

Concatenative based Pashto Digits and Numbers Synthesizer

Sahibzada Abdur
Rehman Abid
Department of
Computer Systems
Engineering,
University of
Engineering &
Technology,
Peshawar, Pakistan.

Nasir Ahmad
Department of
Computer Systems
Engineering,
University of
Engineering &
Technology,
Peshawar, Pakistan.

Muhammad
Akbar Ali Khan
Department of
Computer Systems
Engineering,
University of
Engineering &
Technology,
Peshawar, Pakistan.

Fatima Tuz Zuhra
Department of
Computer Science,
University of
Peshawar, Pakistan.

ABSTRACT

This paper presents a Pashto digits synthesis system that produces the Pashto digits from zero (Sefer) to one thousand (Yaw Zar) using a set of Pashto recorded digits stored in the speech database. Speech database for the Pashto digits synthesizer consist of digits from Sefer (0) to Naha (9), Las (10), Shul (20), Dairsh (30), Celwaikht (40), Panzoos (50), Shpeta (60), Awaya (70), Atya (80), Nawi (90), Sul (100), and Zar (1000). The recording of the Pashto digits has been carried out using Sony PCM-M 10 linear Recorder in a noise free environment. The recording has been done from one person. The recorded audio file of the Pashto digits is split into isolated digits by utilizing Adobe Audition ver 1.0. The Pashto digits are synthesized through concatenative technique using isolated Pashto digits present in the speech database.

General Terms

Speech synthesis, Digits synthesis, Concatenative synthesis

Keywords

Pashto digits, Pashto digits recording, Pashto digits synthesis

1. INTRODUCTION

The world is moving rapidly towards automation and there is an increase in the interaction of humans with the machinery in their daily lives with each passing day, so there is a strong need to make human computer interface more natural and pervasive [1]. To accomplish this target, the development of listening [2], speaking [3] and viewing [4], [5] machines has turned into a very hot area of research. The area of research that converts text into speech is known as speech synthesis or text to speech synthesis (TTS). Speech synthesis produces human like speech artificially. A text-to-speech (TTS) system transfers normal language text, into acoustic signal. In our proposed Pashto digits synthesis, the text that we will be using and given as input are only numbers.

The three basic techniques for synthetic speech production are; formant synthesis, articulatory synthesis and concatenative synthesis. Articulatory synthesizers are physical models based on the full depiction of the physiology of speech generation and the physics of sound production in the vocal apparatus [6]. Formant synthesis is an illustrative, acoustic-phonetic technique to speech synthesis [7]. In formant synthesis fundamental frequency, and noise levels like parameter are changed over time to produce a waveform of a

synthetic speech. Concatenative synthesis is the production of natural sounding synthesized speech waveforms by choosing and concatenating speech units from a database [8]. It is the most simple but efficient method of generating natural and intelligible artificial speech. Formant and concatenative synthesis techniques are most commonly used in the speech synthesis, while the articulatory synthesis technique is the least utilized because of its complex implementation.

In this work the concatenative technique has been used where synthetic speech is produced by concatenating pieces of prerecorded speech segments stored in the speech database. The quality of a speech synthesizer is measured as its capability to be identified and by its likeness to the human speech. The most significant qualities of a speech synthesis system are intelligibility and naturalness. Naturalness means how similar is the generated sound to the human natural speech, where as intelligibility is the ease with which the generated sound is understood by a listener. The ideal speech synthesizer needs to be natural as well as intelligible; therefore the speech synthesis research generally aims to develop techniques that produce speech which is both natural and intelligible.

The paper is organized as follows. Section 2 introduces the Pashto language and its digits. Section 3 describes the Pashto digits recording and its synthesis. Section 4 explains Pashto Digits synthesis results and discussion.

2. PASHTO LANGUAGE AND ITS DIGITS

Pashto is the national language of Afghanistan and one of the widely spoken languages of Pakistan, having an estimated 50-60 million speakers all over the world [9]. Pashto language has three main dialects, Northern Pashto, Central Pashto and Southern Pashto [10]. In Khyber Pakhtoonkhwa (a province of Pakistan) and its bordering areas Kunar and Nangahar (Afghanistan provinces) the Northern Pashto is mostly spoken. The Northern Pashto is also spoken mostly by the Pushtun people living abroad. This dialect is also named as Pakhto, Pukhto and Yusufzai Pukhto.

In Bannu, Waziristan, and Karak areas of the Khyber Pakhtoonkhwa Central Pashto is verbal. It is also called as Waciri (Waziri) and Bannuchi (Bannochi, Bannu). Southern Pashto is spoken in Quetta area of Balochistan province and also in Afghanistan, Iran, Tajikistan, United Arab Emirates, and United Kingdom.

In Pashto language the number system is similar to the most further foreign languages for example Persian [2], Arabic [11] and other languages [9], in which every digit from zero to ten has different name, similarly in the Pashto language case every digits from Sefer (0) to Las (10) has different pronunciation. The numbers subsequent to Las (10), Yawo-las

Table: 1 Pashto digits of Yousafzai/ Northern Pashto dialect

Digits	Pashto words	Pashto digits	Pronunciation
0	صفر	۰	Sefer
1	يو	۱	Yaw
2	دوه	۲	Dwa
3	درے	۳	Dray
4	څلور	۴	Celour
5	پنځه	۵	Penza
6	شپږ	۶	Shpeg
7	اووه	۷	Owa
8	اته	۸	Ata
9	نهه	۹	Naha
10	لس	۱۰	Las
11	يو لس	۱۱	Yawo- Las
12	دوه لس	۱۲	do- Las
13	ديار لس	۱۳	Dyar- Las
14	څوار لس	۱۴	Swaar- Las
15	پنځه لس	۱۵	Penza- Las
16	شپاړس	۱۶	Shparh-as
17	اووه لس	۱۷	Owa- Las
18	اته لس	۱۸	Ata- Las
19	نو لس	۱۹	Noo- Las
20	شل	۲۰	Shul
21	يو ویشت	۲۱	Yaw-Vesht
22	دوه ویشت	۲۲	Dwa—Vesht
23	درے ویشت	۲۳	Dray-Vesht
24	څلور ویشت	۲۴	Celour-Vesht
25	پنځه ویشت	۲۵	Penza-Vesht
26	شپږ ویشت	۲۶	Shpag-Vesht
27	اووه ویشت	۲۷	Owa-Vesht
28	اته ویشت	۲۸	Ata-Vesht
29	نهه ویشت	۲۹	Naha-Vesht
30	دیرش	۳۰	Dairsh
40	څلویښت	۴۰	Celwaikht
50	پنځوس	۵۰	Panzoos
60	شپيته	۶۰	Shpeta
70	اوبه	۷۰	Awaya
80	اته	۸۰	Atya
90	نوی	۹۰	Nawi
100	سل	۱۰۰	Sul
1000	زر	۱۰۰۰	Zar

(11) to Noo-las (19) have similar postfix Las with each number from Yaw (1) to Naha (9) with small variation, whereas from Yaw-Vesht (21) to Naha-Vesht (29) the similar postfix is Vesht with number from Yaw (1) to Naha (9) as a prefix. The same prototype is followed for all numbers upto sul (100). For the numbers greater than hundred i.e. from Yaw-Yaw-Sul (101) to Naha-Nawi-Yaw-Sul (199) in this case for the tens numbers the same criteria has repeated as in case of from Yaw (0) to Sul (100) with addition of common postfix Yaw-sul and the same criteria has been repeated for number upto (999) but the difference is that only postfix changes e.g. number from Yaw-Dwa-Sul (201) to Naha-Nawi-Dwa-Sul (299) have postfix Dwa-Sul, similarly numbers from Yaw-Dray-Sul (301) Naha-Nawi-Dray-Sul (399) Yaw-Celour-Sul (401) Naha-Nawi-Celour-Sul (499) Yaw-Penza-Sul (501) Naha-Nawi-Penza-Sul (599) Yaw-Shpeg-Sul (601) Naha-Nawi-Shpeg-Sul (699) Yaw-Owa-Sul (701) Naha-Nawi-Owa-Sul (799) Yaw-Ata-Sul (801) Naha-Nawi-Ata-Sul (899) Yaw-Naha-Sul (901) Naha-Nawi-Naha-Sul (999) have common postfix of Dray-Sul, Celour-Sul, Penza-Sul, Shpeg-Sul, Owa-Sul, Ata-Sul, and Naha-Sul respectively.

There are various modifications of this prototype amongst different areas/dialects. The digits incorporated in this paper are from the Yousafzai/ Northern Pashto dialect listed in Table 1.

3. PASHTO DIGITS RECORDING AND SYNTHESIS

First of all the Pashto digits have been recorded. The speech database is then developed from the recorded digits. The details of the recording and database development are provided in the following subsections.

3.1 Pashto Digits Recording

The process of Pashto Digits recording has been depicted in Figure 1.

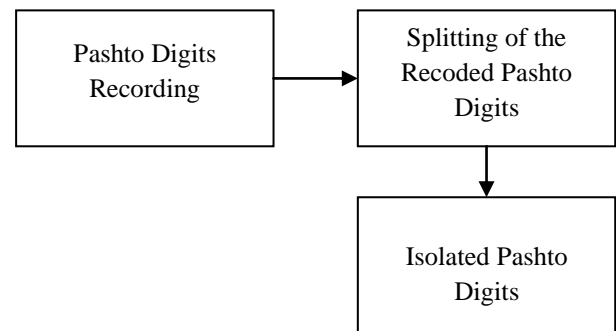


Fig 1: Pashto digits recording

For the recording of the Pashto digits, we have used the Sony PCM-M 10 Linear Recorder. Speaker who could speak the Yousafzai dialect fluently and perfectly were selected for recording of the Pashto digits. The recording was conducted in a noise free environment. The library of the university was used for recording. Before starting the recording of the Pashto digits, a few digits have been recorded for testing of the recorder and to check the sensitivity level of the recorder and to fix the signal to noise ratio. After the adjustment of sensitivity level as well as signal to noise ratio, the recording of Pashto digits was conducted. The speaker has spoken

Pashto digits from Sefer (0) to Sul (100) after that directly spoke Zar (1000). After the recording, the recorded Pashto digits file are split into isolated Pashto digits by using adobe edition ver 1.0 and saved the resulting isolated Pashto digits in .wav format having a sampling rate of 16 kHz, channels mono and 16 bit resolution. By splitting the recorded Pashto digits we have obtained isolated digits from Sefer (0) to naha (9) then Las (10), Shul (20), Dairsh (30), Celwaikht (40), Panzoos (50), Shpeta (60), Awaya (70), Atya (80), Nawi (90), Sul (100), and Zar (1000). The speech database for the Pashto digits synthesis system consists of these few isolated Pashto digits.

3.2 Pashto Digits Synthesis

The Pashto digits synthesis is illustrated in Figure 2.

3.2.1 Text Processing

The text of the input number is given by the user in the string format. The number string is then separated into digits. Based on the digit value and its position, sampled data of respective speech is the selected such as for (33), it will be read as Dray Dairsh.

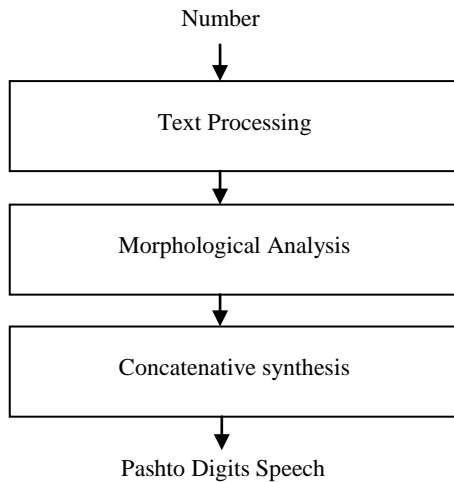


Fig 2: Pashto Digits Synthesis

3.2.2 Morphological Analysis

Pashto digits can be explained in terms of a morphological dictionary, which gives a list of 0 to 9 digits, the matlab result is shown below.

```

lex =
'1'
'2'
'3'
'4'
'5'
'6'
'7'
'8'
'9'
'0'
  
```

3.2.3 Pashto Digits Concatenative Synthesis

In concatenative synthesis we concatenate the isolated Pashto digits from the speech database, which have been created from a few isolated Pashto digits that contains Sefer (0) to naha (9) then Las (10), Shul (20), Dairsh (30), Celwaikht (40), Panzoos (50), Shpeta (60), Awaya (70), Atya (80), Nawi (90),

Sul (100), and Zar (1000). Through concatenation of these few isolated Pashto digits we produce Pashto digits from Yaw-las (11) to Yaw-zar (1000).

4. RESULTS AND DISCUSSION

Figure 3 shows the plot of the signal for the Pashto digit Celour (four), and Figure 4 shows the plot of Dairsh (thirty). These two digits i.e. Celour (four) and Dairsh (thirty) are present in the speech database but Celour Dairsh (thirsh four) is not present in the speech database. By concatenating Celour (4) and Dairsh (thirty), Celour Dairsh (34) is produce and similarly other digits too from Yaw Las (11) upto Yaw Zar (1000). For comparing the plot of synthesized concatenative Pashto digits Celour Dairsh is shown in Figure 5 and the plot of recorded Celour Dairsh (34) shown in Figure 6.

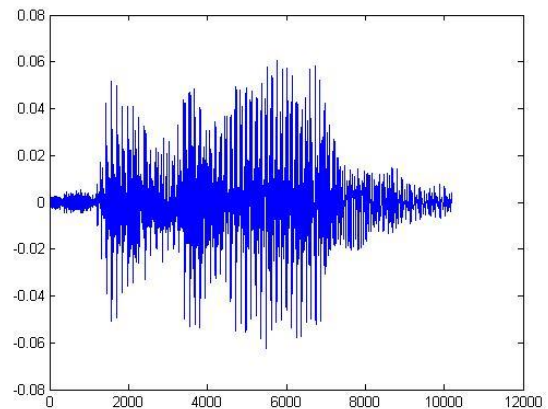


Fig 3: Speech Signal Plot for Pashto Digit Celour (four)

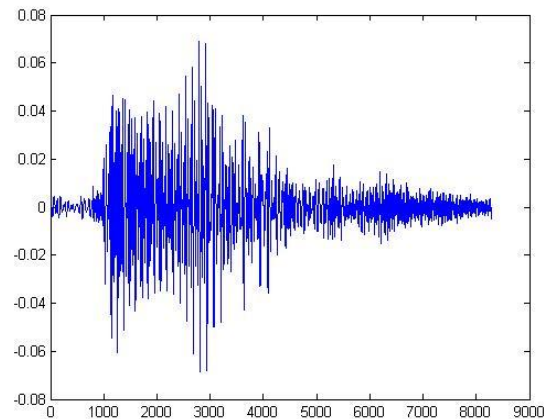


Fig 4: Speech Signal Plot for Pashto number Dairsh (thirty)

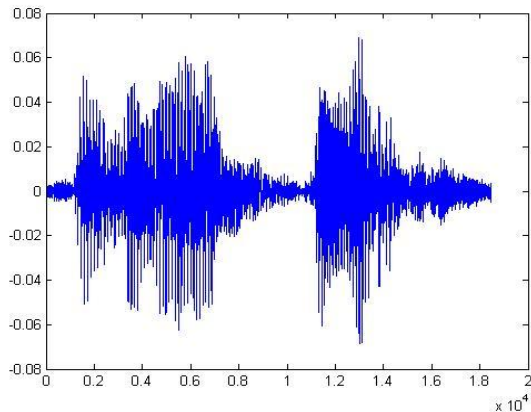


Fig 5: Speech Signal Plot for Concatenated Pashto number Celour Dairsh (thirty four)

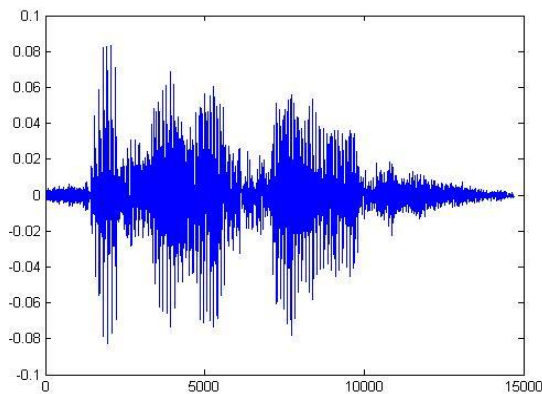


Fig 6: Speech Signal Plot for Recorded Pashto number Celour Dairsh (thirty four)

Synthetically generated Celour Dairsh (34) by the concatenative technique sound naturally just like to the human speech and comparing its plot with that of recorded Celour Dairsh (34), both looks similar.

Similarly the speech for the other Pashto digits and numbers is synthesized by concatenating those few isolated Pashto digits present in the speech database and studied both in terms of intelligibility in term of comparing their plots with those of the natural human speech.

It was found that the synthetic speech produced through the concatenative approach was similar to that of the natural speech both in intelligibility and the resultant signal, however the plot for the synthesized number and recorded number differ slightly at the point of concatenation.

5. CONCLUSION

The research presented in this paper is a first work on the Pashto Speech Synthesis and on Concatenative Pashto Digits Synthesis. The Pashto numbers from sefer (0) to Yaw-zar

(1000) are produced by concatenating isolated digits from a limited isolated Pashto digits database.

6. FUTURE WORK

In this research work the Concatenative based Pashto digits synthesizer has been presented. In future work, the problem of discontinuity of the synthesized number will be considered. The Pashto digits will also be synthesized through formant synthesis technique. The methods will also be extended to the Pashto continuous speech synthesis in the future besides the application of other techniques, such statistical techniques.

7. REFERENCES

- [1] Shriberg, E. (2005), "Spontaneous speech: how people really talk and why engineers should care", proceedings of the INTERSPEECH conference, Lisbon, Portugal, pp. 1781-1784.
- [2] Hejazi, S.A. Kazemi, R. Ghaemmaghami, and S., Sharif (2009), "Isolated Persian digit recognition using a hybrid HMM-SVM", proceeding of the International Symposium on Intelligent Signal Processing and Communications Systems, (ISPACS 2008), Bangkok, Thailand pp.1-4.
- [3] Douglas, B. P. and Janet, M. B. (1992), "The Design for the Wall Street Journal-based CSR Database", Proceedings of the DARPA SLS Workshop, Association for Computational Linguistics Stroudsburg, PA, USA
- [4] Matan, O. Christopher, J. C. B. Cu, Y. L. and John, S. D. (1992), "Multi Digit Recognition using A Space Displacement Neural Network", proceedings of conference of NIPS, vol. 4, Denver, Colorado, USA, pp. 488-495.
- [5] Miciak, M. (2008), "Character Recognition Using Radon Transformation and Principal Component Analysis in Postal Applications", proceedings of the International Multiconference on Computer Science and Information Technology, IMCSIT 2008, Wisla, Poland, pp. 495 – 500.
- [6] Parthasarathy, S. and Coker, C. H. (1992), "Automatic estimation of articulatory parameters", Computer Speech and Language, 6: 37–75.
- [7] Allen, J. Hunnicutt, M. S. and Klatt, D. (1987), "From text to speech the MITalk system", MIT Press, Cambridge, Massachusetts.
- [8] Hunt, A. J. and Black, A. W. (1996), "Unit Selection in a Concatenative Speech Synthesis System Using a Large Speech Database", ATR Interpreting Telecommunications Research Labs. 2-2 Hikoridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan.
- [9] Herbert and Sloan, I. (2009), "A Grammar of Pashto a Descriptive Study of the Dialect of Kandahar, Afghanistan", Ishi Press Intl. pp. 210.
- [10] Paul, L. M. (ed.), (2009), "Ethnologue: Languages of the World", (16th edition), Dallas, Tex.: SIL International. Online version: <http://www.ethnologue.com/>.
- [11] Alotaibi, Y. A. Alghamdi, M. and Alotaiby, F. (2010), "Speech Recognition System of Arabic Digits based on A Telephony Arabic Corpus", proceedings of the ICISP 2010, Canada.