

CHAPTER 8

RESULTS AND DISCUSSION

8.1 SUMMARY AND CRITICISM

The primary objective of this work is to design an efficient Tamil Text to Speech Synthesis system using Annotated Dual Corpora. The research work has involved in identifying speech units that can be used for designing such an efficient synthesizer for the Tamil language. The existing systems use either diphones or syllables. Those systems included an appropriate Prosody model if ever needed. When a dual corpus is used with an appropriate Prosody model, it provides a brilliant Speech Synthesis System. Both the corpora should be annotated with prosodic information to an extent – Diphones need extensively; Syllables need in a minute. The Unit Selection based synthesis provides a framework where fractions of recorded and annotated speech are concatenated together. Those Syllables chosen by the selection algorithm should optimally minimize both the target cost (how close a database unit is to the desired unit - C_t) and the join cost (how well two adjacently selected units to join - C_j). Appropriate entries of diphone database are also concatenated upon the same criteria. When used in the unit selection synthesis technique, the selected speech unit is easy to extract and model into acoustics.

In this thesis, a group delay based segmentation algorithm was used to identify syllables from continuous speech data. Diphone corpus was created by adding appropriate Prosody information with the existing set of Tamil diphones. The intended synthesizer was built by integrating syllable and diphone units with the Festival framework, and using the unit selection



algorithm. This synthesizer is capable of generating natural sounding synthesized speech. It is vivacious in this thesis that using syllables along with diphones at specific places of articulation makes the Synthesizer more intelligent. Though there is an overhead in maintaining two databases, the naturalness of the synthesized speech balances the cost. The thesis also demonstrates that this approach is general and is suitable for other Dravidian languages as well. This thesis has also addressed the issues in building speech synthesizers for embedded devices. We have proposed a Front-end application for Distributed Speech Synthesis. The front end includes provisions for scaling the synthesizer to accommodate some more Dravidian Languages. The users are allowed to select one among the existing languages.

The naturalness of synthetic speech using unit selection concatenation techniques is improved over the years due to the improved control upon the prosody of the generated speech. In real applications, the database may not have units that match both the desired spectral and prosodic features. In this case, a feedback procedure is necessary to include new units with non-existing features.

It is evident that having large amounts of data in a unit selection database would lead to a better synthesis. It becomes more likely that a unit that is closer to the target and more likely to have a better join is always present in the database, thanks to the size of the database. It is important not just to collect every possible unit but instead to cluster the ‘right’ set of units which are most frequently used in a particular language along with their prosodic variants. This phenomenon requires a ‘phonetically balanced’ set of sentences used for creating the speech database. The Tamil synthesizer developed in this thesis has been created using such a phonetically balanced set of sentences. The speech corpus should also be recorded such that the energy levels, the rate of speech and pitch remain uniform across all the units.



8.2 FUTURE DIRECTIONS

The syllables and diphones are good speech units for Indian languages. A dual database scenario discussed in the thesis sets a new trend for database creation. It is now used for other languages in the Dravidian spectrum, which have high demands for TTS systems like Indian English and other regional dialects. The dual-database concept in such languages could be carried out. Due to the context switching process between the databases, overall synthesis time increases, and this leads to the depleted efficiency of the synthesizer. In future, when techniques are employed for overcoming this phenomenon, the quality of the synthesizer will increase.

