Prosodic Analysis for Telugu Script

D.NAGARAJU
Research Scholar, Bharatiyar University,
Coimbatoor ,Tamilanadu,India

Dr.R.J.RAMASREE
Professor & Head of ComputerScience,
RSVP,Tirupati,AP, India

Abstract -- Prosodic analysis is very important task in emotional speech .As a part of our research work Design of Emotional speech synthesis system in Telugu(ESST), need to analyze prosodic parameters like Pitch, duration, phonation etc. we have analyzed Telugu alphabet prosodic parameters by using Praat tool.

Keywords:: ESST, prosodic parameters, praat tool ,emotional speech

I. INTRODUCTION

Telugu is one of the prominent and beautiful Indian languages in the state of Andhra Pradesh. During last few decades emotional speech synthesis systems are developed drastically in European languages. Very few Indian emotional Speech synthesis systems are available, but in Telugu it is very new. In the following sessions we have analyzed how the prosodic parameters are analyzed for Telugu alphabet. Naturalness of speech is dependent on prosodic parameters of alphabet.

II. TELUGU ALPHABET ANALYSIS

The script in Indian languages have generated from the Brahmi script. Aksharas are the basic units to write and express the language. The properties of Aksharas are

(1) Orthographic representation
(2) Syllabic in nature.
(3) The typical forms of are V, CV, CCV and CCCV, thus have a generalized form of C*V. The shape of an Akshara depends on its composition of consonants and the vowel, and sequence of the consonants. Based on pronunciation Telugu alphabet are classified into different categories as shown in the table below.

TABLE 1 :ANALYSIS OF TELUGU ALPHABET

Nature of Alphabet	Alphabet
Short vowels	అ ఇ ఉ ఋ ఎ ఐ ఒ ఔ అం
Long Vowels	ఆ ఈ ఊ ౠ ఏ ఓ అః
Sibilants	శ ష స
Nasals	మ ఙ ఞ ణ న
Voiced fricatives	హ
Voice less	క చ ట త ప
Voice less aspirated	ఖ ఛ ఠ థ ఫ ఱు
Voiced unaspirated	గ జ డ ద బ
Voiced aspirated	ఘ డ ధ భ
Voiced alveolars	ర ఴ ల ళ
Semi vowels	య వ

Short Vowel- Short vowel is pronounced short in length.
Long Vowel- these are pronounced longer in length. Both short and long vowels are sounds produced by breath.
Semi Vowel- it is vowel like a sound that acts as consonant.
Sibilants- characterized by hissing sound. Nasals- these are produced with the voice issuing through the nose.
Voiced fricatives- produced by forcing air through a narrow channel placing two articulators close.
Plosives- pronounced by stopping the airflow using the lips, teeth, and followed by a sudden release of air.
Aspirated- pronounced with an exhalation of breath.
Alveolar- pronounced air flow between the tongue and alveolar ridge.

III. EMOTIONS

We are going to implement three different emotions in our emotional speech synthesis system, they are
   1. *Happy* -happiness is an emotional state of human beings characterized by pleasant state.
   2. *Sad* -sadness is an emotional state of human beings characterized by unhappiness or sorrow.
   3. *Joy* -joy is an emotional state of human beings characterized by a feeling of exceptional good.

IV. PROSODIC PARAMETERS

To get emotions we have modified the following prosodic parameters
Pitch Modification --The following modifications are provided for pitch feature changes.
Pitch level -- The overall level of the F0 contour can be shifted by multiplying all values with a rate factor (rate=0 means no change). This means that high values undergo stronger changes than low values and is chosen to conform with the human logarithmic hearing.
*Pitch range --*The pitch range change is motivated by the peak-feature model mentioned in and achieved by a shift of each F0- value by a percentile of its distance to the mean F0-value of the last syllable. If range=0, all values become the last syllable's mean value. The shifting corresponds to a contrast change in image processing.
Pitch variation -- A pitch variation on the syllable-level is achieved by the application of the pitch range algorithm on each syllable separately. The reference value in this case is the syllable's mean pitch.
*Pitch contour (phrase)--*The pitch contour of the whole phrase can be designed as a rising, falling or straight contour. The contours are parameterized by a gradient (in semitones/sec). As a special variation for happy speech, the "wave model" can be used where the main stressed syllables are raised and the syllables, that lie equally distanced in between, are lowered. It's parameterized by the maximum amount of raising and lowering and connected with a smoothening of the pitch contour, because all F0-values get linearly interpolated.
Pitch contour (syllables)-- A rising, falling or level contour can be assigned to each syllable-type.
Intensity changes Intensity -- An Intensity object represents an intensity contour   linearly spaced time points.
Maximum Intensity -- The maximum value within the specified time domain, it is expressed in dB.
Minimum Intensity -- The minimum value within a specified time domain, it represents in dB.
Mean Intensity -- The mean of the intensity values of the frames within a specified time domain.
*Duration --*Time of utterance of an alphabet.
Phonation Modification Jitter -- In order to simulate Jitter (fast fluctuations of the F0-contour) the F0 values can be displaced by a percentile alternating down and up. The pre-processing guarantees that there is a F0 value for each time frame.
Shimmer -- We can measure shimmer in the Sound editor window, Show pulses from the Pulses menu. We will see blue lines that can be thought of as representing the glottal closures.
*Vocal effort --*Diphone synthesis engines usually don't allow the modification of voice-quality related acoustic features. We are using normal voice quality to design voice database. The change of voice quality can be applied to the whole phrase or specific syllable stress types only.

V. SPEECH RECORDING

ABOUT PRAAT
Praat is an important software package to analyze a sound wave. It is very widely used by top-level researchers, it has most of the good features. The activities that can be performed using praat are discussed below Getting a Sound; there are at least three ways to get a sound:
 1. Recording a sound
 2. Reading a sound from disk
 3. Creating a sound from a formula

RECORDING A SOUND
We used a computer and a microphone to record a speech sound into Praat. To record a sound we performed the following steps:
1. Choose Record mono Sound... from the new menu in the Object window. A Sound Recorder window will appear on our screen.
2. Choose the appropriate input device, namely the microphone.
3. Select 16Khz sampling frequency from sound recorder.
4. Use the Record and Stop buttons to record a few seconds of our speech.

5. Use the Play button to hear what we have recorded.
6. Repeat steps 3 and 4 until we are satisfied with our recording.
7. Click the Save to list button. Our recording will now appear in the Object window, where it will be called "Sound".
8. We can now close the Sound Recorder window.
9. When we saved our sound to the Object window, buttons appeared in that window. These buttons show us what we can do with the sound. Use the Play and View & Edit buttons.

SPEECH ANALYSIS
We have a Sound in the List of Objects, the buttons in the Dynamic menu (the right-hand part of The Object window) will show us what we can do with it.
1. Saving a sound to disk
2. Viewing and editing a sound
Saving a Sound to Disk
There are several ways to write a sound to disk. First, the File menu of the Sound Recorder window contains the command to save the left, the right, or both channels of the recorded sound to any one of standard types of sound files (WAV, AIFC, etc)

VIEWING AND EDITING A SOUND
To see the wave form of a sound that is in the list of objects, select that sound and select View & Edit. A sound editor window will appear on our screen. We see a waveform and probably some analyses below it. We can zoom in and scroll to see the various parts of the sound in detail. We can select a part of the sound by dragging with the mouse. We played a part of the sound by click on any of the rectangles below it. Move a selected part of the sound to another location, by using Cut and Paste from the Edit menu.

VI. RESULTS

PITCH ANALYSIS
We Got pitch from the pitch menu in the sound editor ,we get information about the pitch at the cursor. If a cursor is visible in the window, Get pitch writes to the Info window the linearly interpolated pitch at that time. If a time selection is visible inside the window, get pitch writes to the Info window the mean pitch in the visible part of that selection otherwise, get pitch writes the average pitch in the visible part of the sound.

INTENSITY ANALYSIS
To ask for the intensity at the cursor, or the average intensity in the visible part of the selection, choose Get intensity from the Intensity menu or press the F11 key. We get the same value in dB that is visible to the left or right side of the intensity curve, but with more digits.

TABLE 2 :SAMPLE OUT FOR "అ" ALPHABET

Pitch (Hz)				
Median	Mean	Standard	Min	Max
240.87	244.08	13.46	229.76	269.84

Intensity (db)		
Mean intensity	Min intensity	Maximum
76.827	71.805	77.794

Jitter (%)		Shimmer (%)	
Local	Local absolute(Sec)	Local	Local db
0.800	32.724E-6	3.631	0.355
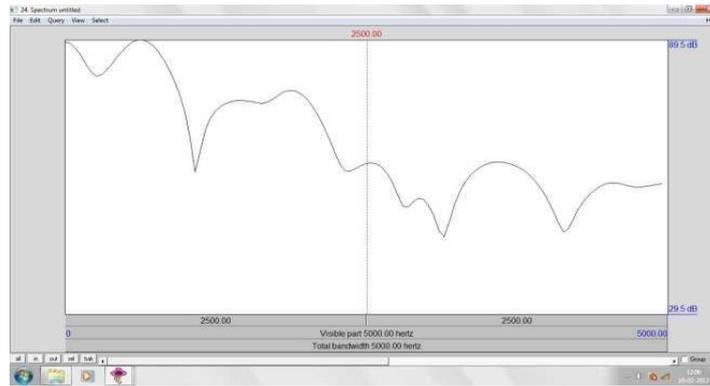
PITCH PARAMETERS FOR VOWELS

TABLE3. PITCH PARAMETERS OF TELUGU VOWELS

Vowels	Pitch (Hz)				Durationseconds
	Median Pitch	MeanPitch	Minimumpitch	Maximumpitch	
అ	240.87	244.08	229.7	269.8	0.11781
ఆ	270.17	268.70	92.03	247.1	0.66161
ఇ	275.02	273.06	258.3	281.0	0.13083
ఈ	265.43	263.34	240.9	280.0	0.39192
ఉ	282.73	294.07	246.2	431.8	0.14785
ఊ	276.53	245.54	93.02	289.7	0.76568
ఋ	300.20	263.55	100.1	320.2	0.59005
ౠ	302.03	293.64	245.7	311.5	0.71982
ఎ	302.23	306.97	225.1	368.4	0.26845
ఏ	232.21	229.35	179.4	239.8	0.47465
ఐ	226.71	214.62	95.04	237.6	0.38933
ఒ	232.88	238.36	191.1	298.9	0.36668

ఓ	245.44	286.87	200.0	475.2	0.49525
ఔ	245.31	206.64	86.25	265.6	0.36668
అం	241.88	240.72	232.8	251.2	0.31703
అః	319.95	317.26	220.9	476.2	0.47051

VII. CONCLUSION

As part my research work so far, we have studied prosodic values for Telugu alphabet. Our next target Is to design and implement a diphone database for telugu language. Diphone database plays a vital role in the implementation of emotional speech synthesis system.

REFERENCES

[1]   Pascal van Lieshout, Ph.D."PRAAT", Oral Dynamics Lab V. 4.2.1, October 7, 2003.

[2]   Scherer K.R. (1996) "Adding the affective dimension: a new look in speech analysis and synthesis" in Proceedings of ICSLP'96

[3]   E. Veera Raghavendra, B. Yegnanarayana, and Kishore Prahallad, "Speech synthesis using approximate matching of syllables," in Proceedings of IEEE SLT Workshop, 2008, pp. 37–40.

[4]   T. Dutoit, V. Pagel, N. Pierret, F. Bataille, and O. Van der Vreken, "The mbrola project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes," Proc. ICSLP'96, Philadelphia, vol. 3, pp. 1393– 1396, 1996.
       NATURAL LANGUAGE PROCESSING, A Paninian Perspective, by Akshar Bharati, Prentice Hall of India.

[4]   Dr. K. V. N Sunitha et al vol 2 (11) 2010, 6269- 6277 ,Bhaashika : Telugu TTS system, International Journal of Engineering Science and Technology.

[5]   Gregor. O. Hofer (2004), Emotional Speech Synthesis, Master of Science, University of Edinburg.

[6]   I.R. Murray, M.D. Edgington, D. Campion, etc. "Rule-Based Emotion Synthesis Using Concatenated Speech," Proc. of ISCA Workshop on Speech and Emotion, Belfast, North Ireland, pp. 173-177, 2000.
       Schroder .M. (2001), "Emotional Speech Snthesis- A Review", Proceedings, Eurospeech 2001.

[7]   T.Dutoit, "An introduction to Text – To – Speech Syntheis", Springer 1997.

ONLINE REFERENCES

1. www.emosamples.syntheticspeech.de
2. www.emotion-research.net
3. http://tcts.fpms.ac.be/synthesis/mbrola.html