Intelligibility and Acoustic Characteristics of Clear and Conversational Speech in Telugu (A South Indian Dravidian Language)

Naresh Durisala · S. G. R. Prakash ·
Arivudai Nambi · Ridhima Batra

Received: 21 November 2009 / Accepted: 13 June 2010 / Published online: 11 April 2011
© Association of Otolaryngologists of India 2011

Abstract The overall goal of this study is to examine the intelligibility differences of clear and conversational speech and also to objectively analyze the acoustic properties contributing to these differences. Seventeen post-lingual stable sensory-neural hearing impaired listeners with an age range of 17–40 years were recruited for the study. Forty Telugu sentences spoken by a female Telugu speaker in both clear and conversational speech styles were used as stimuli for the subjects. Results revealed that mean scores of clear speech were higher (mean = 84.5) when compared to conversational speech (mean = 61.4) with an advantage of 23.1% points. Acoustic properties revealed greater fundamental frequency (f_0) and intensity, longer duration, higher consonant–vowel ratio (CVR) and greater temporal energy in clear speech.

Keywords Clear speech · Conversational speech · Sensory neural · Hearing impaired · Acoustic properties · Intelligibility

Introduction

In everyday conversational speech, people speak easily and freely without paying much attention to their enunciation. However, under various conditions, people tend to speak with more care in an attempt to make their speech “clearer”. This term is denoted as “clear speech”. One condition in which this might occur is when people cannot hear well, such as when talking in a noisy subway or when speaking to someone who has impaired hearing. Research has shown that clear speech is a speaking style that many talkers adopt in order to be understood more easily in difficult communication situations. In a seminal paper, [20] reported clear speech intelligibility advantage for hearing impaired adults. Furthermore, this intelligibility advantage is independent of listener, presentation level and frequency gain characteristics [21, 22]. The available data indicates that the clear speech effect (i.e., the intelligibility difference between the clear speech and conversational speech) is generally robust across a wide range of talkers, listeners and signal degradation conditions and suggests that signal processing schemes that convert conversational speech to a sufficiently close approximation of clear speech could improve speech intelligibility in many situations.

Acoustic comparisons of conversational and clear speech in English have shown that clear speech modifications typically involve enhancement of the overall acoustic salience of the speech signal by means of a decreased speaking rate, longer and more frequent pauses, an expanded pitch range, greater sound pressure levels, more salient stop releases, greater obstruent intensity, increased energy in the 1000–3000 Hz range of long-term spectra, and increased modulation depth of low frequency modulations of the intensity envelope [3, 12, 14, 18, 19, 26, 28].

N. Durisala (✉)
Beltone Hearing Care, House No. 2328, Sushant Lok Phase I,
Near Gold Soukh Jewellery Mall, Gurgaon 122001, Haryana,
India
e-mail: nareshd15@gmail.com

S. G. R. Prakash
Department of Audiology and Speech Language Pathology,
Osmania University, Hyderabad, India

A. Nambi
Department of Audiology, Manipal University, Manipal, India

R. Batra
Ram Manohar Lohia Hospital, New Delhi, India

In addition, the vowel space is expanded in clear speech when compared to conversational speech [6, 8, 11, 16, 19].

In the perceptual domain, intelligibility studies established that naturally produced clear speech enhances intelligibility for various listener populations including adults with normal or impaired hearing [7, 8, 17, 19, 26, 27] elderly adults [23, 9] both native and non-native listeners of the target language [2, 4, 25] and children with and without learning impairments [3].

In the temporal domain [19] found that clear speech is associated with reduced speaking rate, increased number and length of pauses. Although this study identified many acoustical differences between clear and conversational speech, however, it did not attempt to determine which differences were responsible for the high intelligibility of clear speech. In terms of speaking rate [1, 15] found that rate of speech for conversational speech approximately ranged from 160 to 200 words per minute (wpm) and for clear speech it was approximately 90 to 100 wpm. The conversational speaking rates were much higher than those usually observed in natural speech; however, the usual measurements of roughly 140 wpm included hesitations and breathe pauses. Recently, in an in-depth, cross-language exploration of temporal changes for segmental contrasts in Croatian and English clear speech (relative to conversational speech) [24], found similar asymmetrical patterns of lengthening for Croatian phonemically long and short vowels, English vowels preceding voiced and voiceless coda stops, and Croatian and English voice onset time (VOT) for voiced versus voiceless stop categories. However, the duration ratio between the “long” and “short” members of the contrastive categories in both English and Croatian was found to be remarkably stable across the two speaking styles suggesting that, unlike the spectral contrasts, duration contrasts are stable rather than enhanced under conditions of hyperarticulation for clear speech.

Clear speech is of course not a substitute for other well known communication habits like, reducing background noise, avoiding communicating from a different room or with the speaker’s head turned and to have light on the speaker’s face and not behind.

In this study, clear speech is defined to be the style of speech which results from people attempting, with the help of feedback on their intelligibility, to make their speech more intelligible. The term conversational speech will be used to denote speech which is spoken in a style similar to a speaking style used in a natural everyday situation.

As evident from the review of literature, there is a scarcity of studies done wherein a comparison has been made between clear and conversational speech for both perception and production tasks in Indian languages. Hence the current study was undertaken to understand the intelligibility difference between the two speaking styles and also to find the acoustic characteristics which separate the two.

Aim of the Study

The present study is aimed at examining the improvements in intelligibility scores when a speaker attempts to speak clearly and also to understand these improvements in terms of changes in the acoustic parameters of speech.

Method

The present study investigated intelligibility and acoustic characteristics of clear and conversational speech for which the methodology used is given below.

Intelligibility

Listeners

Twenty listeners with post-lingual bilateral moderately severe sensorineural hearing loss (SNHL) with an age range of 17 to 40 years were included in the study. All the listeners were native Telugu speakers and had no history of any significant neurological or cognitive listening deficits. Informed consent was obtained from the subjects.

Stimuli

Listeners were tested on 40 Telugu sentences. Each sentence contained a minimum of one content word and a maximum of three content words with 83 total content words. Sentences were chosen rather than isolated words because it would permit the talker to manipulate prosodic and phonological variables when speaking clearly and conversationally. Intelligibility was measured by scoring the total number of correct content words identified. A word was marked incorrect if a single sound in the word was omitted or misidentified.

Talker

Sentences were spoken by a female native Telugu speaker (aged 24 years). Although she had no experience of any professional training, a method was developed to produce clearest possible speech by providing her auditory feedback and informal training. After several trials, the clearest speech produced was used as stimuli for the subjects.

Procedure

Recordings were made in a sound treated room for both clear and conversational speech sentences by using digital acquisition software with a sampling rate of 44100 Hz at 32 bit rate. Microphone was placed at a distance of 15 cm

from the mouth. The talker typically recorded both styles of speaking in an alternative manner. While recording conversational speech, the talker was instructed to recite the stimulus material in the same manner in which she spoke in ordinary conversation. During the course of recording, the talker was notified if she abruptly altered her speaking rate or began to speak mechanically. While recording clear speech, the talker was instructed to speak as clearly as possible, as if she was trying to communicate in a noisy environment or communicating with a hearing impaired listener. The talker was also instructed to enunciate consonants more carefully and with greater vocal effort than in conversational speech and to avoid slurring words together. All the hearing impaired listeners were tested individually using MAICO double channel diagnostic clinical audiometer. Air conduction and bone conduction thresholds were obtained prior to the start of the testing. All the subjects who had air-borne gap within 10 dB and 'A type' tympanogram were included in the experiment. Prior to the start of the experiment each and every subject was presented with a puretone to determine their most comfortable level. All stimuli were presented bilaterally from a laptop (Compaq) which was routed to the supra aural headphones through the audiometer. All listeners were informed about the nature of the sentence materials and instructed to write down the responses whatever they thought they heard. Listeners were asked to guess the sentences if they were not understood. A score of one was given for every correct response of content word.

Acoustic Characteristics

Acoustic Analysis

To isolate the characteristics that contribute to the highly intelligible speech, acoustical parameters of the female Telugu speaker were analyzed at a global level by using PRAAT (version 4.4.12) and MATLAB version (6.5) software packages. This could help in identifying factors responsible for the intelligibility advantage of clear speech over conversational speech. The relative differences were computed and compared statistically by using paired *t*-test.

Data Analysis

Intelligibility

Each participant in perception test received a keyword correct score out of 83 content words for the 40 sentences they heard. A score of one was given for every correct response of the content word and marked as wrong even if they were transcribed incorrectly. Percent correct words

were calculated and paired *t*-tests were employed for statistical analysis.

Acoustic Characteristics

Fundamental Frequency (F0)

Using the PRAAT software package, fundamental frequency (F0) values were extracted from the voiced portions of 20 content words which were selected randomly. Statistical comparisons of F0 average were performed via paired *t*-tests. The difference variable was $D = F0_{\text{clear}} - F0_{\text{conv}}$, where $F0_{\text{clear}}$ was the fundamental frequency of the content word spoken clearly, and $F0_{\text{conv}}$ was the fundamental frequency of the content word spoken conversationally.

Intensity

To compare the relative power of words spoken in clear and conversational mode, intensity for each word in both modes was computed. Statistical comparisons of clear and conversational level for each word were performed via paired *t*-tests, where the difference variable was $D = I_{\text{clear}} - I_{\text{conv}}$, where I_{clear} was the intensity of the content word spoken clearly, and I_{conv} was the intensity of the content spoken conversationally.

Duration

The relative durations of 20 content words spoken clearly and conversationally were also computed and compared statistically via paired *t*-test. In this case, the difference variable was $D = D_{\text{clear}} - D_{\text{conv}}$, where D_{clear} was the duration of the content word spoken clearly, and D_{conv} was the duration of the content spoken conversationally.

Consonant–Vowel Ratio (CVR)

Must attention in the literature [9, 17] has been given to consonant–vowel ratio (CV ratio) or ratio of consonant power to the power in the nearest adjacent vowel. In the present study consonants were grouped into five classes: stops, fricatives, affricates, liquids, and, nasals. The intensity of these consonants and the relative intensity of the adjacent vowel were analyzed by using PRAAT software. Relative CV ratio was computed and compared statistically via paired *t*-test.

Temporal Modulation Spectrogram

Studies have shown that temporal envelope modulations are thought to be an important factor for enhancing speech intelligibility [12]. Key words were excised from both the

clear and conversational speech sentences and were down sampled to 20 kHz sampling rate. The temporal envelope modulations and modulation spectrogram for these words were analyzed by using MATLAB software.

Results

Intelligibility

The average speech perception scores (percentage correct words) are represented in Table 1 and graphical representation is shown in graph 1. It is clearly evident that listeners obtained higher scores for sentences spoken in clear speech mode when compared to conversational speech. Paired *t*-tests were employed and the results revealed a statistically significant difference ($P < 0.001$) between both speaking modes with an intelligibility advantage of 23.1% percentage points. The possible reasons for this increased intelligibility scores in clear speech could be greater temporal modulation depth, higher spectral energy and greater duration when compared to conversational speech. This increase in duration can contribute to better temporal processing for individuals with perceptual deficits. The increase in perception scores for clear speech supports the results of the previously reported studies [20].

Acoustic Analysis

Fundamental Frequency

Table 2 and graph 2 depict the fundamental frequency for clear speech and conversational speech. A significant mean

Table 1 Mean, standard deviation, significance of mean difference in speech intelligibility scores for clear and conversational speech

Speech Intelligibility scores (SIS)	Mean	Standard deviation	t-value
SIS <i>clear</i>	84.5	5.5	**11.9
SIS <i>conv</i>	61.4	8.4	

*** significant at $p < 0.001$

Graph 1: Mean Speech Intelligibility scores of Clear and Conversational speech

Table 2 Mean, standard deviation and significance of mean difference for fundamental frequency (F0) between clear and conversational speech

Fundamental frequency (F0)	Mean	Standard Deviation	t-value
F0 <i>clear</i>	244.8	22.3	4.00*
F0 <i>conv</i>	223.9	8.8	

*** significant at $p < 0.05$

Graph 2: Mean Fundamental Frequency of Clear and Conversational Speech

difference of 20.9 Hz was found between the two speaking modes. Paired *t*-test also revealed statistically significant difference at $P < 0.05$ between the two speaking modes. This increase in fundamental frequency suggests that a larger amount of laryngeal tension occurs while speaking clearly [19].

Intensity

The mean intensity values spoken by a female talker in both speaking modes is given in Table 3 and the graphical representation is shown in graph 3. These results indicate that the relative power of clear speech is higher (5.5 dB).

Table 3 Mean standard deviation, significance of mean difference for intensity between clear and conversational speech

Intensity	Mean	Standard deviation	t-value
I <i>clear</i>	73.0	2.1	**12.2
I <i>conv</i>	67.5	2.5	

*** significant at $p < 0.001$

Graph 3: Mean Intensity of Clear and Conversational Speech

The statistical analysis also revealed a significant mean differences at $P < 0.001$ between the two speaking modes. The increase in intensity suggests that during clear speech, talker automatically tends to speak louder when compared to conversational speech [10]. Also stated that the clear speech is produced at higher intensities than conversational speech.

Duration

The mean duration values for 20 content words spoken in clear and conversational speech is revealed in Table 4. As it was seen in the findings for fundamental frequency and intensity, the average duration of clear speech words was also greater by 0.25 s when compared to average duration of the conversational speech. A statistically significant difference at $P < 0.001$ between the two speaking modes was eminent. Previous studies [3, 13] who investigated average sentence duration between both speaking styles, also found large differences. The possible reason for those large differences in duration between the two speaking styles in their studies could be attributed to the additional pause duration between the words within the sentences. These silent pauses between the words were longer in clear speech which contributes as an additive factor for increase in the overall duration of the sentences.

Consonant–Vowel Ratio (CVR)

The mean CVR values spoken in both speaking styles are tabulated in Table 5. Like the above acoustic parameters, the CV ratio for clear speech was found to be higher when compared to conversational speech. Paired t -test also revealed that there was statistically significant difference between both speaking modes at $P < 0.05$. In clear speech,

Table 4 Mean, standard deviation, significance of mean difference in duration between clear and conversational speech

Duration	Mean	Standard deviation	t -value
D <i>clear</i>	0.68	0.18	9.6**
D <i>conv</i>	0.43	0.16	

** Significant at $P < 0.001$

Table 5 Mean, standard deviation and significance of mean difference for Consonant–Vowel ratio between clear and conversational speech

Consonant–Vowel ratio	Mean	Standard deviation	T
Cv <i>clear</i>	0.98	0.43	2.98*
Cv <i>conv</i>	0.95	0.59	

* Significant at $P < 0.05$

the difference in intensity between consonants and adjacent vowels was lesser and also the intensity of consonants was higher when compared to conversational speech. This may contribute to increased speech intelligibility in clear speech. Unlike in this study, [13] analyzed CVR for different classes of consonants and found little differences for any of the consonants between speaking styles, but affricates achieved 2–3 dB larger CVR in clear speech compared to conversational speech.

Temporal Amplitude Modulation Spectrogram

Amplitude modulation spectrogram was analyzed by using MATLAB (version 6.5) software which is depicted in Fig. 1. The two modulation spectrograms in the figure depicts that clear speech energy is more at higher frequencies when compared to conversational speech whereas for lower frequencies it is concentrated almost equally. This energy is extensively higher in the bottom figure particularly at frequencies between 8 and 9 kHz. This greater energy at higher frequencies in clear speech may contribute to increased intelligibility.

Discussion

The present study aimed at investigating the intelligibility differences in clear and conversational speech among stable sensorineural hearing loss individuals and also to objectively analyze the acoustic factors contributing to those differences.

The results of intelligibility test revealed consistent clear speech intelligibility advantage across all the sentences. The data provides strong evidence, that the talker successfully maintained clear speech articulatory modifications across all the sentences which led to an increased intelligibility for all the hearing impaired listeners. Likewise the previous research studies conducted on clear speech [3, 4, 7, 8, 17, 19, 27] the current study also focused on isolated words. The connection between the temporal organizational change from conversational to clear speech and its relation with increased intelligibility is yet to be fully established [13]. For instance, found that clear speech produced at normal/conversational speaking rates increased the intelligibility, suggesting that, the temporal changes are not essential for enhancing the intelligibility. However, this does not mean that, when present, the temporal changes of slow clear speech are not important for the intelligibility benefit. Listeners may use the temporal cues when available either to further enhance intelligibility or to process the speech signal in a different manner (weighing available cues differently)[5]. The present data suggest few possible temporal-related mechanisms that may underlie the

Fig. 1 Amplitude modulation spectrogram for word **maɪsaɪlɒu** (top left clear, top right conv) and for word **tʃettu** (bottom left clear and right conv)

intelligibility increase. First, the overall speaking rate in conversational speech was found to be faster when compared to clear speech. This decrease in speaking rate may be related to greater articulatory precision in clear speech contributing to better speech intelligibility and was achieved through lengthening of consonants and vowels in clear speech. An informal observation found that unlike in clear speech, the number of pauses and silent gaps were found to be dropped or co-articulated in conversational speech. Second, the temporal restructuring across all the sentences may further contribute to higher clear speech intelligibility compared to conversational speech intelligibility. In the present paper this was notified by comparing the duration of both styles of speaking. This suggests that in order for listeners to successfully interpret speech signal and derive information, the timing related cues also plays a predominant role which can be measured as durational ratios.

Conclusions

The present study was carried out to observe the enhancement of the intelligibility scores when a speaker tries to speak clearly and also to know these enhancements in terms of changes in the acoustic parameters of speech.

For the intelligibility study, the mean score of clear speech was 84.5 with standard deviation of 5.5 and mean score for conversational speech was 61.4 with standard deviation of 8.4, respectively. The advantage of clear speech was found to be 23.1 percentage points. The results of the present study are in congruence with the previous investigations.

The present study has also revealed several acoustic differences between clear and conversational speech. The findings are as follows:

- The value of fundamental frequency (F0) tends to be higher in clear speech when compared to conversational speech with a significant mean difference of 20.8 Hz.
- The intensity of clear speech was elevated when compared to conversational speech with a significant mean difference of 5.5 dB.
- The duration of clear speech was greater in contrast with conversational speech with a significant difference of 0.24 s.
- The CVR in clear speech was higher with a mean difference of 0.3 dB when compared to conversational speech.
- Clear speech has more temporal energy at higher frequencies when compared to conversational speech.

In conclusion, clear speech can be used as an effective measure to improve speech perception for the hard of hearing individuals. Understanding the acoustic characteristics of clear speech can contribute not only to basic science, but also, to the development of improved hearing aids, speech training aids, speech synthesizers and speech recognition systems. The present speech perception results, together with acoustic results suggest innovative signal processing schemes in hearing aids to improve speech performance in individuals with hearing loss. Future research can be carried out by using individuals with various degrees of SN hearing loss particularly on children and older age population. Based upon that, it may be worthwhile for teaching parents, clinicians and other caregivers who come into frequent contact with children with hearing impairment and other speech perception difficulties to monitor their own speech in an effort to adopt a clear speaking style on a routine basis.

Acknowledgments We would like to sincerely thank Dr. R. Rangasayee, Director of Ali Yavar Jung National Institute for the Hearing Handicapped, Secunderabad, for providing an opportunity to carry out this research work. We would also like to acknowledge our participants without whom the study would not have been possible. This study was conducted at Ali Yavar National Jung National Institute for the Hearing Handicapped, Southern Regional Centre, Secunderabad, India.

References

1. Abrams MH, Goffard SJ, Kryter KD, Miller GA, Miller J, Sanford FH (1944) Speech in noise: a study of the factors determining its intelligibility (ORSD Report No. 4023). Harvard University, Psychoacoustic Laboratory, Cambridge
2. Bradlow AR, Alexander J (2007) Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust Soc Am* 121:2339–2349
3. Bradlow AR, Kraus N, Hayes E (2003) Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res* 46:80–97
4. Bradlow AR, Bent T (2002) The Clear speech effect for non-native listeners. *J Acoust Soc Am* 112:272–284
5. Drullman R, Festen JM, Plomp R (1994) Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am* 95:2670–2680
6. Ferguson SR, Kewley-Port D (2007) Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J Acoust Soc Am* 50:1241–1255
7. Ferguson SH (2004) Talker differences in clear and conversational speech: vowel intelligibility for normal hearing listeners. *J Acoust Soc Am* 116:2365–2373
8. Ferguson SR, Kewley-Port D (2002) Vowel intelligibility in clear and conversational speech in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 112:259–271
9. Gordon-Salant S (1987) Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects. *J Acoust Soc Am* 97:3099–3111
10. Huber JE, Stathopoulos ET, Curione GM, Ash TA, Johnson K (1999) Formants of children, women and men: the effects of vocal intensity variation. *J Acoust Soc Am* 106:1532–1542
11. Hefler KS (1998) Auditory and auditory-visual of clear and conversational speech by older adults. *J Am Acad Audiol* 9:234–242
12. Johnson K, Flemming E, Wright R (1993) The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69:505–528
13. Krause JC, Braida LD (2004) Acoustic properties of naturally produced clear speech at normal speaking rates. *J Acoust Soc Am* 115:362–378
14. Krause JC, Braida LD (2002) Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility. *J Acoust Soc Am* 112:2165–2172
15. Liu S, Rio ED, Bradlow AR, Zeng FG (2004) Clear speech perception in acoustic and electric hearing. *J Acoust Soc Am* 116:2374–2383
16. Mac Clay H, Osgood CE (1967) Hesitation phenomena in spontaneous English speech. In: Jakobovits LA, Miron MS (eds) *Readings in psychology of language*. Prentice-Hall, Englewood Cliffs, pp 305–324
17. Montgomery AA, Edge RA (1988) Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults. *J Speech Hear Res* 31:386–393
18. Payton KL, Uchanski RM, Braida LD (1994) Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J Acoust Soc Am* 95:1581–1592
19. Perkell JS, Zandipour M, Matthies ML, Lane H (2002) Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *J Acoust Soc Am* 112:1627–1641
20. Picheny MA, Durlach NI, Braida LD (1989) Speaking clearly for the hard of hearing III: an attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *J Speech Hear Res* 32:600–603
21. Picheny MA, Durlach NI, Braida LD (1986) Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *J Speech Hear Res* 29:434–446
22. Picheny MA, Durlach NI, Braida LD (1985) Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversational speech. *J Speech Hear Res* 28:96–103
23. Pisoni DB, Bernacki RH, Nusbaum HC, Yuchtman M (1985). Some acoustic-phonetic correlated of speech produced in noise. In: *Proceedings of 1985 international conference on acoustics, speech, and signal processing*, 1581–1584
24. Sandra GS (1987) Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *J Acoust Soc Am* 80:1599–1607
25. Smiljanic R, Bradlow AR (2008) Stability of temporal contrasts across speaking styles in English and Croatian. *J Phonetics* 36:91–113
26. Smiljanic R, Bradlow AR (2007) Clear speech intelligibility: listener and talker effects. Available via <http://www.icphs2007.de>
27. Smiljanic R, Bradlow AR (2005) Production and perception of clear speech in Croatian and English. *J Acoust Soc Am* 118:1677–1688
28. Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI (1996) Speaking clearly for the hard of hearing IV: further studies of the role of speaking rate. *J Speech Hear Res* 39:494–509