PERCEPTUAL ADAPTATION TO NATURAL FAST AND TIME COMPRESSED SPEECH IN YOUNG INDIAN ADULTS

[1]George Nooromplakal, [2]Ganesh A.C. & [3]Subba Rao T. A.

Abstract

It is well known from the previous research that the listener's can perceptually adapt to time compressed speech in isolated conditions. The ability of the listener to perceptually adapt in background noise has not yet thoroughly investigated in Indian languages. The present study was aimed to investigate the perceptual adaptation to natural fast sentences before and after time compressed speech in the presence of noise in Kannada language. A total of 24 normal hearing adults participated in the present study within the age range from 20 to 30 years. All participants were divided into two groups with 12 individuals in each group.

The sequence for Group I was normal rate, natural fast sentences, time compressed sentences and the sequence for Group II was normal rate, time compressed sentences, natural fast sentences. In each condition 25 sentences were presented to both groups. The adaptation effect was calculated using reaction time and word accuracy measurements. Result of the present study shows adaptation of listeners to natural fast speech. Accuracy and reaction time responses clearly indicated that the participants in II group had the advantage of time compressed speech presented before natural fast speech due to availability of spectral information. This also confirms the view that auditory system adapts to the stimulus presented in a particular context over a period of time. Clinical implication is discussed in terms of training Cochlear Implant children and also adults with stroke in whom speech perception is a challenging task.

Key words*: Speech perception, Accuracy, Reaction time, adaptation effect, Cochlear Implant*

Speech perception refers to a process of extracting meaningful words from a constant, fast changing acoustic signal of speech. It is the process by which the sounds of language are heard, interpreted and understood (Alho, 2010). As adult listeners, perception of message from the speaker is quite effortless. However, the simplicity of everyday conversation masks the perceptual and cognitive complexities concerned in perceiving speech. Upon examination of the speech stimulus, it is obvious that this everyday skill demands a perceptual ability of the listener (Holt, 2008). The perception of acoustic/phonetic characteristics of speech varies as a function of speaker, rate of speech, prosody, and so forth. Human speech perception is flexible and thus rapid adaptation takes place whenever there is a change in the speaker, speech rate or speech conditions (Sebastian-Galles, Dupoux, Costa & Mehler, 2000).

Perceptual adaptation is where our brain adapts to the perception that it receives and is the means by which the brain accounts for the difference that the subject may witness. The actual estimation of varying speech rates of speakers in the daily listening conditions accounts for better speech perception skills and improved communication act. There are quite number studies supports that there is adaptation effect to time compressed speech, noise-vocoded speech, foreign-accented speech and synthetic speech (Bradlow & Bent, 2007). Time-compressed speech is a technique used, often in perceptual tasks, to make recorded speech enclose more words in a given time, yet still be comprehensible. Within a speech act, speakers repeatedly vary their speech rate (Miller, Grosjean & Lomanto, 1984). These variations result in variations of co articulation and assimilation (Browman & Goldstein, 1990; Byrd & Tan, 1996), deletion of segments (Ernestus, 2002; Koreman, 2006), reduced vowel durations (Lehiste, 1970; Max & Caruso, 1997) and reduction of unstressed vowels (Lisker, 1973). The consequences of these variations force the listeners to use a normalization process which involves short term automatic compensations (Green, Cook, & Crawford, 1994; Miller et al., 1984; Miller & Liberman, 1979). For effective speech processing, listeners must be able to adjust to extreme changes in speech rate rapidly. When variations in speech rate are minimal, as in natural speech, listeners can accomplish this task without apparent effort. However, for extremely fast rates, adaptation becomes considerably more

[1]Clinical Specialist, Advanced Bionics India Pvt Ltd, Email: georgethomas_n@yahoo.co.in, [2]Lecturer, Dr M.V. Shetty College of Speech and Hearing, (MVSCSH), Mangalore-15, Email: gani0017@gmail.com, & [3]Professor and Principal, Dr M.V. Shetty College of Speech and Hearing, Mangalore-15, Email: drtasr.slp@gmail.com

difficult. In some of the previous studies the adaptation effect to fast speech was calculated by manipulating the temporal aspects of the signal while preserving the spectral components (Golomb, Peelle & Wingfield, 2007; Pallier, Dupoux, & Galles, 1998)

Studies have shown that listeners can adapt up to 38% compressed speech (Dupoux & Green, 1997). Janse (2004) proved that the processing for natural fast sentences is more complex than the artificially time compressed speech. It may due to natural fast speech lack both spectral and temporal domains (Koreman, 2006; Wouters & Macon, 2002).

Natural fast sentences also possess coarticulatory features which make them complicated to perceive. Peelle and Wingfield (2005) demonstrated that perceptual learning is comparable in young and older adults but maintenance and transfer of this learning decline with age. Time-compressed speech has a temporal and a segmental processing advantage over naturally produced fast speech (Janse, 2004). Studies suggest that listeners adapted to one type of fast speech facilitate adaptation and/or general performance for the other type (Adank & Janse, 2009).

Hence perception of natural fast speech is a complex task which the listeners achieve with compensatory processes. Perception to natural fast speech can also be achieved by training the listener with artificially time compressed speech which possesses fluctuations in temporal aspects and allows auditory system to get adapted to the temporal variations.

Even though there are many studies related to perceptual adaptation and time compressed speech, still there is a scarcity of Indian studies. Salian, Dhamani and Rajashekhar (2008) studied the perception of time compressed speech which had temporal envelope and fine structure cues in adults and children using cochlear implants. Not many studies are available in Indian languages on the perceptual adaptation to natural fast sentences. The present study is aimed at carrying out as an extension of the study of Adank and Janse (2009) to investigate perceptual adaptation to natural fast sentences before and after time compressed speech in the presence of noise in Kannada language.

The study aims at exploring the performance of young adults in perception and perceptual adaptation of sentence stimuli in natural fast and time compressed speech mode in the presence of background noise.

Method

Participants

Twenty four native speakers of Kannada (8 males and 16 females) with an age range of 20 to 30years (Mean age 23.6years) participated in the study. The participants had no speech, language or hearing impairment or neurological /psychological disease. The subjects reported to have no prior experience with the time compressed speech. Prior to the testing, Hearing evaluation of all subjects were done using GSI-61 Clinical audiometer and the thresholds were confirmed to be within normal limits.

Stimulus

The stimuli included 75 Kannada Quick Sin sentences (Avinash, Meti & Kumar, 2010) recorded by a twenty four year old male speaker who was exposure to native Kannada language for almost 20 years. The selected sentences were recorded in two different modes; Normal mode and Natural fast mode, in a sound treated room using a Frontech external microphone (SSD-HP-202) into a Sony Vaio (EB24) laptop computer. The PRAAT software, 5.1.22 version (Boersma & Weenink, 2008), was used to record the sound at sampling rate of 44100 Hz. In the Normal mode, the speaker was instructed to read the sentences in quiet using a Normal speech rate. Next, the natural fast sentences were recorded by instructing the speaker to read the sentences four times in quick succession. Out of these, the second (or third in few occasions) sentence was selected as the fastest but still most fluent of the four sentences produced. Zero crossing was performed for all the sentences to get the absolute onset and offset timing of sentence. On an average, sentences in Normal speed consisted of 4.81syllables per second. However, natural fast sentences were of 11.37 syllables per second. Thereby, natural fast sentences were produced at 42.56% of the duration of normal speed sentences.

As a next step, Pitch Synchronous Overlap and Add (PSOLA) in PRAAT, was used to obtain the Time compressed sentences. Each sentence was time compressed using a Compression ratio, obtained by comparing the Normal and Natural fast sentences. For instance, if the ratio between a selected normal and natural fast sentence is 40%, then the time compression will be done at 40% for that particular sentence. Also, the intensity levels of all sound files were scaled to 70 dB Sound Pressure Level. Lastly, four talker babble at 0 dB SNR was added to all the recorded sentences using a Matlab in order to produce a more naturalistic auditory condition.

Procedure

Each participant (tested individually) was seated comfortably in the room and given appropriate instructions. A total of 5 sentences were given as practice trials in the beginning. These trials were not included in the test. To measure the adaptation, a total of 75 sentences (25 in each mode, Normal, natural fast and time compressed) were presented in two experimental designs,
1) Normal speech - Natural fast - Time compressed mode and
2) Normal speech - Time compressed - Natural fast mode

Stimulus presentation and reaction time measurements were done using DMDX software (version 4.0.3.0), with Frontech external microphone (SSD-HP-203) into a Sony Vaio (EB24) laptop computer. Each participant's reaction time and word frequency accuracy was calculated.

The significance differences between three experimental conditions was assessed by a repeated measure ANOVA test using SPSS software (version 17)

Results

Accuracy

Twenty four subjects participated in the study. The study aimed at measuring the response accuracy and mean reaction time for subject's response. The accuracy of responses was measured as the number of words repeated correctly, per sentence.

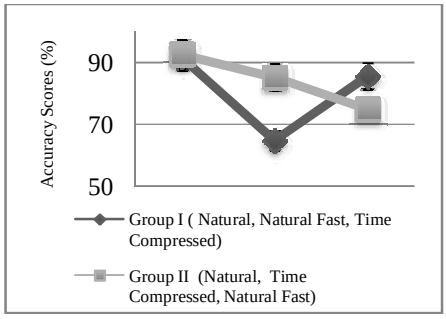

Figure 1*: Shows the mean accuracy scores of subjects for sentences in normal speed, artificially time compressed mode and normal speech rate across Group I and Group II*

The mean percentage correct scores for speech perception is showed in Figure 1 for Normal rate Speech (Mean scores- group 1- 85.45 & Group 2- 85.76), Time Compressed Speech (Mean score- Group1- 91.81 & Group 2- 92.57) and Natural fast Speech (Mean scores- Group 1-64.67 & Group 2- 74.78).

Thereby the overall scores indicated Normal speech mode as being better than other two modes in both the experimental conditions. Moreover, the accuracy scores of sentences in natural fast mode were poorer than time compressed mode when compared to the normal mode sentences. In addition, adaptation effect was observed as the test preceded the twenty-fifth sentence, especially in natural fast mode. It was observed that the responses improved in accuracy with increase in time. Although, in case of the time compressed speech, adaptation effect was less prominent as the exposure time increased.

The ANOVA results indicated a significant main effect of condition for both first [$F (1.74, 19.14) = 295.21$, $p<0.001$] and second experimental conditions [$F (1.76, 19.36) = 65.58$, $p<0.001$]. For each condition, the results were compared with the other two conditions using Pair wise comparisons with Boneferroni adjustments. The data showed statistical significance ($p <0.05$) by the post hoc test when the Normal speech mode was compared with Natural fast mode and Time compressed mode conditions, in both experimental designs.

Reaction Time Measurement

The mean reaction time scores is showed in Figure 2 for Normal rate Speech (Mean scores- group 1- 333.86 ms & Group 2- 351.50ms), Time Compressed Speech (Mean score- Group1- 714.07 & Group 2- 735.32) and Natural fast Speech (Mean scores- Group 1-953.17 & Group 2- 784.97).

Reaction time measurements were longer for both natural fast and time compressed speech when compared with normal sentences. In both natural fast and time compressed speech, listeners took more time to react correctly to the presented stimuli. Reaction time became further poor as the sentences were pooled with speech babble. In some, the reaction time analysis shows that listeners responded fastest to the normal-speed sentences, slower for the time-compressed sentences and slowest for the natural fast sentences. Overall the reaction time for group II was less for natural fast sentences. However the reaction times were not significantly different for the order of stimuli presentation. Whereas adaptation to time compressed speech did not show improved accuracy over trials, but showed decreased reaction times over trials. Regulation to natural fast speech was found both in

improved accuracy and in considerable decreased reaction times over trials

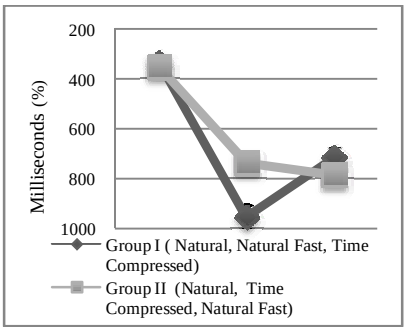

Figure 2: Average reaction times in millisecond for Normal sentences, Time compressed and Natural fast across Group 1 and Group 2.

Discussion

The study aimed at observing individual adaptation to natural fast sentences presented in degraded conditions and thus comparison of this adaptation to time compressed sentences. Two groups of participants were involved in the study and were presented with sentences in three specified modes but in different order for each experimental condition.

Results of the present study shows adaptation of listeners to natural fast speech. Natural fast speech involves greater variation in spectral as well as temporal aspects compared to artificially time compressed speech (Adank & Janse, 2009; Bent, Buchwald, & Pisoni, 2009). Thus, natural fast speech is always difficult in perception relative to both time compressed and normal speech sentences. This is possibly due to its spectral and temporal variations. Also natural fast speech possesses greater amount of co articulatory postures which makes it difficult to process in the natural listening environment. The scores of accuracy obtained for natural fast speech were considerably low when compared to other two modalities. Addition of speech babble at 0dB SNR made the task quite difficult and further responses measured also shows an increased reaction time. Performance of speech in artificially time compressed mode was better than the natural fast mode, as artificially time compressed speech has variations only in temporal aspects whereas its spectral contents were intact. But adaptation process in artificially time compressed speech was less seen than that of natural fast speech and the accuracy scores in artificially time compressed speech were almost similar to that of normal speech mode. The reason for this may be explained in terms of better spectral residue in the speech content. The better performance was also reflected in reaction time measurements where it was better and reduced than that of natural fast mode.

The present study also attempted to compare the perceptual adaptation of natural fast speech in relation with artificially time compressed speech. The second group (Normal- time compressed- natural fast) showed higher accuracy for the natural fast materials. The main reasons for such results is that listeners from this group already being adapted to temporally altered speech material (time compressed speech) than natural fast sentences which has alterations in both spectral as well as temporal aspects. This phenomenon confirms the view that auditory system adapts to the stimulus presented in a particular context over a period of time.

The listeners in the first group (normal- natural fast- time compressed) didn't show any considerable adaptation from natural fast speech to time compressed speech, as the accuracy rates were almost similar for time compressed speech materials in both groups. But both groups showed reduction in reaction time measurements over sentences. In the study, the time-compression ratio varied according to the natural fast sentences. It is unclear how this may have affected the adaptation in participants. Studies show that phonetic inconsistency in the experiment will assist in perceptual learning (Logan, Lively & Pisoni, 1991). Compression ratio can also lead to decline in the performance (Dupoux & Green, 1997), where as there are contradictory studies on positive relation between performance and compression ratio (Golomb et al., 2007).

Study confirms previous research done by Janse (2004) and Adank and Janse (2009). The scores obtained for natural fast speech in the present study were lower when compared to study done by Adank and Janse (2009), which can be attributed to the introduction of speech babble at 0dB SNR. According to Pavlovskaya & Hochstein (2004), if an easy task is followed by a complicated task, there will be a transfer of learning. However same thing won't happen when an complicated task is followed by an easier task. This view was supported from the study as the listeners in the second group who were exposed to time compressed speech before natural fast speech performed well and not the first group which was presented with natural fast speech prior to time compressed speech. Another factor which might contribute for the poor performance in natural fast sentence was coarticulatory effect and assimilation, which results in deletion of segments, decreased vowel durations and reduction of unstressed vowels.

Clinical Implications

According to literature, geriatric population and persons with stroke generally have more trouble with natural fast speech. Thus introducing the time compressed speech in the rehabilitation approach can facilitate perceptual abilities to process naturally produced fast speech. In the perception of natural fast sentences in the environment are multifaceted to children with Cochlear Implantation. Hence training these children with time compressed speech in the clinical setting as well as in home environment will improve their speech perception abilities. Further, effect of time compression speech on adaptation have to be evaluated in individual with hearing loss and other pathologies which impairs speech perception abilities. Future research can also focus on comparison of speech perception abilities across adults, geriatrics and pediatric group, especially in children with cochlear implant and adults with stroke.

Conclusions

The present study focused to measure the perceptual adaptation of natural fast sentences in young adults and correlated this adaptation with respect to time compressed speech sentences. The results showed perceptual adaptation effect of young adults on natural fast speech as well as on time compressed speech. Exposure to time compressed speech prior to natural fast speech significantly increased the word accuracy and reduced response reaction time. The study provides us an insight about how listeners process a naturalistic distortion in the speech signal and in turn helps in learning more about human's general speech comprehension ability. The results of the present study show the importance of adaptation effect to degraded signal in the normal hearing individuals.

References

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *Journal of acoustic society of America, 126, 5,* 2649-2659. doi: org/10.1121/1.3216914

Alho, J. (2010). Role of the speech motor system in speech perception. Retrieved from http://lib.tkk.fi/Dipl/2010/urn100236.pdf.

Avinash, M. C., Meti.R.., & Kumar, U., (2010). Development of Sentences for Quick Speech-in-Noise (QuickSin) Test in Kannada. *Journal of Indian Speech and hearing Association, 24* (1), 59-65.

Bent, T. V., Buchwald, A., & Pisoni, B.D. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *Acoustic society of America,126, 2660-2669.*

Boersma, P., & Weenink, D. (2008). Praat: Doing Phonetics by computer (Version5.0.17) [Computer Program]. Retrieved from http://www.praat.org/

Browman, P., & Goldstein, L. (1990). Gestural specification using dynamically defined articulatory structures. *Journal of phonetics, 18,* 299-320.

Byrd, D., & Tan. (1996). A Phase Window Framework for Articulatory Timing. *Phonology, 13,* 139-169. doi: 10.1017/S0952675700002086

Bradlow, A. R., & Bent, T. (2007). Perceptual adaptation to non-native speech. *Cognition 106,* 707-729, doi:10.1016/j.cognition.2007.04.005

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24,* 263 – 282.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effect of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance. 23,* 914-923.

Ernestus, M. (2002). The recognition of reduced word forms. *Brain and Language, 81, 3,* 162-173, doi:10.1006/BRLN.2001.2514

Golomb, D. J., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *Acoustic Society of America, 121, 1701- 1708.* doi: org/10.1121/1.2436635.

Green, P., Cook, M., & Crawford, M. (1994), Handling missing data in speech recognition. *International Conference on Spoken Language Processing (ICSLP-94), Yokahama, Japan*

Holt, L. L. (2008). Speech Perception within an auditory cognitive Science framework. *Carnegie Mellon University Research Showcase, 1,* 1-7. doi: 10.1111/j.1467-8721.2008.00545.x

Janse, E. (2004). Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication, 42, 2,* 155-173, doi:10.1016/J.SPECOM. 2003.07.001

Koreman, J. (2006). Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of Acoustic society of America 119, 1,* 582-596, doi: ORG/10.1121/1.2133436

Lehiste, I. (1970). Suprasegmentals. *The MIT press, 55 Hayward Street Cambridge, MA 02142 USA.*

Lisker, L. (1973). On explaining vowel duration variation. *Haskins laboratories, New haven, con, SR 37/38.*

Logan, J. S., Lively, S. E., & Pisoni, D. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report *Journal of acoustic society of America. 89,* 874–886.

Max, L., & Caruso, A. J. (1997). Acoustic Measures of Temporal Intervals Across Speaking Rates; Variability of Syllable- and Phrase-Level Relative Timing. Journal of Speech, Language, and Hearing Research,.40, 1097-1100.

Miller, J. L., & Liberman A. M. (1979). Some effects of later-occurring Information on the perception of stop consonant and semivowel. *Attention, Perception & Psychophysics. 25*, 457-465. doi: 10.3758/BF03213823.

Miller, L. J., Grosjean, F., & Lomanto, C. (1984). Articulation Rate and Its Variability in Spontaneous Speech: A Reanalysis and Some Implications. *Phonetica, 41*, 215-225. doi: 10.1159/000261728.

Pallier, C., Dupoux, E., & Galles, N. S. (1998). Perceptual adjustment to time-compressed speech: a cross-linguistic study. *Memory and cognition 26*, 844-851, doi: 10.3758/BF03211403.

Pavlovskaya, M., & Hochstein, S. (2004). Transfer of perceptual learning effects to untrained stimulus dimensions. *Journal of Vision, 4, 416*. doi: 10.1167/4.8.416

Peelle, J., & Wingfield, A. (2005). Dissociations in Perceptual Learning Revealed by Adult Age Differences in Adaptation to Time-Compressed Speech. *Journal of Experimental Psychology: Human Perception and Performance, 31, 6*, 1315–1330, doi: 10.1037/0096-1523.31.6.1315.

Salian, S. K., Dhamani, I., & Rajashekhar, B. (2008). *Role of temporal envelope and fine structure cues on time compressed speech recognition in adults and children using cochlear implants.* Cochlear Implant Group of India Conference, India.

Sebastian-Galles, N., Dupoux E., Costa A., & Mehler J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception and Psychophysics, 62*, 834-842

Wouters, J., & Macon, M. (2002). Effects of prosodic factors on spectral dynamics Analysis. *Journal of Acoustic Society of America, 111*, 417-427, doi: ORG/10.1121/1.1428262