Punjabi Text-To-Speech Synthesis System

Parminder SINGH[1] *Gurpreet Singh LEHAL*[2]
(1) GURU NANAK DEV ENGINEERING COLLEGE, Ludhiana, Punjab, India
(2) PUNJABI UNIVERSITY, Patiala, Punjab, India
parminder2u@gmail.com, gslehal@gmail.com

ABSTRACT

Speech based interface can play a vital role for the successful implementation of computerized systems for masses. As a tool for this purpose, effort has been made for the development of a Text-To-Speech (TTS) synthesis system for Punjabi language written in Gurmukhi script. Concatenative method has been used to develop this TTS system. Syllables have been reported as good choice of speech unit for speech databases of many languages. Since Punjabi is a syllabic language, so syllables has been selected as the basic speech unit for this TTS system, which preserves within unit co-articulation effects. System involves development of algorithms for pre-processing, schwa deletion and syllabification of the input Punjabi text, as well as speech database for Punjabi. A syllable based Punjabi speech database has been developed that stores articulations of syllable sounds at starting, middle and end positions of the word for producing natural sounding synthesized speech.

KEYWORDS : Speech synthesis, Punjabi speech database, Punjabi syllables.

Proceedings of COLING 2012: Demonstration Papers, pages 409–416,
COLING 2012, Mumbai, December 2012.

1 Introduction

Text-To-Speech (TTS) synthesis system has extensive range of applications in everyday life. In order to make the computerized systems more interactive and helpful to the users, especially physically and visibly impaired and illiterate masses, the TTS synthesis systems are in great demand for the Indian languages. Concatenative speech synthesis technique has been used for the development of this system. Punjabi is a syllabic language (Singh, 2002), so syllables has been selected as the basic speech units and output waveform is generated by concatenating the syllable sounds, which preserves within unit co-articulation effects (Raghavendra, Desai, Yegnanarayana, Black, & Prahallad, 2008; Narayana & Ramakrishnan, 2007). Syllable sounds in different contexts have been marked in pre-recorded sound file and stored in the speech database to get natural sounding synthesized speech.

1.1 Research background

1.1.1 Punjabi language

Punjabi is an Indo-Aryan language spoken by more than hundred million people those are inhabitants of the historical Punjab region (in north western India and Pakistan) and in the Diasporas particularly Britain, Canada, North America, East Africa and Australia. It is written from left to right using Gurmukhi (an abugida derived from the *Laṇḍā* script and ultimately descended from Brahmi script) as well as Shahmukhi (a version of the Arabic script) scripts. This TTS system for Punjabi language has been developed for Gurmukhi script. In Gurmukhi script, which follows the *"one sound-one symbol"* principle, the Punjabi language has thirty eight consonants, ten non-nasal vowels and same numbers of nasal vowels (see Figure 1) (Singh & Lehal, 2010).

ਸ ਹ ਕ ਖ ਗ ਘ ਙ
ਚ ਛ ਜ ਝ ਞ ਟ ਠ ਡ ਢ ਣ
ਤ ਥ ਦ ਧ ਨ ਪ ਫ ਬ ਭ ਮ
ਯ ਰ ਲ ਵ ੜ ਸ਼ ਖ਼ ਗ਼ ਜ਼ ਫ਼ ਲ਼ } Consonants

ਇ ਈ ਏ ਐ ਅ ਆ ਔ ਉ ਊ ਓ Non-nasal Vowels
ਇੰ ਈਂ ਏਂ ਐਂ ਅੰ ਆਂ ਔਂ ਉਂ ਊਂ ਓਂ Nasal Vowels

FIGURE 1 – Punjabi consonants and vowels

1.1.2 Punjabi syllables

Defining syllable in a language is a complex task. There are many theories available in phonetics and phonology to define syllable. In phonetics, the syllables are defined based upon the articulation (Krakow, 1999). However in phonological approach, the syllables are defined by the different sequences of the phonemes. So, combination of phonemes gives rise to next higher unit called syllable. Further, combination of syllables produces larger units like morphemes and words. So, syllable is a unit of sound which is larger than phoneme and smaller than word. In every language, certain sequences of phonemes and hence syllables are recognized. Using these phonetic sequences and hence structures, all possible syllables can be formed those have been discovered so far in ancient and recent

literary works. In addition, all theoretically possible syllables can be composed that may or may not yet be used in a language, but valid in the sense that these follow rendering rules for the language at present (Joshi, Shoff, & Mudur, 2003). A syllable must have a vowel, without vowel, syllable cannot exist. In Punjabi seven types of syllables are recognized (Singh, 2002) – V, VC, CV, VCC, CVC, CVCC and CCVC (where V and C represents vowel and consonant respectively), which combine in turn to produce words. The occurrence of syllables of last type CCVC is very rare, and has not been considered in the present work.

Punjabi language has thirty eight consonants, ten non-nasal vowels and same numbers of nasal vowels; so, the above said seven syllable types results 11,27,090 syllables in Punjabi with non-nasal vowels and the same number of syllables with nasal vowels and thus giving total of 22,54,180 syllables in Punjabi.

1.1.3 Schwa deletion in Punjabi language

Schwa is a mid-central vowel that occurs in unstressed syllables. Phonetically, it is a very short neutral vowel sound, and like all vowels, its precise quality varies depending on its adjacent consonants. Each consonant in Punjabi (written in Gurmukhi script) is associated with one of the vowels. Other vowels, except schwa ('ਅ' the third character of Punjabi alphabet and written as [ə] in International Phonetic Alphabet (IPA) transcription), are overtly written diacritically or non-diacritically around the consonant; however schwa vowel is not explicitly represented in orthography. The orthographical representation of any language does not provide any implicit information about its pronunciation and is mostly ambiguous and indeterminate with respect to its exact pronunciation. The problem in many of the languages is mainly due to the existence of schwa vowel that is sometimes pronounced and sometimes not, depending upon certain morphological factors. In order to determine the proper pronunciation of words, it is necessary to identify which schwas are to be deleted and which are to be retained. *Schwa deletion* is a phonological phenomenon where schwa is absent in the pronunciation of a particular word, although ideally it should have been pronounced (Choudhury, Basu, & Sarkar, 2004). The process of schwa deletion is one of the complex and important issue for grapheme-to-phoneme conversion, which in turn is required for the development of a high quality text-to-speech (TTS) synthesizer. In order to produce natural and intelligible speech, the orthographic representation of input has to be augmented with additional morphological and phonological information in order to correctly specify the contexts in which schwa vowel is to be deleted or retained (Narasimhan, Sproat, & Kiraz, 2004).

Mostly phonological schwa deletion rules have been proposed in literature for Indian languages. These rules take into account morpheme-internal as well as across morpheme-boundary information to explain this phenomenon (Narayana & Ramakrishnan, 2007). The morphological analysis can improve the accuracy of schwa deletion algorithm which is a diachronic and sociolinguistic phenomenon (Singh, 2002; Singh & Lehal, 2010). The syllable structure and stress assignment in conjunction with morphological analysis can also be used to predict the presence and absence of schwa (Tyson & Nagar, 2009).

Vowels, except schwa ([ਅ]), are represented diacritically when these come along with consonants (known as half vowels), otherwise as such. The consonant sound varies according to the vowel attached to consonant. For example, consonant [ਸ] conjoined with

vowel [ਈ] (having diacritic ੀ) results a single orthographic unit “ਸੀ”, having pronunciation of a consonant-vowel sequence /ਸ+ਈ/ (/si/) however when this consonant comes with vowel [ਆ] the resulting single unit [ਸਾ] will be pronounced as /ਸ+ਆ/ (/sā/).

Consonants represented in orthography without any attached diacritic, basically have the associated inherent schwa vowel that is not represented diacritically. While pronouncing any written word, the speaker retains the intervening schwa vowel associated with a consonant where required and eliminate it from pronunciation where it is not required. In Punjabi, inherent schwa following the last consonant of word is elided. For example, Punjabi word “ਸੜਕ” ([sədəkə] means road) pronounced as \ ਸ ਅ ੜ ਕ \ (\s ə d k\) is represented orthographically with only the consonant characters [ਸ], [ੜ] and [ਕ]. Schwa following the last consonant [ਕ] is deleted as per rule said above and deletion of schwa following the second consonant [ੜ] makes the word monosyllabic of type CVCC (Consonant-Schwa-Consonant-Consonant).

2 Implementation

The working of this Punjabi TTS system can be divided into two modules: Offline Process and Online Process. These two subparts are discussed in the following subsections.

2.1 Offline process

Offline process of this TTS system involved development of the Punjabi speech database. In order to minimize the size of speech database, effort has been made to select a minimal set of syllables covering almost whole Punjabi word set. To accomplish this all Punjabi syllables have been statistically analyzed on the Punjabi corpus having more than hundred million words. Interesting and very important results have been obtained from this analysis those helped to select a relatively smaller syllable set (about first ten thousand syllables (0.86% of total syllables)) of most frequently occurring syllables having cumulative frequency of occurrence less than 99.81%, out of 1156740 total available syllables (Singh & Lehal, 2010). An algorithm has been developed based on the set covering problem for selecting the minimum number of sentences containing above selected syllables for recording of sound file in which syllable positions are marked. The developed Punjabi speech database is having starting and end positions of the selected syllable-sounds labeled carefully in pre-recorded sound file. As the pronunciation of a syllable varies depending on its position (starting, middle or end) in the word, so separate entries for these three positions has been made in the database for each syllable. In order to increase the naturalness of the system a good number of most frequently occurring words of corpus have been stored in the database as such.

2.2 Online process

Online process is responsible for pre-processing of the input text, schwa deletion, syllabification and then searching the syllables in speech database. First module, *Pre-processing* involves expansion of abbreviations, numeric figures and special symbols etc. present in the input text to the full word form, so that these should be spoken correctly.

Second module, *Schwa Deletion* is an important step for the development of a high

quality Text-To-Speech synthesis system. During utterance of words not every schwa following a consonant is pronounced. For proper pronunciation of word, schwas which are not to be uttered have to be identified in the orthographic form of word for deletion. A rule based schwa deletion algorithm has been developed for Punjabi. These rules are based on mainly three parameters: grammatical constraints, inflectional rules and morphotactics of Punjabi language, which play important role for identification of schwa to be deleted for correct pronunciation of input word. For example, the vowel-consonant pattern for Punjabi word "ਮਰਦ" ([mərədə] means man) is CCC. Grammatically, there must be schwa vowel following each consonant in Punjabi but the word's pronunciation specifies the existence of schwa [ə] /ਅ/ sound after the first consonant only. So, schwa following the first consonant [ਮ] will be retained, however schwa vowels following the second [ਰ] and third [ਦ] consonants will be deleted. The accuracy of developed schwa deletion algorithm is about 98.27% (Singh & Lehal, 2011).

Third module, *Syllabification* of the words of input text is a challenging task. A universal tendency for syllables to have onsets has long been claimed in phonological theory. This preference is built into rule based approaches to syllabification formulated, for example, as the onset first principle (Kahn) or the core syllabification principle (Clements) (Chiosain, Welby & Espesser, 2012). A rule based syllabification algorithm has been developed, which syllabifies the input word into the corresponding syllables. Syllabification rules are mostly language specific. For syllabification the phonotactic constraints of Punjabi have been followed. So, rules have been devised based on the position of vowels and consonants in a word, to segment input word into corresponding syllables. For example, table 1 below shows syllabification of some words; where C, V and n stands for consonant, vowel and nasal respectively.

Input word	**CV pattern of input word (after Schwa deletion)**	**Output of Syllabification module**	**CV pattern of word syllables**
ਆਇਆ	VVV	ਆ ਇ ਆ	V V V
ਇੰਗਲੈਂਡ	VnCCvnC	ਇੰਗ ਲੈਂਡ	VnC CvnC
ਸਮਾਜਿਕ	CSCvCvC	ਸ ਮਾ ਜਿਕ	CS Cv CvC
ਕਰਵਾਉਣ	CSCCvVC	ਕਰ ਵਾ ਉਣ	CSC Cv VC
ਰਾਜਨੀਤਕ	CvCCvCC	ਰਾਜ ਨੀਤਕ	CvC CvCC
ਟੂਰਨਾਮੈਂਟ	CvCCvCvnC	ਟੂਰ ਨਾ ਮੈਂਟ	CvC Cv CvnC
ਪਾਕਿਸਤਾਨ	CvCvCCvC	ਪਾ ਕਿਸ ਤਾਨ	Cv CvC CvC
ਅਧਿਕਾਰੀਆਂ	VCvCvCvVn	ਅ ਧਿ ਕਾ ਰੀ ਆਂ	V Cv Cv Cv Vn
ਵਿਦਿਆਰਥੀਆਂ	CvCvVCCvVn	ਵਿ ਦਿ ਆਰ ਥੀ ਆਂ	Cv Cv VC Cv Vn

TABLE 1 – Output of the syllabification module.

The developed syllabification module has been tested on about first 10,000 most frequently used words of Punjabi, selected from a Punjabi corpus having about 1,04,42,574 total words and 2,32,565 unique words. It has been found that accuracy of the syllabification module is about 97.89%.

Syllables of input text are first searched in the Punjabi speech database for corresponding syllable-sound positions in recorded sound file and then these syllable sounds are

concatenated. Normalization of the synthesized Punjabi sound is done in order to remove discontinuities at the concatenation points and hence producing smooth and natural sound. Synchronous-OverLap-Add (SOLA) method is used to achieve a shorter or longer playback time than original waveform. A good quality sound is being produced by this TTS system for Punjabi language.

Conclusions

A fairly good quality Punjabi Text-To-Speech synthesis system has been developed for Punjabi. During the development of this TTS system, it has been observed that for a concatenative speech synthesis system, the important features that must be taken care of are: selection of basic speech unit for concatenation, statistical analysis of selected speech units on corpus, corpus must be carefully selected and unbiased; and marking of the speech units in recorded sound file. The last one is most important and quality of the output speech depends, how carefully speech units are marked in recorded sound file. Correct schwa deletion is very important for natural pronunciation of the output synthesized speech. Schwa basically controls articulation of the sound wave and hence pronunciation. Syllabification process is language specific and involves grammatical rules of the language as well as knowledge of how local inhabitants syllabify a word during its utterance in a natural way. Syllabification with good accuracy is providing a strong base for high quality of the output synthesized speech.

References

Chiosain, M.N., Welby, P. & Espesser, R. (2012). Is the syllabification of Irish a typological exception? An experimental study. *Journal of Speech Communication*, 54(1): 68–91.

Choudhury, M., Basu, A. & Sarkar, S. (2004). A Diachronic Approach for Schwa Deletion in Indo Aryan Languages. In *Workshop of the ACL Special Interest Group on Computational Phonology (SIGPHON), Association for Computations Linguistics*, pages 20–27, Barcelona.

Joshi, R.K., Shoff, K. & Mudur, S.P. (2003). A unified phonemic code based scheme for effective processing of Indian Languages. In *23rd Internationalization and Unicode Conference*, Prague.

Krakow, R.A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics*, 27: 23–54.

Narasimhan, B., Sproat, R. & Kiraz, G. (2004). Schwa-Deletion in Hindi Text-to-Speech Synthesis. *International Journal of Speech Technology*, 7(4): 319–333.

Narayana, M.L., Ramakrishnan, A.G. (2007). Defining syllables and their stress in Tamil TTS corpus. In *Workshop on Image and Signal Processing (WISP-2007)*, pages 92–95, IIT Guwahati, India.

Raghavendra, E.V., Desai, S., Yegnanarayana, B., Black, A.W., & Prahallad K. (2008). Global syllable set for building speech synthesis in Indian languages. In *IEEE Workshop on Spoken Language Technologies (SLT-2008)*, pages 49–52, Goa, India.

Singh, P. (2002). *Sidhantik Bhasha Vigeyan*, 4th Edition, Madan Publications, India, pages 371–372.

Singh, P. & Lehal, G.S. (2010). Statistical Syllables Selection Approach for the Preparation of Punjabi Speech Database. In *5th International Conference for Internet Technology and Secured Transactions (ICITST-2010)*, pages 1–4, London, UK, IEEE.

Singh, P. & Lehal, G.S. (2011). A Rule Based Schwa Deletion Algorithm for Punjabi TTS System. In *International Conference on Information Systems for Indian Languages (ICISIL-2011)*, vol. 139, pages 98–103, Patiala, India, Springer.

Tyson, N.R. & Nagar, I. (2009). Prosodic Rules for Schwa-Deletion in Hindi Text-to-Speech Synthesis. *International Journal of Speech Technology*, 12(1): 15–25.