of speech of clinical conditions. The spABR results of the current study in the HI population has indicated alterations in the neural signaling velocity and synchronicity in the processing of speech acoustics at sub cortical level may be due to compromised hearing.

However, further studies are needed to compare the relationship of sub cortical and cortical responses to this speech sound in NH and HI. Further, the stimulus can be employed to assess its efficacy to establish its usefulness relationship at the brainstem level to categorize clinical population with language disorders, speech, learning and auditory processing and the consequences of auditory treatment and speech and hearing therapy and speech processing with rehabilitation technologies.

Conclusions

This is the first study to use Hindi speech stimulus to record the neurophysiologic encoding of speech at brainstem for clinical applications. The study describes the explicit method of construction and generation of speech stimulus

to record neurophysiology of sub cortical functioning objectively for various clinical applications. The speech sounds of 40 ms duration containing five formants were synthesized for measurement of spABR. The synthesized speech syllable *ldal* having sufficient acoustic information to be auditory identifiable, distinguishable and quantifiable acoustic parameters was considered as most ideal input signal to record the neurophysiologic responses.

The usefulness of stimulus *ldal* in recording sub cortical responses was validated on adults with normal hearing and hearing impairment. We obtained clear and significant differences between the response values for measures of spABR to temporal and spectral features of the stimulus in individuals with normal hearing and hearing impairment. The auditory brain stem response measures indicate precise encoding of acoustic features of the stimulus, evident in both the onset and FFR portions of the response in normal hearing groups suggesting a representative speech syllable in the normal auditory system.

However, the study has exhibited statistically significant longer latencies and shorter amplitude of discrete peaks in HI population as compared to NH group. The variations in response measures are most likely due to compromised hearing ability imply that alterations in brainstem responses have clinical significance to identify the subjects with possible auditory processing difficulties.

The study reports reliable timing and amplitude aspects of neurophysiologic responses to stimulus *ldal* in normal hearing adults and altered timing and amplitude aspects of sub cortical responses to stimulus in the hearing impaired population. Hence, it can be inferred that the altered potentials evoked in the auditory system is expected in clinical population having auditory processing difficulties. Thus, it can be concluded that spABR provide an objective means for examining auditory processes involved in encoding speech stimulus.

Implications of Constructed Speech Stimuli

The findings of this present study are very vital. However, data can be collected using the stimulus on large group of normal hearing population to establish the norms to provide a baseline for evaluation and comparison of sub-cortical responses evoked by speech stimulus between normal and clinical population. Apart from this:

These speech sounds can be used in assessment of perceptual abilities of voiced and voiceless sounds in infants and young children.

It can be used for monitoring benefit with speech correction strategies of voiced and unvoiced stop syllables.

It can also be used as input stimulus in analysis of electro acoustic characteristics of digital hearing aid and validating the hearing aid fitting.

Acknowledgments The authors submit their sincere thanks to the participants and parents for their patience and valuable time.

Compliance with Ethical Standards

Conflict of interest Authors have no conflict of interest. However, the corresponding author is pursuing Ph. D on the same topic under Maharashtra University of Health Sciences, Nasik – 422004, Maharashtra, India.

Ethical Approval “All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.”

Informed Consent Informed consent was obtained from the individual subjects involved in the study.

References

1. Banai K, Hornickel J, Skoe E, Nicol T, Zecker SG, Kraus N (2009) Reading and sub cortical auditory function. *Cereb Cortex* 19:2699–2707
2. Bradlow AR, Kraus N, Nicol TG, McGee TJ, Cunningham J, Zecker SG, Carrell TD (1999) Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *J Acoust Soc Am* 106:2086–2096
3. Chandrasekaran B, Kraus N (2009) The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47:236–246
4. Cunningham J, Nicol T, Zecker SG, Bradlow AR, Kraus N (2001) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol* 112:758–767
5. Eimas PD, Miller JL (1980) Contextual effects in infant speech perception. *Science* 209:1140–1141
6. Galbraith GC, Arbagey PW, Branski R, Comerchi N, Rector PM (1995) Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6(17):2363–2367
7. Gelfand SA (2009) *Essential of audiology*, 3rd edn. Thieme Publisher, New York
8. Gorga MP, Kaminski JR, Beauchaine KL et al (1989) Auditory brainstem responses from children three months to three years of age: II. Normal patterns of response. *J Speech Hear Res* 32:281–288
9. Hall JW (2007) *New handbook of auditory evoked responses*. Pearson, Boston, MA
10. Hood LJ (1998) *Clinical applications of the auditory brainstem response*. Singular Publishing Group Inc, San Diego
11. Hornickel J, Anderson S, Skoe E, Yi H, Kraus N (2012) Sub cortical representation of speech fine structure relates to reading ability. *Neuroreport* 23(1):6–9
12. Houtgast T, Steeneken HJM (1973) The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica* 28:66–73
13. Jacobson T (1985) *The auditory brainstem responses*. College Hill Press, San Diego

14. Jewett DL, Williston JS (1971) Auditory-evoked far fields averaged from the scalp of humans. *Brain* 94:681–696
15. Johnson KL, Nicol T, Kraus N (2005) Brainstem response to speech: a biological marker of auditory processing. *Ear Hear* 26(5):424–434
16. Johnson KL, Nicol T, Zecker SG, Kraus N (2007) Auditory brainstem correlates of perceptual timing deficits. *J Cogn Neurosci* 19(3):376–385
17. Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E et al (2008) Brainstem encoding of voiced consonant: vowel stop syllables. *Clin Neurophysiol* 119(11):2623–2635
18. Johnson KL, Nicol T, Zecker SG, Kraus N (2008) Developmental plasticity in the human auditory brainstem. *J Neurosci* 28:4000–4007
19. King C, Warrier CM, Hayes EA, Kraus N (2002) Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Letter* 319:111–115
20. Kraus N, Nicol T (2003) Aggregate neural response to speech sounds in the central auditory system. *Speech Commun* 41:35–47
21. Kraus N, McGee TJ, Carrell TD, King C (1996) Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 273:971–973
22. Krishnan A (2002) Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 166:192–201
23. Krishnan A, Swaminathan J, Gandour JT (2009) Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci* 21(6):1092–1105
24. Liberman AM, Delattre PC, Cooper FS, Gerstman LJ (1954) The role of consonant–vowel transition in the perception of the stop and nasal consonant. *Psychol Monogr* 68(8):1–13
25. Liberman AM, Cooper FS, Shankweiler DS, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461
26. Lisker L, Abramson AS (1964) A cross language study of voicing in initial stops: acoustical measurements. *Word* 20:384–422
27. Maddieson I (1984) *Patterns of sound*. Cambridge University Press, Cambridge
28. Musiek FE (1991) Auditory evoked responses in site-of lesion assessment. In: Rintelmann WF (ed) *Hearing assessment*. PRO-ED, Austin, pp 383–427
29. Palmer A, Shamma S (2004) Physiological representations of speech. In: Greenberg S, Ainsworth WA, Popper AN, Fay RR (eds) *Speech processing in the auditory system: Springer handbook of auditory research*, vol 18. Springer, New York (pp- xiv, 476)
30. Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. *Clin Neurophysiol* 115:2021–2030
31. Sininger YS (1993) Auditory brain stem response for objective measures of hearing. *Ear Hear* 14:23–30
32. Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31(3):302–324
33. Song JH, Banai K, Russo N, Kraus N (2006) On the relationship between speech- and non- speech-evoked auditory brainstem responses. *Audiol Neurotol* 4:233–241
34. Starr A, Picton TW, Sininger Y, Hood LJ, Berlin CI (1996) Auditory neuropathy. *Brain* 119:741–753
35. Stevens KN (1980) Acoustic correlates of some phonetic categories. *J Acoust Soc Am* 68:836–842
36. Wang S, Mannell R, Newall P, Zhang H, Han D (2007) Development and evaluation of mandarin disyllabic materials for speech audiometry in China. *Int J Audiol* 46(12):719–731
37. Werker JF, Tees RC (1984) Developmental changes across childhood in the perception of non-native speech sounds. *Can J Psychol* 37:278–286
38. Wible B, Nicol T, Kraus N (2004) Atypical brainstem representation of onset and formant structure of speech sound in children with language based learning problems. *Biol Psychol* 67:299–317
39. Wible B, Nicol T, Kraus N (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain* 128(2):417–423
40. Wong PC, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10(4):420–422