Speaking tool in Tamil for vocally disabled

Shashikiran K[1], Abhinava[1], Swapnil Belhe[2], A G Ramakrishnan[1]

[1]MILE Lab, Dept of Electrical Engineering, Indian Institute of Science, Bangalore 560 012.

[2]Gist Group, CDAC Pune.

Shashikiran@mile.ee.iisc.ernet.in, abhinav@mile.ee.iisc.ernet.in, ramkiage@ee.iisc.ernet.in

swapnilb@cdac.in

Introduction:

It has always been a challenge to bridge the gap between vocally disabled and the masses. The development of sign language has only been partially successful in bridging the gap. It requires persons conversing to know the sign language. Our work is a conscious effort to overcoming this pitfall. The proposed methodology is a combination of two different entities namely (Online Handwriting Recognition) OHR and (Text to Speech) TTS. OHR deals with recognition of handwritten words online. In an on-line handwriting recognition system, machine recognizes the writing while the user writes. The OHR output is in Tamil Unicode & becomes the input to TTS. Unicode is a globally accepted encoding format which makes our application viable to be used in various circumstances. TTS is a system that takes Unicode text and produces natural and intelligible speech in that language. This enables the patients, who had laryngectomy and tracheotomy as well as the vocally challenged to communicate effectively. As vocal disability may be congenital or because of ailments like oral or throat cancer, our method serves equally to both.

The tool is based on a hand-held, Tablet PC based on Intel Atom processor. The user can write one sentence at a time on the screen using the stylus and then click the button “Speak”. The sentence is recognized and then converted into speech and spoken out. Thus, the patient can call the attention of the nurses or his relatives in another room easily.

Details of the OHR module:

In Online handwriting recognition, a machine recognizes, as a user writes on a pressure sensitive screen with a stylus. The stylus captures information about the position of the pen tip as a sequence of points in time. The sequence of point between a PEN DOWN and PEN UP signal defines a stroke. This spatio-temporal information of the character being traced is the only input available to the online recognition system. Also given a character, one can capture the different writing styles using the information from the stylus.

Given a Tamil word, we first run a segmentation algorithm to identify the individual symbols. This algorithm segments word level data into symbol level data, as the modeling of the

data is done at the symbol level. The recognition is performed at this level and results are concatenated to form the words. The extracted symbols are subject to the following preprocessing modules: smoothing to remove noise, resampling to a fixed number of points for speed normalization and size normalization.

Once symbols are brought to a standardized form, a set of seven features namely ,

Pre-processed X & Y co-ordinates: Preprocessed data points (x,y) are themselves good features
Pen direction angle: At each sample point, the direction of pen tip movement from that point to the next point can be used as a feature
Normalized first derivatives of X & Y: Derived at each sample point of the preprocessed Tamil symbol, are also used.
Normalized Second derivatives of X & Y: Same as above.

The preprocessing techniques & features are discussed elaborately in Rituraj et al [2]. These features are then fed to the SDTW classifier for recognition of the Tamil symbol.

Statistical Dynamic Time warping (SDTW): In SDTW, a reference character is represented by a sequence Q=(Q1;Q2;Q3;Qlq) of statistical quantities (states), as shown in Fig 1.

These statistical quantities include

1) Discrete probabilities that statistically mode the transitions between states. We have empirically used 20 states in our work.
2) A continuous probability density function that models the feature distribution at each state. We have modeled this distribution as a multivariate Gaussian distribution for each of the 20 states.

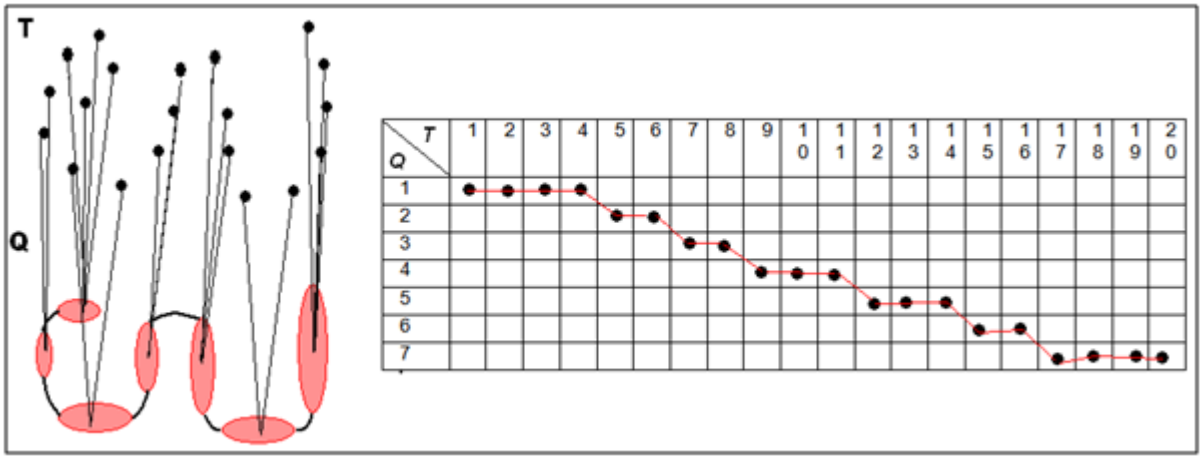

Fig 1. Transitions between states in SDTW

While testing, the SDTW distance of test pattern to the reference model of each class is computed. and the test pattern is assigned the label of the class giving minimum SDTW distance. The definition of SDTW distance is different from that of DTW and is given in Claus et al

[1].Fig.1 shows how the matching takes place between the reference model and Test pattern. Matrix in Fig.1 shows the DTW path (path of best SDTW matching).

SDTW distance is the negative log state optimized likelihood of pattern T generated by the model Q, with optimal state sequence S, given by the Viterbi algorithm. So, models in SDTW frame work are similar to HMMs of particular type with state prior probabilities π=(100…0)T and are left to right models with step size of at most 1 and with null transitions (transitions that allow change in state without observation change i.e. transitions (0,1)). So the models in SDTW frame work can be trained by algorithms used for training HMMs. In our work, we used segmental K-means algorithm [3] for training SDTW model parameters.

Unicode Generation: Based on rules derived from the language, valid Unicode string is generated from the output labels of the classifier. This string is the input for the TTS module.

Details of the TTS module:

The TTS is based on concatenative waveform synthesis [4]. There are 1026 phonetically rich sentences spoken by a professional Tamil speaker, which has been segmented and annotated at the phone level. The input text is passed through a grapheme to phoneme conversion module, after text normalization. Certain pause rules are added based on a preliminary POS tagger and several rules. The phonemic text is parsed into the basic units for concatenation and the units are searched for, from the synthesis speech database, based on context and prosodic parameters. The selected speech units are concatenated in the waveform domain using a pitch continuity metric [5] and a pitch synchronous concatenation methodology [6]. It takes directly Unicode Tamil text as the input and produces a .wav file as the output. This wav file is directly played by the tablet PC and hence, when the user writes a word, it will be read after a second by the system.

References:

1. Clauss Bahlmann and Hans Burkhardt , ”The writer independent online handwriting recognition system Frog On Hand and cluster generative statistical dynamic time warping”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.26, No. 3, pp 299-310, March 2004.
2. Rituraj Kunwar et al. A HMM Based Online Tamil Word Recognizer, 8thTamil Internet Conference, 2009, University of Cologne, Germany.
3. Rakesh Dugad and U. B. Desai ”A tutorial on hidden Markov models”, SPANN- 96.1 May 1996.
4. K Parthasarathy, “A research bed for unit selection based text to speech synthesis system”, MS thesis, Dept. of electrical Engineering, IISc Bangalore, Feb. 2009.
5. Vikram Ramesh Lakkavalli, Arulmozhi P and A G Ramakrishnan, “Continuity Metric for Unit Selection based Text-to-Speech Synthesis”, Proc. SPCOM 2010, Bangalore, July 23-26.