International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

Extraction of Di-phones for Telugu ::Issues and solutions

D.NAGARAJU
Research Scholar, Bharathiar University,
Coimbatoor ,Tamilanadu,India,
e-mail:dubisettynagaraju@gmail.com,

Dr.R.J.RAMASREE
Professor Department of Computer Science,
RSVP, Tirupati, AP, India
e-mail: rjramasree@yahoo.com

Abstract- This paper describes a method for extraction of diphones to generate diphone database for concatenative text to speech systems. Diphone is an adjacent pair of phones. Diphone is a very important resource for both text to speech [TTS] and speech to text [STT]. Consider the pronunciation of కాకి -kaaki. It consists of phonemes క్ [k], ఆ [a], ఆ [a], క్[k], ఇ[i]. The diphones generated while pronouncing the above are a) – క్ [-k] b) క్ఆ [ka], c) ఆ ఆ [aa] d) ఆ క్ [ak] e) క్ ఇ [ki] f) ఇ- [i-]. It is attempted to generate all diphones of given sound.

Generation of diphones involves a) annotation of speech signal with corresponding transliterated scheme b) segmentation of the speech signal into diphones. Diphone concatenation has the advantage of simplicity and relatively small database of speech, when compared to the other methods.

Key words: Diphone synthesis – annotation – segmenting-Mbrola –TTS- Telugu etc

I. INTRODUCTION

In concatenative speech synthesis, corpus generation involves selection of diphones. Diphones are extracted manually from recorded speech database. This paper describes the extraction of diphones and testing of diphone quality for natural speech signal.

Diphone concatenation has the advantage of simplicity and relatively small database of speech, when compared to the other methods [7].

II.RELATED STUDIES

Italian Text to speech synthesis system is having about 1100 diphone units [12]. Italian TTS used automatic extraction of diphones. Maori TTS system uses the Mbrola supported diphone database [13].

Extraction of phone and diphones are complex and most rigorous operation. Success or failure of diphone database is depends on diphones. Quality of phone is an important aspect here, because diphones are extracted from stable phone. To get quality phone concentrate on spectral, phase, fundamental frequency, duration, voicing characteristics [10]. In automatic annotation select smallest possible set of sentences to get maximum diphones[11].

III.DIPHONE SYNTHESIS

Diphone synthesis explained in the following terms

A. phone set
B. Diphone set
C. Audio collection.
D. Praat tool
E. Diphone extraction

A. PHONE SET

In Telugu language sound is a combination aksharas. Akshara is the fundamental linguistic unit of telugu language. To extract diphones of the given telugu sound, It is necessary to list the phonemes in Telugu language. Aksharas consists of consonants (C) and vowels (V) i.e. Vowel-V- అ, CV- క, CVV- కై , CCV- క్ష . There are 16 Vowels and 36 consonants in the Telugu language [4].

Ex: 1.అ –a and ఆ- aa

ఇ – i and ఈ – ii etc

Ex 2. అమ్మ -amma

This word consists of two characters. The first letter is the single vowel అ-'a'. The second letter is a combination of two consonants and a vowel ◌ం- ◌ం- అ 'm-m-a'.

Each character in Telugu has a primary and a secondary form. The primary form is used when this symbol forms the base of a character; the secondary form is used when it is combined into another base. So in order to add another consonant to an existing consonant-vowel combination, just add the secondary form of that new consonant to the existing symbol.

a) Acchulu- అచ్చులు

a aa i ii u uu R Ru e ea ai o oe ou aM a@h

అ ఆ ఇ ఈ ఉ ఊ ఋు ఋూ ఎ ఏ ఐ ఒ ఓ ఔ అం అః

b) Hallulu- హల్లులు

k kh g gh ~m

క్ ఖ్ గ్ ఘ్ ఙ్

c ch j jh ~n

చ్ చ్ జ్ ఝ్ ఞ్

T Th D Dh N

ట్ ఠ్ డ్ ఢ్ ణ్

t th d dh n

త్ థ్ ద్ ధ్ న్

p ph b bh m

ప్ ఫ్ బ్ భ్ ం

Y r l v S sh s h lh ksh ~r

య్ ర్ ల్ వ్ శ్ ష్ స్ హ్ ళ్ క్ష్ ఱ్

fig:1 List of phones in Telugu.

Allophones- .Same phone have different pronunciations are known as allophones examples are

1. /i/ - పిల్లి -pi̲lli – unaspirated

   - పిల్ల -pilla – pi^{h}̲lla-aspirated

2. /u/ - పుట్టుక –pu̲TTuka- unaspirated

   - పుట్ట- puTTa -pu^{h}TTuka - aspirated

B. DIPHONE SET

To create the diphone database of a certain language, it is necessary a list of the phonemes in that language. There are 16 vowels and 36 consonant and hence a total of 53 phonemes in the Telugu language[including SIL]. P is the number of phones in any language, and then maximum possible diphones are P^2. Included SIL as phone then diphones is P^2-1. Total diphones of the Telugu language are 53^2-1= 2808. SIL-SIL diphone not possible in any language.

C. AUDIO COLLECTION/ RECORDING CORPUS

In speech recording, selection of speaker is an important task. Few people have better voice for synthesis than others. In general people with, clearer, more consistent voices are better than others. Professional speakers are in general better for synthesis then non-professional. We are selected a professional speaker.

The recordings were made at semi-professional recording studio in two sessions. Recordings are made without interruption and maintaining constant pitch. The signals were sampled at 16 KHz and quantified 16 bits per sample.

D. PRAAT TOOL

Praat is a software package, useful to develop, analysis and reconstruct of speech signal in phonetics. It was designed by Paul Boersma and David Weenink of the University of Amsterdam[1]. It is freely available for most platforms. It includes articulator synthesis, spectrographic analysis etc. it operates on various operating systems like Windows, Unix, Linux and Mac [2]. Advantage of praat is easy interface and default options try to learn. We can do make waveforms generation, intensity contour, pitch tracks, recordings, edit recorded sound extract sounds, get pitch, intensity, draw a plot etc [3].

E. DIPHONE EXTRACTION

Use either manually or Automatic system to extract diphones from sound. We are using manual system to extract diphones from sound. Diphones are speech units that begin in the middle of the stable state of a phone and end in the middle of the following one. Middle phone is more stable than edge of the phone. Stable parts in phones are a) for stops one third [1/3], b) phone silence –one quarter [1/4] and c) for other diphones -50%[1/2] [6].

In first stage we are searching for phones and diphones in an annotated sentence (sample as shown in figure). In second stage cut the identified diphones from annotated recordings. Architecture of extraction diphones as shown in figure1.

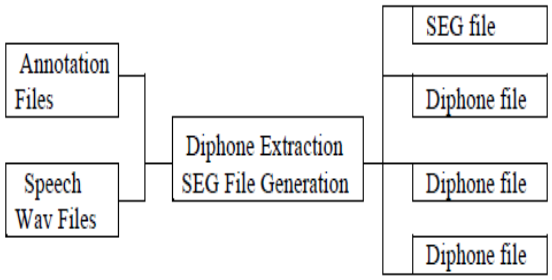
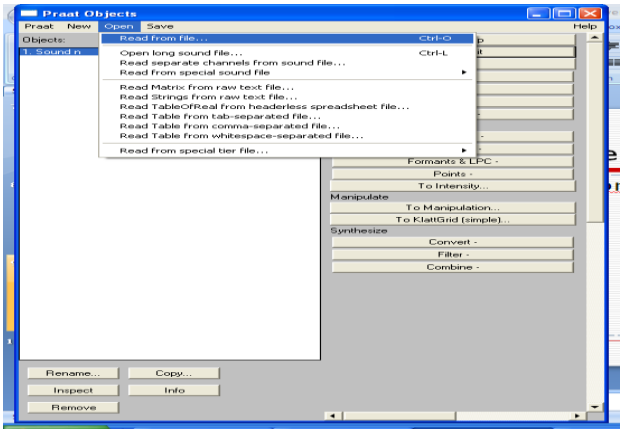


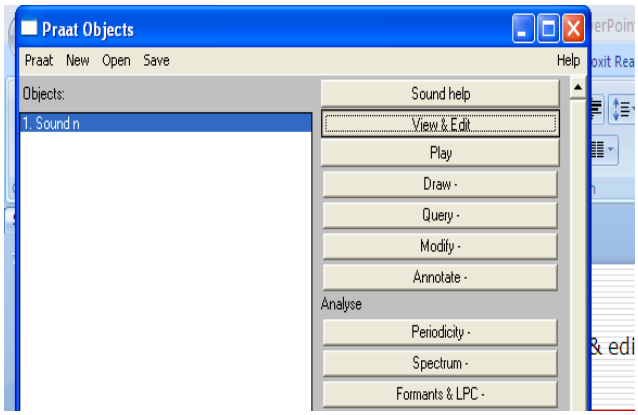
Fig1: Architecture of diphone extraction

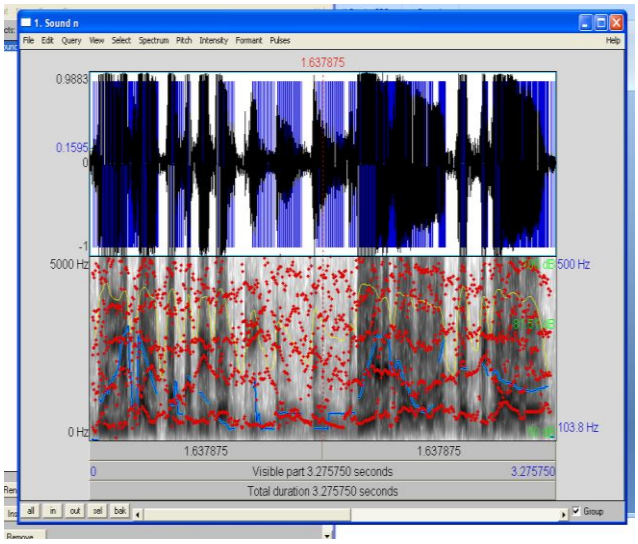
Step by step procedure to extract diphones is given below.

Step1. Open praat tool and read wav file

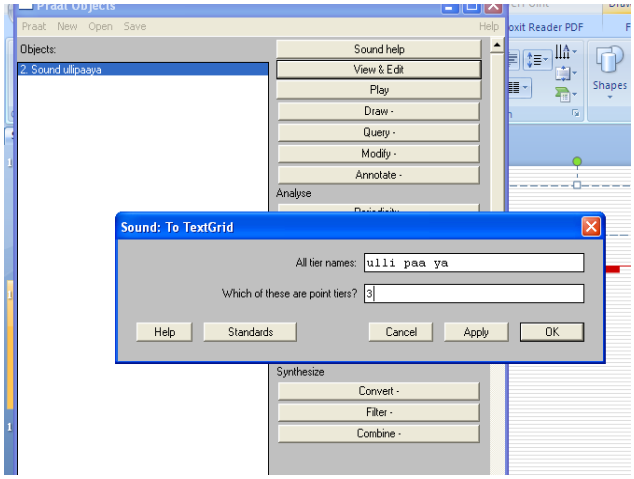


Step2. Click on view and edit command

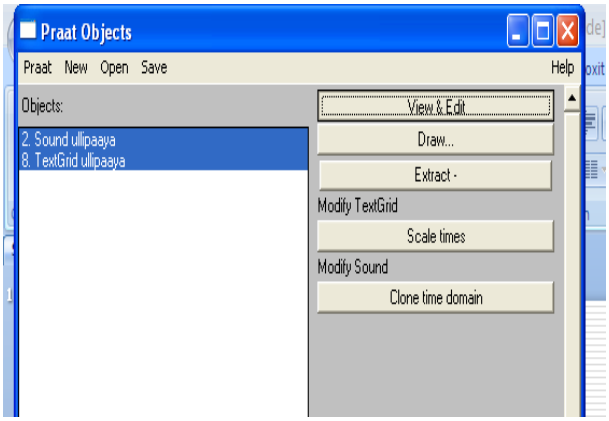


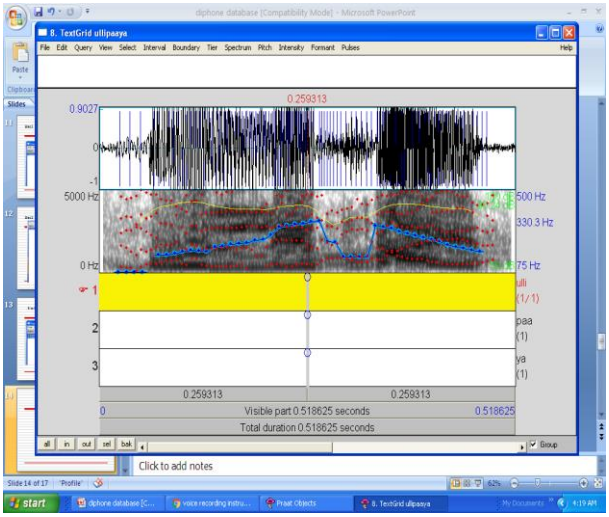


Step3. Create annotation wav file

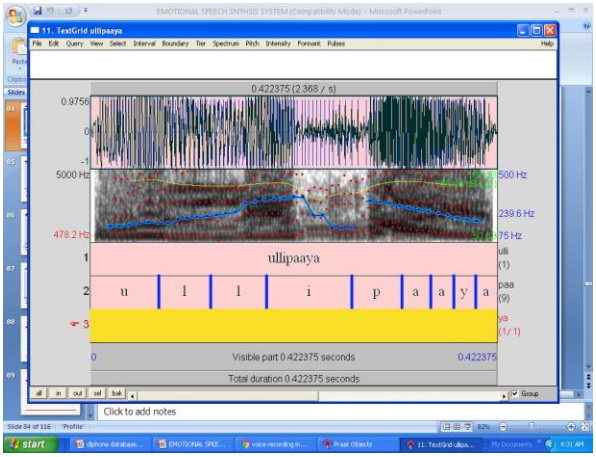


Step 4. View & Edit both sound and text grid files

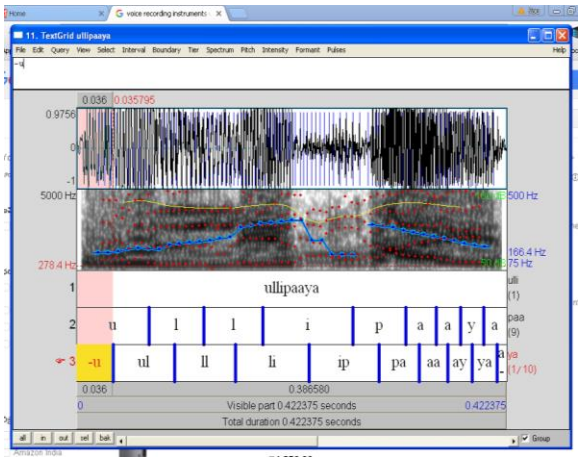




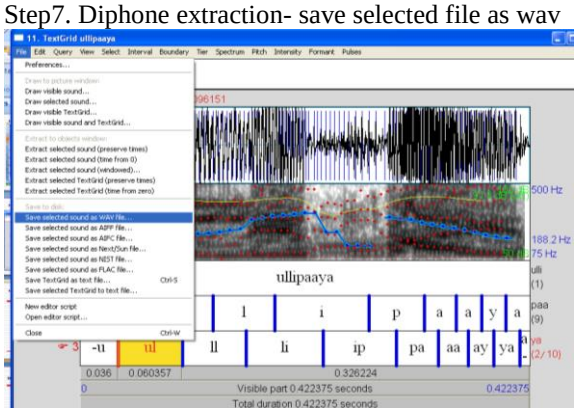
Step 5. Phone selection



Step6. Diphone selection



Step7. Diphone extraction- save selected file as wav



IV. QUALITY CONTROL –CHECKING AND CORRECTING DIPHONES

After alignment and labeling is finished. The results are visually inspecting / checking every one or randomly or all individual [8]. Common mistakes are a) mislabeling b) mispronunciation of phones and c) wrong position of pitch marks etc. Mislabeling corrects to test each and every diphone[9]. Mispronunciation solved to synthesis and listen each prompt, if it is wrong extraction of diphone, go back and select correct pitch marking.[10]

V. CONCLUSIONS

In this study 106 sentences/622 words are recorded and 1356 natural [non emotion] diphones are extracted. By using Telugu natural diphones a diphone database was generated. This database is used to get emotional speech after adding emotional parameters.

REFERENCES

[1]http://web.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421 .pdf
[2] https://en.wikipedia.org/wiki/Praat
[3]http://ec-concord.ied.edu.hk/phonetics _and_ phonology /wordpress/learning_website/chapter_1i ntroduction_ new .htm #1.2.1
[4] http://www.omniglot.com/writing/telugu.htm
[5]http://timesofindia.indiatimes.com/city/hy der aba d/Telangana-slang-can-do-without-31-Telugu-let ter s -Telangana-University-professor-says/artic le sho w/ 30744511.cms
[6] nlp.stanford.edu/courses/lsa352/lsa352.lec4.p pt
[7] H. Timothy Bunnell, Steven R. Hoskins, and Debra M. Yarrington. A biphone constrained concatenation method for diphone synthesis. http://www.isca-speech.org/archive_ op en /arc hive _ papers/ssw3/ssw3_171.pdf
[8] Kevin A. Lenzo , Alan W Black , Diphone collecrion and synthesis. https://www.cs.cmu.edu/ ~awb/ papers/ ICSLP2000_diphone.pdf
[9] Building Synthetic Voices Alan W Black Kevin A. Lenzo. http://festvox.org/bsv/bsv.pdf
[10] Thomas Ewender and Beat Pfister . Automatic ceating a diphone set from a speech database. http://www.tik.ee.ethz.ch/file/07a8 e26e6e4d188bf9a6e661b9c2523d/Ewender:11.pdf
[11] Jolanta Bachan-Efficient diphone database creation for Mbrola, a multilingual speech synthesizer. http://mechatronika.polsl.pl/owd/ pdf2010 /303.pdf
[12] Automatic diphone extraction for an Italian text-to-speech synthesis system. http://mirlab.org/conference_papers/International_Conference/Eurospeech%201997/pdf/tmb/a0688.pdf
[13] Speech data analysisi for diphone construction of a Maori online TTS synthesizer http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.7636&rep=rep1&type=pdf