National Conference on "Advanced Technologies in Computing and Networking"-ATCON-2015
Special Issue of International Journal of Electronics, Communication & Soft Computing Science and Engineering, ISSN: 2277-9477

Syllable Specific Unit Selection Cost Function Using a Tone Modeling Technique for Automatic Phonetic Segmentation of Hindi Speech Using HMM

Miss.Nutan S.Raut   Dr. Mrs.Sujata N.Kale   Dr. V. M. Thakare

***Abstract* —This paper presents a technique of improving tone correctness in speech synthesis of a tonal language based on an average-voice model trained with a corpus from nonprofessional speakers speech. Unit selection-based concatenative synthesis is one of the widely used speech synthesis approaches. This approach overcomes the limitations of other synthesis techniques such as articulatory synthesis and formant synthesis.The automatic phoneme segmentation framework evolved imitates the human phoneme segmentation process.Public announcements are generally made by creating a library of a set of prerecorded standard sentences for giving only routine information to the passengers of various transport systems like roadways, aircraft and ships.**

***Keywords* — Text to speech synthesis,target cost,concatentation cost,unit selection, Hidden Markov models.**

I. Introduction

Unit selection-based concatenative synthesis is one of the widely used speech synthesis approaches. In this approach, depending on text input, prerecorded speech segments are selected from a large database and concatenated to produce speech. Unit selection framework selects the most appropriate natural segments of speech using syllable-specific concatenation cost and target cost. The automatic phoneme segmentation framework evolved imitates the human phoneme segmentation process. Public announcements are generally made by creating a library of a set of pre-recorded standard sentences for giving only routine information to the passengers of various transport systems like roadways, aircraft and ships.The goal of this paper is modeling tones to improve the tone correctness of synthetic speech.Quantized F0 symbols in the proposed technique were utilized in two different ways based on phone and sub-phone boundary information.

.II. Background

Festival speech synthesis engine by default uses phones as the basic units. The choice of unit size depends on the characteristics of the language. Hybrid systems have been developed by combining unit selection and statistical parametric synthesis approaches. In order to exploit the merits of the two approaches, hybrid systems have been developed.Hybrid systems have been developed by combining unit selection and statistical parametric synthesis
approaches.In order to exploit the merits of the two approaches, hybrid systems have been developed[1]. A more flexible system that can synthesize speech using only a small amount of automatically labeled speech data from nonprofessional speakers is required to widen the applications of speech synthesis. Introducing generative F0 models is another approach to avoiding the problem with insufficient amounts of training data. T-Tilt model, which is an expansion of Tilt model, was proposed to model the F0 contours of tonal languages[2]. Hidden Markov model is an extension of discrete Markov model in which states are hidden, The hidden Markov models may be used to represent the sequence of sounds within a section of speech.
Phoneme, an elemental speech sound, can be modeled by an individual HMM. In this paper work is done on the automatic phonetic segmentation is to modify an HMM based recognizer to perform the task of phonetic segmentation., in order to reduce manual efforts and speed up the segmentation process[3].

. III. Related Work Done

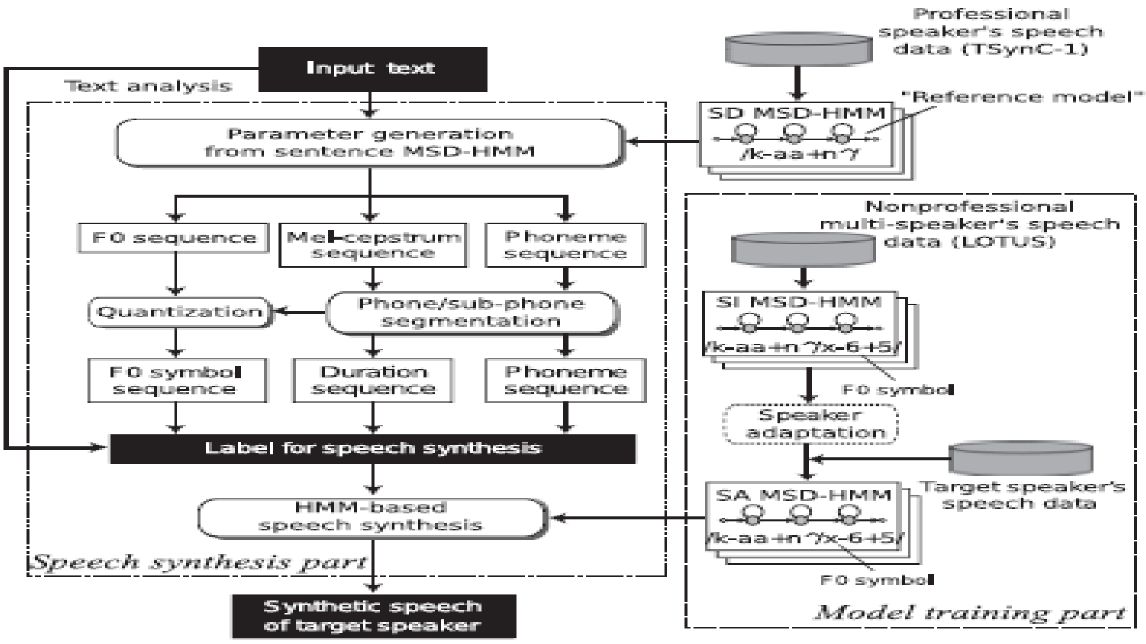
Concatenation cost gives an estimate of the quality of the join between candidate units $ui-1$ and ui for the desired target units. It predicts the degree of perceived discontinuity. Concatenation cost can be split into sub costs, where each of the subcosts, indicates one continuity metric.
In the Festival speech synthesis engine, concatenation cost is calculated based on optimal coupling technique, but the problem in that is the cost of concatenating two units at their best position of join is determined. To overcome this problem the proposed segmental concatenation cost is defined based on the type of segments present at the syllable joins[1]. Two approaches study to obtain the tonal context labels, i.e.,phone-based and sub- phone-based quantized F0 symbols,but the generated phone-based F0 symbol sequence is insufficient to represent correct information on tone contours, so to overcome this problem, propose subphone- based F0 symbols where quantize the F0 contour based on units smaller than phones[2].Hidden Markov model is an extension of discrete Markov model in which states are hidden,The hidden Markov models may be used to represent the sequence of sounds within a section of speech.Performing manual segmentation is considered to be a precise method of obtaining segmented database and information about its contents(labeling). But, the main limitation of this is, manual segmentation is time consuming, labor intensive and prone to error. So, in this paper work is done on the automatic phonetic segmentation is to modify an HMM based recognizer to perform the task of phonetic segmentation., In order to reduce manual efforts and speed up the segmentation process, it is desirable to design an efficient method for automatic phonetic labeling, especially when the size of the speech corpus is large[3].

IV. TONE MODELLING TECHNIQUE

Major methodology works on the waveform generation module or back end of syllable TTS, which selects the units from the database based on a unit selection algorithm selects the units from the database [1]. Two approaches study to obtain the tonal context labels, i.e.,phone-based and sub-phone-based quantized F0 symbols , The generated phone-based F0 symbol sequence is insufficient to represent correct information on tone contours[2].To solve this problem, propose subphone- based F0 symbols where quantize the F0 contour based on units smaller than phones.Performing manual segmentation is considered to be a precise method of obtaining segmented database and information about its contents (labeling). But, the main limitation of this is, manual segmentation is time consuming, labor intensive and prone to error. Most of the previous methods do not elaborate the issue of error analysis, such as what categories of phonemes tend to be error-prone and how to deal with them[3].

V. OVERVIEW OF TTS SYATEM

In this paper Proposed Methodology using Model training & Speech synthesis. In Model training first, an average voice model is trained using multiple data from nonprofessional speakers' speech. The spectrum and F0 are modeled with multi-stream HMMs in which the output distributions for the spectral and F0 parts are modeled using a continuous probability distribution for the former and a multi-space probability distribution. & in speech synthesis Reference model of a professional speaker to generate the proposed labels.When an input text is given, the conventional context labels are automatically generated from the given text by text analysis and a synthetic F0 contour is generated from the reference model using conventional labels shown in fig. The F0 context labels are created using the F0 quantization.fig shows overview of proposed TTS system[2].



VI. ANALYSIS & DISCUSSION

The parser takes the text as the input, extracts the phonemes and arranges them according to the sequence of occurrence in the given sentence, and the list of the phonemes is given as the output. The overall segmentation results in terms of percentage of phonemes falling into different ranges of deviation. It provides the number of tokens occurring for the every phoneme present in the database. The three average-voice models were constructed using different tonal context labels, i.e., conventional, phone-based F0 symbol and sub-phonebased F0 symbol. The reference model was trained using 2500 sentences from the TSynC-1 speech data uttered by a professional speaker, where the conventional tonal context labels were used. The tonal context labeling propose can generate an F0 contour closer to that of natural speech than the other techniques. The test-set consisted of 10 sentences. All syllables present in the test-set were available in the training database. Unit selection cost functions, namely concatenation cost and target cost, are proposed for syllable based synthesis. Concatenation costs are defined based on the type of segments present at the syllable joins.

CONCLUSION

Separate sets of concatenation costs and target costs are proposed appropriate to syllables.Concatenation costs are proposed which is based on the type of segments present at syllable joins. A technique of modeling tones to improve the tone correctness of synthetic speech. Quantized F0 symbols in the proposed technique were utilized in two different ways based on phone and sub-phone boundary information.The phoneme boundary estimate by 36 dimensions HMM, . a comparison has been made against hand-segmented and hand-labeled phonemes.For each of the phoneme occurring in the speech database,the difference between estimated start-end points & actual start-end points was computed and average was calculated. This is called deviation and is used as a measure of the performance of the segmentation.

FUTURE WORK

The proposed cost metrics have reduced perceptual discontinuities at the syllable joins. In addition to linear regression models, different nonlinear models such as ANN(artificial neural network) and SVM(support vector machines)can be explored for further improvement.The segmentation performance,& the better segmentation rates can be obtained by increasing the number of training speech sentences for future work.

REFERENCES

[1] N. P. NARENDRA and K. SREENIVASA RAO "Syllable Specific Unit Selection Cost Functions for Text-to-Speech Synthesis", ACM Transaction on Speech & Language Processing,VOL.9,NO..3, Nov-2012.

[2] Vataya Chunwijitra , Takashi Nose, Takao Kobayashi, ,"A tone-modeling technique using a quantized F0 context to improve tone correctness in average-voice-basedspeech synthesis" ScienceDirect,Vol.54,No.2,PP.245-255, Feb-2012.

[3] Archana Balyan ,S. S. Agrawal , Amita Dev," Automatic phonetic segmentation of Hindi speech using hidden Markov model", ,vol.27, No.4, PP.543-549 ,Nov-2012.

.

AUTHOR'S PROFILE

Miss.Nutan S. Raut has completed B.E. Degree in Computer Science and Engineering from Sant Gadge Baba Amravati University, Amravati, Maharashtra. He is persuing Master's Degree in Computer Science and Information Technology from P.G. Department of Computer Science and Engineering, S.G.B.A.U. Amravati. His current research interest is focus on the area Text to speech synthesis.(e-mailid: nutan.raut@yahoo.co.in)

Dr.Mrs.Sujata N.Kale is Assistant Professor in PG department of Applied Electronics,SantGadgebaba Amravati University. She has 21 years of teaching experience. She received her doctoral degree in Machine Intelligence from SantGadgebaba Amravati University,Amravati. Her current research interests focus on the areas of image processing, signal processing,neuralnetworks.(sujatankale@rediffmail.com)

Dr. V. M. Thakare

Dr. Vilas M. Thakare is Professor and Head in Post Graduate department of Computer Science and engg, Faculty of Engineering & Technology, SGB Amravati university, Amravati. He is also working as a coordinator on UGC sponsored scheme of e-learning and m-learning specially designed for teaching and research. He is Ph.D. in Computer Science/Engg and completed M.E. in year 1989 and graduated in 1984-85.

He has exhibited meritorious performance in his studentship. He has more than 27 years of experience in teaching and research. Throughout his teaching career he has taught more than 50 subjects at various UG and PG level courses. He has done his PhD in area of robotics, AI and computer architecture. 5 candidates have completed PhD under his supervision and more than 8 are perusing the PhD at national and international level. His area of research is Computer Architectures, AI and IT. He has completed one UGC research project on "Development of ES for control of 4 legged robot device model.". One UGC research project is ongoing under innovative scheme. At PG level also he has guided more than 300 projects/discretion. He has published more than 150 papers in International & National level Journals and also International Conferences and National level Conferences. He has also successfully completed the Software Development & Computerization of Finance, Library, Exam, Admission Process, and Revaluation Process of Amravati University. Also completed the Consultancy work for election data processing. He has also worked as member of Academic Council, selection Committee member of various Other University and parent university, Member of faculty of Engineering & Science, BOS (Comp. Sci.), Member of IT Committee, Member of Networking Committee, Member of UGC, AICTE, NAAC, BUTR, ASU, DRC, RRC, SEC, CAS, NSD etc committees. . He has also worked chairman of many committees like BOS, Monitoring and Control, New Installations, Curriculum design and developments etc. He has organized more than 50 Summer schools / STTP/ Conferences / Seminar /Symposia / Workshop /Orientation Program/Training/Program / Refresher Courses. He is member and fellow of Learned Societies like Institute of Engineers, Indian Society of Technical Education ISTE, Computer Society of India CSI, etc. He has delivered more than More than 70 Keynote addresses and Invited talks delivered in India and abroad at the occasion of International & National level Technical/social events, International Conferences and National level Conferences, and also acted as session chairs many times. 3 times he has received National Level excellent paper award at National Conference, Gwalior and at other places. He has also received UGC fellowship and a major UGC project. (e-mailid: vilthakare@yahoo.co.in)