IJCSC International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 91-95

Rule Based Age Detection System

Gurpinder Kaur[1] & Arjan Singh[2]

[1]Department of Computer Science & Engg., SUSCET, Tangori
[2]Department of Computer Science & Engg., BBSBEC, Fatehgarh Sahib
Email: [1]gkmann2k3@yahoo.co.in, [2]bhinderas@rediffmail.com

ABSTRACT

Speech has an important role in communication and it is one of the fundamental methods of conveying emotion, age, intensions on a par with facial expression. Prosodic phenomena are specific to spoken language where prosody is the set of features of speech output that includes the pitch contour, duration, fundamental frequency (f0), the pausing, the speech rate, the emphasis on words and many other features. In this paper we found out various attributes of male and female speech corpus of different age groups. We have also make an attempt to find speech parameters while speaking a word and a sentence (Punjabi Language). Analysis has been done and results are evaluated and comparisons have been made on the basis of various parameters of speech of different age groups. This paper briefly describes two studies of speaker age. The first one analyzed and compared a large number of potential age cues, and the second study proposed a method for the synthesis of speaker age.
Keywords: Prosody, Formants, Speech Corpora

1. INTRODUCTION

Numerous acoustic features of speech undergo change with ageing. Age-related variation has been found in duration, F0, sound pressure level (SPL), acoustic correlates of voice quality and spectral energy distribution (phonatory and resonance) [9, 1, 10, 7]. Still, few attempts have been made to establish the relative importance of the various features. Moreover, despite the growing need for voice variation in terms of speaker-specific qualities in speech synthesis applications like spoken dialogue systems and voice prostheses, very few attempts have been made to simulate age using speech synthesis. The complex nature of human speech, which comes from the fact that it varies depending on the speaking style and emotion of the speaker, makes it difficult to find age of the speaker. Intelligibility, naturalness and variability are three features used to compare synthetic speech with human speech [6].

The objective of this work is:

Analysis of speech of a speaker to find various factors affecting the pitch with respect to age.

1. To find out the attributes of speech when a speaker says a word or when a speaker speaks a sentence.
2. Compare various attributes of speech output of different age groups.
3. Draws graphs of various outputs.

Studies on Swedish speech [4] showed that prosodic parameters, fundamental frequency and duration, were not enough to synthesize age. Increasing the parameter space by including voice quality parameters, spectral energy distribution, harmonics-to-noise ratio and articulatory precision has been shown to improve the recognition results for Swedish speech [5]. [6] had worked on to find out whether perceived age is a distracter for voice disguise and speaker identification or not.

2. BACKGROUND AND PREVIOUS STUDIES

In age estimation, the accuracy depends, among other things, on the precision required and on the duration and type of the speech sample (prolonged vowel, read speech etc.). The less acoustic information present in a speech sample [7], the more difficult the task. Speaker and listener characteristics, including gender, age group, the speaker's physiological and psychological state, and the listener's experience or familiarity with similar speakers (dialect etc.) may also influence the accuracy Consequently, some speakers may be more difficult to judge than others. Unfortunately, these studies are often difficult to compare due to differences in the stimuli as well as in the method. Differences concern Language, Stimulus Duration, Type of Speech (Prolonged Vowels, Whispered Vowels, Single Words, Read, Backward Or Spontaneous Speech Etc.), Sound Quality, Speaker Age and Gender, Listener Age and Gender, Recognition Task, Result Measure (Correlation, Absolute Mean Error, % Correct Etc.).

3. PURPOSE AND QUESTIONS

The purpose of this study was to determine how stimulus duration and two different stimulus types (isolated word and sentence) influence estimation of female and male speaker age by answering the following questions:

1. In what way does various speech parameters affect the accuracy of estimation of speaker's age?
2. Is there a difference between perception of female and male speaker age with respect to stimulus duration and type?

4. ANALYSIS

The purpose of the study was to automatically extract and analyze a large number of acoustic features in various segments from a large speech corpus. The aim was to identify the most important acoustic correlates of female and male speaker age.

4.1. Speech Material

The speech samples consisted of 40 female and 40 male versions of the one Punjabi word (Waheguru) by different age groups and one sentence (Man Jeetey Jag Jeet) in Punjabi. The word and sentence has been used because they contained phones with tendencies to contain age-related. The word was produced by 40 female and 40 male speakers, taken from the Gurumukhi corpus and recorded by similar equipment and conditions.

- Test 1: about 10-30 seconds of sentence.
- Test 2: about 3 seconds of one isolated word.

4.2. Method and Procedure

First, a number of Praat [8] scripts were used to normalize word for SPL and transcribe them into phoneme, plosive closure, voice onset time (VOT) and aspiration segments. Then, another Praat script extracted 50 acoustic features (in 8 groups) from the words and same from sentence. Some (e.g. syllables and phonemes per second, jitter and shimmer) were extracted only for the whole file, while others (e.g. F0, formant frequencies and segment duration) were extracted for word and sentence. Table 1 offers an overview of which segments were analyzed in each feature group. Most features were extracted using the built-in functions in Praat. The corpus generated is analyzed manually to normalize the parameters to some extent. F0 and duration are not sufficient to detect age. Interestingly, very different results were obtained by different studies.

Table 1
Segments Analyzed in Each Feature Group

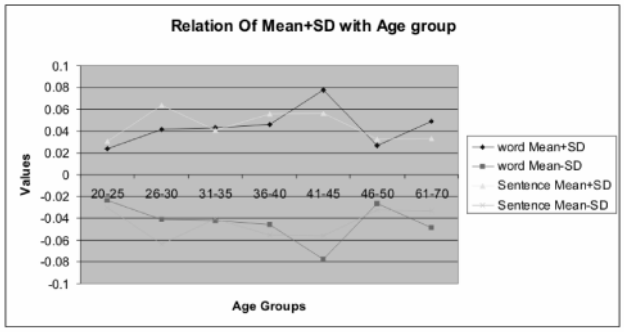
No	*Feature Group*	*Segments Analyzed*
1.	Phonemes/Sec	Word & Sentence
2.	Mean SPL	Word & Sentence
3.	Mean ± SD	Word & Sentence
4.	HNR and NHR	Word & Sentence
5.	F0 (Hz, semitones)	Word & Sentence
6.	Segment Duration	Word & Sentence
7.	Jitter and Shimmer	Word & Sentence
8.	Formant Frequencies(F1-F3)	Word & Sentence

The work has been focused to provide input wave file for pre-processing applications. The graphical representation of speech signal as a function of time is called a waveform. The following more informative characteristics can be extracted or associated with speech signal: energy, pitch, formants, speaking rate, spectrum, cepstrum and transcription [3].

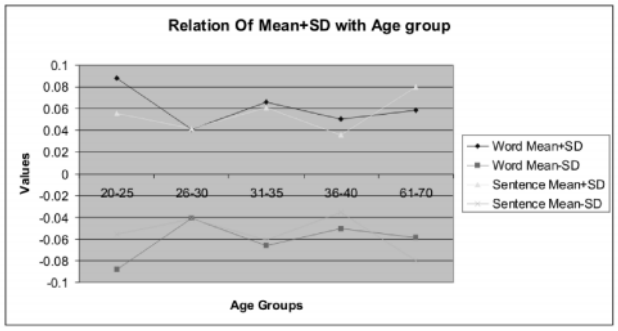
4.3. Results of Analysis

4.3.1. Comparison (Word vs. Sentence) of Female Speech Corpus with Age

1) Relation of Mean ± SD w.r.t Age Groups



(a) Female



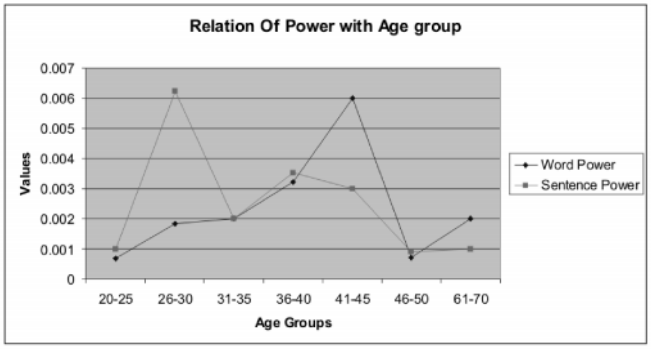
(b) Male

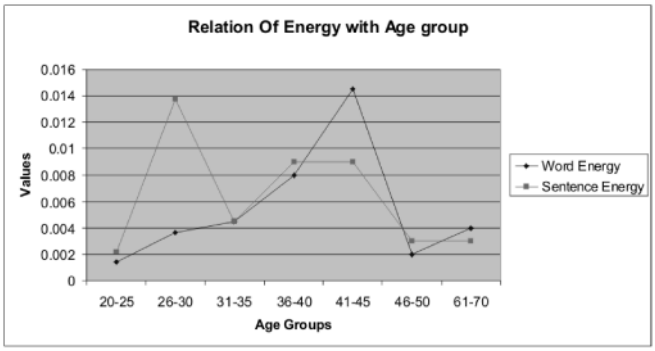
Figure 1: Relation of Mean ± SD w.r.t Age Groups

a) In figure 1(a),the high values of Mean ± SD in case of sentence is at the age of 26-30 years while in word it is higher in 41-45 age groups while in figure 1(b), the range is wider at the age of 61-70 years while in case of word it is wider in 20-30 age groups.

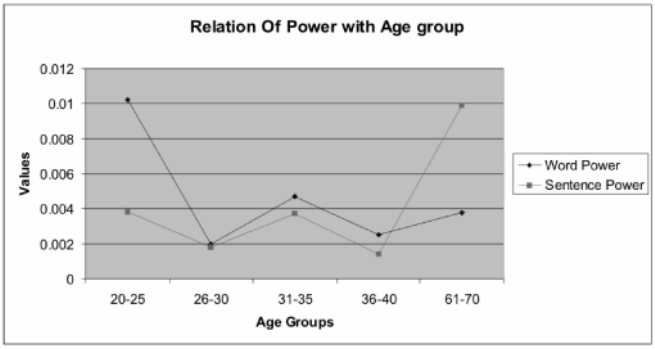
b) After 45 years, the slope starts decreasing gradually to age 50 and then increases slightly in older age in figure 1(a) and vice versa in figure 1(b).

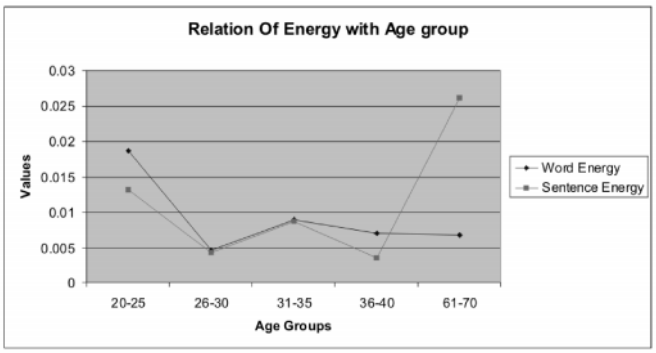
2) Relation of Power and Energy w.r.t Age Groups





(a) Female





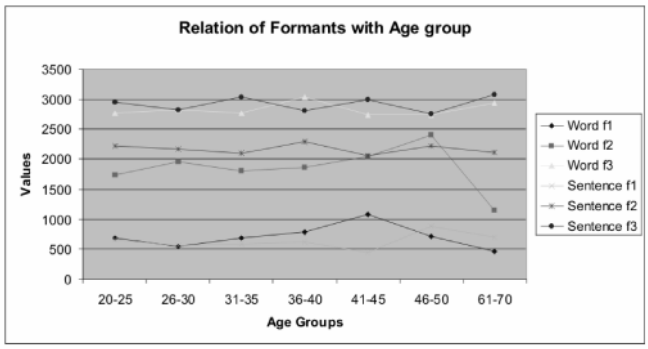
(b) Male

Figure 2: Relation of Power and Energy w.r.t Age Groups

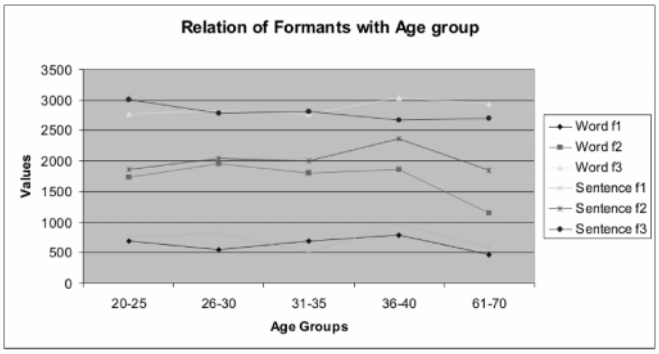
a) In figure 2(a), the high values of Power and Energy are obtained at the age of 26-30 in case of sentence whereas in case of word the values are high at the age of 41-45 while the high values of Power obtained at the age of 61-70 in case of sentence whereas in case of word the values are high at the age of 20-30 in figure 2(b).

b) In case of sentence, the values increases up to middle age and then decreases but the results are reverse in case of word in figure 2 (a) whereas in 2(b), The values of Power remain same at middle ages for both word and sentence.

3) Relation of Formants w.r.t Age Groups



(a) Female



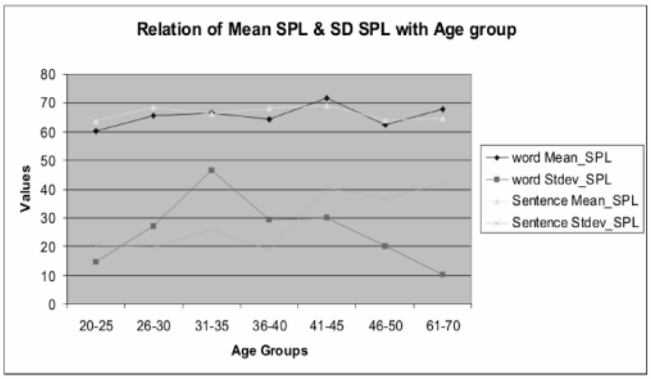
(b) Male

Fig. 3: Relation of Formants w.r.t Age Groups

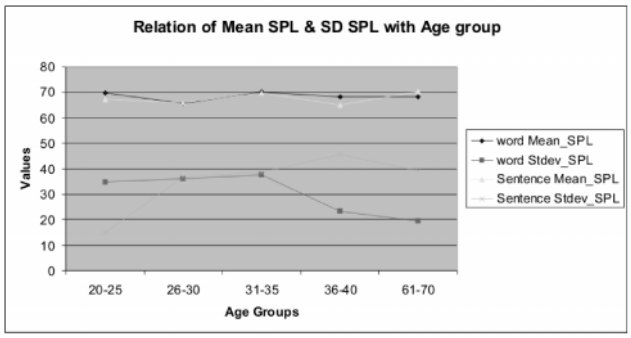
a) The values of F1 formant differs at the age of 41-45 years in female corpora while values of F1 formant differs at the age of 26-30 years in male.

b) In figure 3(a), the difference is large at the age of 61-70 years and in figure 3(b), difference is large at the age of 40-70 years for F2 formant.

4) Relation of Mean SPL and SD SPL w.r.t Age Groups



(a) Female



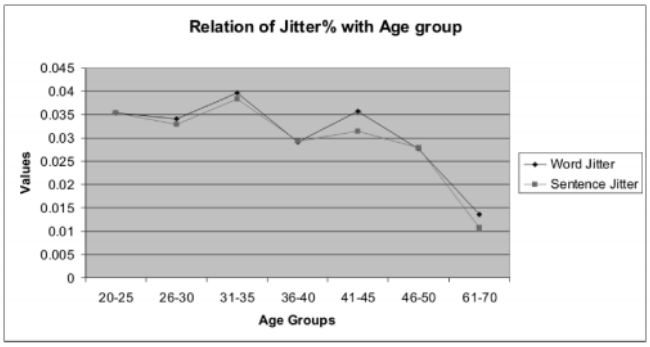
(b) Male

Fig. 4: Relation of Mean SPL and SD SPL w.r.t Age Groups

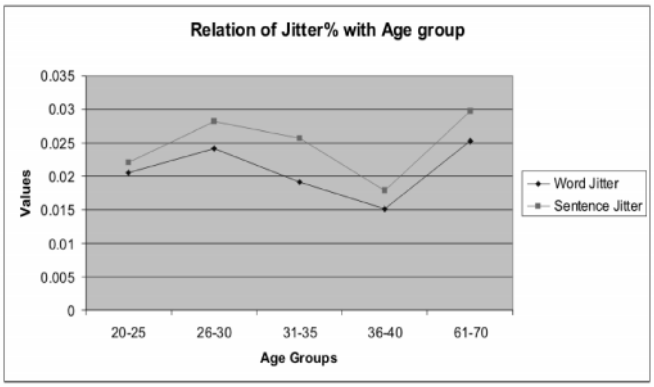
a) The values of SD SPL (sound pressure level) increases large at the age of 30-35 years in case of word.

b) The values of SD SPL (sound pressure level) increases large at older age in case of sentence.

5) Relation of Jitter% w.r.t Age Groups



(a) Female



(b) Male

Fig. 5: Relation of Jitter% w.r.t Age Groups

a) The values of Jitter are higher at younger age but it largely decreases as the age increases in figure 5 (a).

b) The slope in case of jitter increases slightly and the decreases up to 40 years and then increases in older age in figure 5(b).

5. IMPLEMENTATION OF ANALYSIS

5.1. Implementation

The prosodic features are extracted using tool named Praat's programmable scripting language [8]. The reason of choosing Praat as platform is that it provides an existing suite of high quality speech analysis routines, such as pitch tracking. Additional reasons for using Praat include:

(a) Praat is a public domain, widely used speech analysis toolkit that is supported on a variety of platforms (e.g., Windows, Macintosh, Linux, and Solaris).

(b) It provides a variety of valuable data structures, such as TextGrid, PitchTier, and Table, to represent various types of information used for extracting prosodic features.

(c) It provides a built-in programmable scripting language for calling Praat's commands and extending its capability.

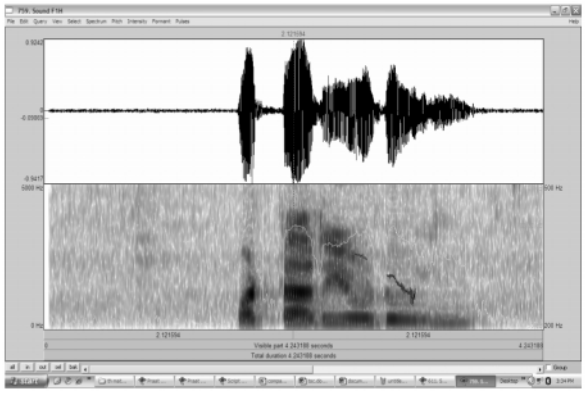
(d) Additions to Praat functionality can be immediately adopted into the tool. This is especially useful for incorporating new prosodic features. Hence, Praat is an ideal platform for building a public domain prosodic feature extraction tool that can be used and extended.

5.2. Sentence

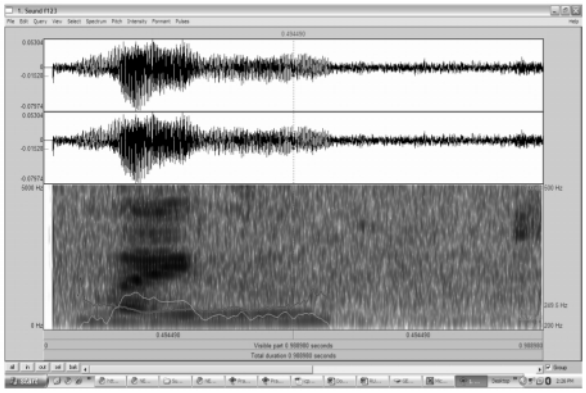
The starting and ending times of sentence is determined. Sentence is also represented in Text Grid Interval Tiers. : The study rely on Praat's autocorrelation based pitch tracking algorithm to extract raw pitch values, using gender dependent pitch range. The raw pitch contour is smoothed and the voiced/unvoiced regions are determined and stored in a TextGrid IntervalTier. Praat's pitch stylization function is used to stylize raw values over each voiced region. Both raw values and stylized values are represented in PitchTiers. The pitch slopes values are generated based on the stylized pitch contour, and are stored in a TextGrid IntervalTier. Intensity values are computed for each frame and stored in an IntensityTier. Since there is no intensity stylization function in Praat, The study choose to represent intensity values in a PitchTier, and apply the pitch stylization function to stylize the intensity contour.

6. WORKING

The first step of this system is input of wave/sound file. This is done with the help of microphones. The figure 13 shows the pitch, intensity, spectrum and duration of the source wave file of sentence.



(a) Sentence



(b) Word

Figure 6: Plot of Wave File of Sentence Showing Pitch, Intensity, Spectrum and Duration

7. RESULTS AND DISCUSSIONS

The vocal tract shape (formant structure) and prosody characteristics such as duration and intonation patterns are studied in this work. It is also discussed that how these characteristics are used collectively, for estimating age. For performing this study, one Punjabi sentence and single Punjabi word is recorded by forty male and forty female speakers of various age groups. For each utterance, duration, average pitch, pitch contour, average frame energy and energy contour (gain contour) are computed. Using these values, comparisons have made for word and sentence. Results are found more fruitful in case of sentence. More parameters are computed while considering sentence.

8. FUTURE WORK INVOLVES

(1) improved parameter extraction for formants,

(2) better interpolation Algorithms, and

(3) expansion of the system to handle more speakers as well as a larger and more varied speech material.

Further research with a larger material is needed to identify and rank the most important age-related parameters. If further developed, the prototype system may well be used in future studies for analysis, modeling and synthesis of speaker age and other speaker-specific qualities, including dialect and attitude. The phonetic knowledge gained from such experiments may then be used in future speech synthesis applications to generate more natural-sounding synthetic speech and age.

REFERENCES

[1] Amerman, J. D., Parnell, M. M. 1992, "Speech Timing Strategies in Elderly Adults", *Journal of Voice 20, 65– 67.*

[2] Charniak, E., "A Maximum-entropy-inspired Parser," Proc. of the 2000 Conference of the North American Chapter of the Association for Computational Linguistics, pp. 132-139, 2000.

[3] L. Rabiner, and B.-H. Juang *Fundamentals of Speech Recognition,* Prentice Hall, Englewood Hills, NJ, 1993.

[4] Schötz, S. (2004) "Prosodic Cues in Human and Machine Estimation of Female and Male Speaker Age" *Proceedings of the IXth Conference*, Lund, 2004. Frankfurt am Main: P. Lang, pp 215-223.

[5] Schötz, S. (2004) "Some Acoustic Cues to Human and Machine Estimation of Speaker Age" *Proceedings of Fonetik 2004*, Stockholm. pp. 40-43.

[6] Traunmüller, H.; Eriksson, A., 2000. "Acoustic Effects of Variation in Vocal Effort by Men, Women and Children", JASA, 107: 3438-3451.

[7] Schötz, S. (2007) "Analysis and Synthesis of Speaker Age" *Proceedings of ICPhS XVI 2007*, Saarbrücken, Germany pp 640-644.

[8] Boersma, P. and Weenink, D. (2005), "Praat: Doing Phonetics by Computer (Version 4.3.04) [Computer Program]. Retrieved 8 March 2005, from http://www.praat.org/.

[9] Ryan, W. J. 1972, "Acoustic Aspects of the Aging Voice", *Journal of Gerontology*, **27**, 256–268.

[10] Xue, S. A., Deliyski, D. 2001, "Effects of Aging on Selected Acoustic Voice Parameters: Preliminary Normative Data and Educational Implications", *Educational Gerontology*, **21**, 159–168.