Implementation of Concatenation Technique for Low Resource Text-To-Speech System-based on Marathi Talking Calculator

Monica Mundada¹, Sangrampsing Kayte¹, Pradip K. Das²

¹Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
²Department of Computer Science and Engineering IIT Guwahati, Assam, INDIA
monicamundada5@gmail.com, bsangrampsing@gmail.com, pkdas@iitg.ac.in

Abstract
The indulgent acquaintance of mathematical basic concepts creates the pavement for numerous opportunities in life for every individual, including visually impaired people. The use of assertive technology for the disabled section of the society makes them more independent and avoid barriers in the field of education and employment. This research is focused to design an Android-based application i.e. talking Calculator for low resource based Marathi native language. The novelty of this work is to develop both, the application and the Marathi number corpus. Marathi is an Indo-Aryan language spoken by approximately 69.99 million speakers in India, which is the third widely spoken language after Bengali and Telugu but as they lack in linguistic resources, e.g. grammars, POS taggers, corpora, it falls into the category of low resource languages. The front end part of the application depicts the screen of a basic calculator with numerals displayed in Marathi. During runtime, each number is spoken as the specific key is pressed. It also speaks out the operation which is intended to be performed. The concatenation synthesis technique is applied to speak out the value of decimal places in the output number. The result is spoken out with proper place value of a digit in Marathi. The performance of the system is measured to the accuracy rate of 95.5%. The average run time complexity of the application is also calculated which is noted down to 2.64 sec. The feedback and review of the application is also taken from real end-user i.e. blind people.

Index Terms: Visually challenged, Assertive Technologies, text-to-speech, Digital Signal Processing, Concatenative Synthesis, Unit Selection, Di-phone Synthesis.

1. Introduction
The invention of new technologies makes life easier for humans, but for disabled people it is, a blessing. The learning and understanding of Mathematical concepts cultivates thoughts and reasoning skills. The study of the subject is noteworthy in the human civilization and assigned importance from primary to secondary and higher-secondary level of education [1]. Mathematics teaches the systematic way of thinking with help of numerical and spatial aspects of the objects. But the learning of mathematics is fundamentally different than reading and writing other subjects and languages [2]. In traditional time the learning of Mathematics has been unapproachable to visually impaired and blind students because it is enriched with visually defined concepts and information. As per the World Health Organization (WHO) [3] 285 million people are considered to be visually impaired world-wide: Out of which 39 million people are blind and 246 million have low vision [4]. Calculators are the tools which are used to solve simple and complex problems faced by students, teachers, academicians and people working at home, school and industry. But the scenario is different for visually impaired people with these normal calculators [2].

Marathi is a low resource language which trails in the online corpus, Part of Speech Tagging (POS) and building of lexicons [6]. The primary aim of this research article is to survey alternatives which can advance the computational facilities of the visually impaired related to Marathi language. A text-to-speech (TTS) system converts normal language text, into speech [7]. The aim of speech synthesis is to project the design of a machine having an understanding and natural sounding voice for communication [8]. Speech synthesis systems leads with the conversion of the input text into its corresponding linguistic or phonetic representations and then synthesize the sounds corresponding to those representations [9]. With the input being a plain text, the generated phonetic representations need to be augmented with information about the intonation and rhythm that the synthesized speech should have. Here for ease, the speech corpus developed consist of only numbers [10]. Using the concatenation synthesis technique in Android platform, we have developed a low resource based Marathi Talking calculator [11]. So this Marathi talking calculator speaks out each number key as it pressed. It also speaks the operation to be executed and the result in Marathi with the correct digit place value. It also has two important key functions i.e. clear and go back along with voice feedback. So this gives perfect indication about the overall procedure of the application. The synthesized voice is accepted and correct, which is verified with the subjective Mean Opinion Score (MOS) test [12]. The application is also verified and feedback is applied with real blind students. This research article focuses to perform the basic calculation in Marathi language with important concept of voice. This will help rural illiterate people and also visually impaired people to
derive the basic mathematical functions. The average run time complexity of the application is calculated as 2.64 sec.

2. Database Corpus for Numbers
Most Indian languages are phonetic in nature [13]. Marathi consists of Devanagari script and spoken mainly in Maharashtra, India [14]. There are about 12 vowels and 32 consonants [15]. The peculiarity of this language is that it retains the pronunciation of some Sanskrit alphabets. The arrangement of phonemes is according to place and manner of articulation of the native script [16]. The database is designed for low resource Marathi language with total size of 121 recordings. This includes Marathi numerals from 1 to 100, place values of digits, and the four arithmetic operations. This also includes recording for clear and go back in Marathi. The database is developed at a professional recording studio, Silver Oak Advertisers, Aurangabad. The transcribing and labeling of each speech file phonetically is done using the Wave Surfer. This tool is enriched with features like playback of speech file with variable length delay between repetitions. It also adds phonemic, orthographic, tone and annotations to transcription in an interlinear format. It can also plot waveform, pitch plot, spectrum and various F1 vs. F2 displays [17]. Figure 1 shows the transcription files of speech input number “Ek”. Figure 2 depicts the Marathi numerals with pronunciation and arithematic operators.

2.1. Literature Survey
A lot of research and study is focused for developing synthesizers for Marathi language using various techniques. The Blizzard challenge 2015 [18] [19] covers the synthesis for various Indian languages including Marathi. Also research is going in institutes like IIT-M, C-DAC, TIFR, IITD, IT Devanagari script and spoken mainly in Maharashtra, India [14]. There are about 12 vowels and 32 consonants [15]. The peculiarity of this language is that it retains the pronunciation of some Sanskrit alphabets. The arrangement of phonemes is according to place and manner of articulation of the native script [16]. The database is designed for low resource Marathi language with total size of 121 recordings. This includes Marathi numerals from 1 to 100, place values of digits, and the four arithmetic operations. This also includes recording for clear and go back in Marathi. The database is developed at a professional recording studio, Silver Oak Advertisers, Aurangabad. The transcribing and labeling of each speech file phonetically is done using the Wave Surfer. This tool is enriched with features like playback of speech file with variable length delay between repetitions. It also adds phonemic, orthographic, tone and annotations to transcription in an interlinear format. It can also plot waveform, pitch plot, spectrum and various F1 vs. F2 displays [17]. Figure 1 shows the transcription files of speech input number “Ek”. Figure 2 depicts the Marathi numerals with pronunciation and arithematic operators.

3. Speech Synthesis for Marathi Numerals using Concatenation
Concatenative synthesis uses different length pre-recorded samples derived from natural speech [22]. This method is categorized into two types i.e. Unit and di-phone synthesis [23]. Di-phones are defined as speech units that initiate in the middle of the stable state of a phone and end in the middle of the following one. The number of di-phones is determined by the possible combinations of phonemes in a language [24]. In di-phone synthesis, only one example of each di-phone is contained in the speech database. The quality of the resulting speech is generally not as good as that from unit selection but more natural-sounding than the output of formant synthesizers [25]. Di-phone synthesis suffers from the robotic-sounding quality [26]. Unit selection synthesis uses unit as database for speech synthesis. This method requires large recording of database [27]. The unit selection technique gives naturalness due to the fact that it does not apply digital signal processing techniques to the recorded speech, which often make recorded speech sound less natural [28]. Figure 3 represents the block diagram of unit selection method.

Figure 1: Transcription file for speech input number “Ek”

Figure 2: Marathi numerals with pronunciation and arithematic operators

Figure 3: Block diagram of unit selection text- to-speech (TTS) synthesis. After [29].

4. Building of Marathi Talking Calculator
TTS system can be used to speak text messages from emails, SMS, web pages, news, articles, blogs, talking books and toys, games, man-machine communications, etc. Internet revolution made phones smart which became a fundamental part of life. There are number of speech driven applications available on smart phones [30]. Android operating system is gaining lots of attention as it provides access to various features of the phone like location sensor, TTS and many more. Android being open source delivers free development tools, which encourages people to use the android system [31]. These features of Android attracted developers to build systems which can be easily accessible to common people and also for low resource
languages. A talking calculator has a built-in speech synthesizer [32] that reads aloud each number, symbol or operation key a user presses in Marathi. Figure 4 describes the workflow for Android-based Marathi talking calculator. The main feature of the application is the ability to talk in Marathi. The attractive feature of APP is it contains two special buttons for clear and go back which again speaks out when pressed. This again gives the clear instruction to the user for performing the particular operation. The minimum system specification for implementing Marathi Talking calculator Android operating system with any version and 512MB RAM. The size of developed application is 7 MB. The front end of the application is designed like a basic calculator, the numerals are presented in Marathi along with the basic arithmetic operation i.e. Addition, Subtraction, Multiplication and Division. A real time image of the application is shown in Figure 5.

5. Experimental Analysis

The implementation and GUI of the application is designed using Android studio and Java script environment [33]. The system is trained with the Marathi recorded number corpus. The concatenation technique is applied for the generation of output speech. The run time analysis of the application is calculated as the time taken for each operation to be performed from the initial state to the final calculated and synthesized output. We have calculated the run time analysis for each of the basic operation i.e. Addition, Subtraction, Multiplication and Division. Total 50 readings are stored for the same. The average run time analysis for the overall application execution along with the result output speech synthesized in Marathi is 2.64 sec. Figure 6 depicts the run time complexity of input numbers, operation and the output result.

5.1. Performance Evaluation

The performance of the synthesized speech system is evaluated in terms of naturalness and intelligibility parameters [34]. There are various performance evaluation parameters i.e. subjective and objective, but here we opted the Mean Opinion Score (MOS) parameter for the evaluation of the synthesized speech. MOS is a test that has been used for decades in telephony networks to obtain the human users view of the quality of the output. It is calculated for the synthesized speech produced after the production from the application. The study also focused on the production from the application. It was counseled to the volunteers that they have to provide score ranging from 1 to 5 (Excellent 5, Very good 4, Good 3, Satisfactory 2 and Not understandable 1) for understandability. We have applied MOS test for all the four operations in the application i.e. addition, subtraction, multiplication and division using 100 subjects. The performance of the system is found to have an accuracy rate of 95.5% with the MOS test observations. The real-end user of this Marathi Talking calculator are blind people. So our intention of designing of this application is to benefit such sections of the society. The research study shows that visually impaired people have a good sense of hearing as compared to people with other disabilities. So we have also reviewed our application from such a cluster of people. We have tested our application with Blind students from classes 5-9 at Tarawati Bafna Blind School, Aurangabad. These students are well aware of handling mobiles and its operations. We have performed MOS tests with such group of 10 students for basic operations. So the accuracy of the application as per the visually impaired people is 93.25%.

6. Conclusion

This research study is intended for the development of low resource Marathi talking calculator which benefits visually impaired people to perform basic calculations in their native language. Using Concatenation method, the result of output speech is synthesized. The assertive devices will be beneficial if the synthesized voice sounds more natural. The study also focused on the development of this application is to benefit such sections of the society. The research study shows that visually impaired people have a good sense of hearing as compared to people with other disabilities. So we have also reviewed our application from such a cluster of people. We have tested our application with Blind students from classes 5-9 at Tarawati Bafna Blind School, Aurangabad. These students are well aware of handling mobiles and its operations. We have performed MOS tests with such group of 10 students for basic operations. So the accuracy of the application as per the visually impaired people is 93.25%.
feedback we received from blind students is to develop a multi-
lingual calculator.

7. Acknowledgements
We are thankful to Dr. Bharti Gawali Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada Universi-
ty, Aurangabad for providing the guidance for the research. We also show our gratitude for School authorities and students of Tarawati Bafna Blind School, Aurangabad for providing their valuable co-operation and support for performing MOS tests.

8. References
[2] M. Niss, “Mathematical competencies and the learning of math-
ematics: The danish kom project,” in 3rd Mediterranean confer-
ce on Mathematical education, Athens, Greece, 2003, pp. 115–
124.
[5] M. Post, C. Callison-Burch, and M. Osborne, “Constructing par-
allel corpora for six Indian languages via crowdsourcing,” in 7th
Workshop on Statistical Machine Translation, Montreal, Canada.
[6] B. B. Ali and F. Jarray, “Genetic approach for Arabic par-
415–420.
language telephone speech corpus,” in 2nd International Confer-
ce on Spoken Language Processing, Banff, Alberta, Canada, 1992.
ment of Marathi speech interface system,” in Advanced Comput-
speech synthesis,” in IEEE International Conference on Acous-
tics, Speech and Signal Processing, Seattle, WA, USA, vol. 1,
synthesis,” in 8th European Conference on Speech Communica-
[14] W. Bright, “The devanagari script,” The worlds writing systems,
quality systems: The role of perceptual contrast,” Language. JSTOR,
tool,” in 6th International Conference on Spoken Language Pro-
cessing, Beijing, China, 2000.
mura, “The NAIST text-to-speech system for the blurred chal-
lenge,” in Proc. Blizzard Challenge workshop, Dresden Germany,
2015, pp. 1–4.
“The IRISA Text-To-Speech system for the Blizzard challenge
2016,” in Blizzard Challenge 2016 workshop, Cupertino, United
[20] R. Kumar, S. Kishore, A. Gopalakrishna, R. Chitturi, S. Joshi,
S. Singh, and R. Sitaram, “Development of Indian language
speech databases for large vocabulary speech recognition sys-
tems,” in 10th International Conference Speech and Computer
(SPECOM), Patras, Greece, 2005, pp. 1–4.
[21] P. P. Shirishrimal, R. R. Deshmukh, and V. B. Waghmare, “In-
dian language speech database: a review,” International Journal
17–21, 2012.
[22] B. Duggan and M. Deegan, “Considerations in the usage of text
to speech (TTS) in the creation of natural sounding voice en-
abled web systems;” in 1st international symposium on Infor-
mation and communication technologies, Trinity college dublin,
[23] D. O’Shaughnessy, L. Barbeau, D. Bernardi, and D. Archambault,
“Diphone speech synthesis,” Speech Communication, Elsevier,
adult cross-language speech perception,” The Journal of the Acous-
own general purpose unit selection speech synthesizer,” Edin-
[26] S. Kayte, “A text to speech system for Marathi using English
Language,” International Journal of Engineering Science and
speech synthesis system using a large speech database,” in IEEE
International Conference on Acoustics, Speech, and Signal Pro-
Telephone speech corpus for research and development,” in IEEE
International Conference on Acoustics, Speech, and Signal Pro-
[29] M. Beutnegel, A. Conkie, J. Schroeter, Y. Stylianou, and
A. Syrdal, “The AT&T next-gen TTS system,” in The Journal of
healthcare applications for smartphones,” BMC medical infoma-
[31] J. Higginbotham and S. Jacobs, “The future of the android operat-
ing system for augmentative and alternative communication,”
Perspectives on Augmentative and Alternative Communication,
[32] L. M. G. Duhaney and D. C. Duhaney, “Assistive technology:
Meeting the needs of learners with disabilities,” International
[33] M. Palmieri, I. Singh, and A. Cicchetti, “Comparison of cross-
platform mobile development tools;” in 16th IEEE International
Conference on Intelligence in Next Generation Networks (ICIN),
[34] S. N. Kayte, M. Mundada, S. Gaikwad, and B. Gawali, “Per-
formance evaluation of speech synthesis techniques for English
Language,” in Proceedings of the International Congress on
253–262.