

Efficient Model for Numerical Text-To-Speech Synthesis System in Marathi, Hindi and English Languages

G. D. Ramteke and R. J. Ramteke

School of Computer Sciences, North Maharashtra University, Jalgaon (MS-India).
Email: shyam.ramteke@gmail.com, rakeshj.ramteke@gmail.com

Abstract—The paper proposes a numerical TTS-synthesis system in Marathi, Hindi and English languages. The system is in audible forms based on the sounds generated from several numeric units. A hybrid technique is newly launched for a numerical text-to-speech technology. The technique is divided into different phases. These numerical phases include pre-processing, numeric unit detection, numeric and speech unit matching; speech unit concatenation and speech generation. In order to enhance the syntactic generation of audible forms in three languages, two discipline tests were performed. The prosodic test has been obtained for evaluating on the statistical readings. Overall quality issue (OQI) test is a subjective test which is performed by various persons who are aware of three mentioned languages. On the basis of two distinct parameters of OQI test, all scores are positively provided. Initial parameter compromises with listening quality. The second parameter, awareness rate improves a level of the intelligibility. The ultimate satisfactory results of artificial sound generation in three unrelated languages were touched to humankind voice.

Index Terms—NTTS-System, Hybrid Speech Synthesis, Digital Signal Processing, Prosodic Analysis, Overall Quality Issue.

I. INTRODUCTION

The electronic synthesis of a natural voice is an intention for phonologists and engineers. Both experts desire to provide enormous helpful applications to man-machine interaction such as Talking Toys for Kids, Speaking Cat App on Mobile for enjoyment to youngsters, Talking Dictionary for improvement their vocabulary, Announcement of Bus-stop as well as Railway-station for warning to Passengers and so on [1]. Nowadays, research is very enthusiastically improving on TTS-Synthesis system. In India, Text-To-Speech (TTS) synthesis system plays a vital role of information technology for 18-standard languages e.g. Marathi, Hindi and so on. The domain would be extended in order to enhance the synthesized speech. Some of the present technologies for text-to-speech synthesis usually produce a speaking style machine by the voice which is to close human [2, 3]. The machine of speaking style involves two stages: number

processing and generation of the speech process. The number processing stage is extracted linguistic units using the rules. On the other hand, the signal generation stage is extracted acoustic units from speech using prosodic analysis. The present paper focuses on a new emergent approach for non-natural text-to-speech synthesizer. The approach is a hybrid technique. It has been borrowed from concatenative and rule-based speech synthesis techniques. The emergent technique is worked on the supersegmental which is one type of the prosody. The content of the prosody can be divided into three parts: length, pitch detection and intensity of the speech signals [4, 5]. The prosodic features are to apply on the numeric style. The form of numeric style is converted into the stream of phonemes using the rules of Indian numbers system [6]. Alternatively, the synthesizer can incorporate a model of the vocal tract. In an efficient model, the characteristics of person's voice are to produce a completely "synthetic/artificial" voice output [8]. The speech signals are normalized through the PRAAT tool which is used for speech analysis and developed by Paul Boersma and David Weenink, University of Amsterdam. With the help of natural-sound and intelligibility, the superiority of a text-to-speech synthesizer is considered. Natural-sound is like the accent of the person and intelligibility has caliber to be understood. The agenda of an intelligible text-to-speech is permitted to general public with visual impairments for listening. The task of reading makes it very easier to work on a personal computer [13, 24].

The major objective of a numeric text-to-speech engine is to execute the conversion of Indian number system into artificial voice form. Various voice forms would syntactically be generated in Marathi, Hindi and English languages.

The paper is summarized as follows; the next section deals with earlier TTS-work. Section-III describes Marathi, Hindi and English languages. Section-IV discusses on two components of NTS-system: Non-Natural Language Processing and Digital Signal Processing. Section-V proposes a model for the numerical text-to-speech in three unrelated languages. Section-VI gives the description of experimental work. The result and discussion explain in Section-VII. Finally, the present paper concludes on the efficient model.

II. EARLIER TTS-WORK

Text-To-Speech (TTS) system is a branch of the speech research field. A number of researchers have been working on the TTS-technology since last few decades. William A. Ainsworth^[1] developed the system for conversion English text into spoken form using a small amount of data. The performance of English TTS-system was achieved because most of the longer words in English are uttered on the basis of rules. The results of the system appeared to be encouraging.

Bhuvana Narasimhan et al.^[5] have proposed the Hindi TTS-model for schwa-deletion using concatenative technique. There were different issues of schwa pronunciation in Hindi: every schwa following the consonant is not produced within the word; the schwa deletion can be blocked for the presence of a morpheme boundary in multimorphemic words. Pamela Chaudhury et al.^[9] dealt with the model for Telugu conversion into spoken form. The results of Telugu TTS-system were good for intelligibility and fair for voice quality. But in 2012, Lakshmi Sahu et al.^[17] have presented the corpus-driven TTS-system based on concatenative synthesis method for two Indian languages: Hindi and Telugu. The system has been enriched with a couple of voices (male and female). The samples took from North India for Hindi languages and another from South India for Telugu language.

Soumya Priyadarsini Panda et al.^[22] have dealt with the conversion natural language text into a spoken waveform or artificial production of speech for Odia, Bengali and Hindi languages. The model was based on concatenative speech synthesis algorithm. It worked well for most of the characters in the three Indian languages. Saleh M. Abu-Saud^[26] has implemented the ILA-Talk system which was used for the multilingual TTS-system. It composed of the analysis phase which was categorized into two major cases: Case 1 was related to a number of training examples in the number words selected from the dictionary. Case 2 was concerned with the length of the number of characters in the training example.

In the previous work^[28], we have developed the TTS-work based on phonetic and voice processing in Marathi and Hindi language. The system was based on unit selection process using synthesis-by-rules for converting Devnagari phonetics into synthesized speech. The model was achieved to understand the phonetics without reading it. In this paper, the numerical TTS-work for Marathi, Hindi and English languages has been extended.

III. BASIC OF MARATHI, HINDI AND ENGLISH

India is the multilingual country in all over the world. Language is a way of communication among people. The communication way can be in the form of written or spoken [1]. The writing style and the spoken form are consistently unable to maintain. A couple of forms are pivotal for Indian students or novices to make a systematic study of the Marathi, Hindi and English sound

system. English is spoken over a large part of the world. It is spoken by educated people in India. In various regions of India, persons have shaken off the heavy features. Especially, English language can be spoken in spite of the regional languages [8]. The good speakers or bad speakers of English can be recognized in all over India. The terms ‘good’ and ‘bad’ refers to the level of tone of approximation to local English and standard Indian English. Also, the additional things are referred to qualities of clear, effective and intelligible speech [10]. Each language has an own pattern of speaking and writing. The writing shape of Hindi and Marathi languages uses Devnagari script. Marathi is a native language of Maharashtra state in India [16]. Hindi is a national language of India. Both languages include vowels, consonants and numbers. A couple of these languages are one of the 23-Indian constitutional languages [17]. In Maharashtra region, Marathi is an official language used in government and private sectors [20]. The paper is focused on Indian number system in three individual languages. In addition to English numerals for the cardinal form, Table-1 reveals few samples of two languages in an Indian script. The numbers are utilized in various patterns of writing except date format. The numbers of Marathi, Hindi or English language are disposed of the voice analysis and synthesis intention [21, 22].

Table 1. Cardinal Numerals in Indian Style for Marathi, Hindi and English Lanaguges

Numeral in English Figure	Devnagari Numeral in Symbolic Form	Pronunciation of the Number		
		Marathi	Hindi	English
0	०	शून्य (Shunya)	शून्य (Shunya)	Zero
1	१	एक (Ek)	एक (Ek)	One
2	२	दोन (Don)	दो (Do)	Two
3	३	तीन (Teen)	तीन (Teen)	Three
4	४	चार (Chaar)	चार (Chaar)	Four
5	५	पाच (Paach)	पाँच (Paanch)	Five
6	६	सहा (Saha)	छः (Chah)	Six
7	७	सात (Sat)	सात (Sat)	Seven
8	८	आठ (Aath)	आठ (Aath)	Eight
9	९	नऊ (Nau)	नौ (Nao)	Nine

IV. NUMERIC-TO-SPEECH SYSTEM

Numeric material emphasizes the efforts of very significant calculation. In addition to express the feelings, the importance of speech is to share the knowledge in the form of message. A numeric-to-speech system (NTS engine) is categorized into two discipline components. Starting component is Non-Natural Language Processing (NNLP) and another component is Digital Signal Processing (DSP) [9]. Fig. 1 exposes a functionality of numeric-to-speech synthesizer.

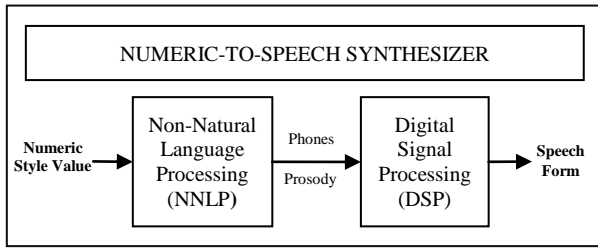


Fig.1. A General Functional Diagram of a Numeric-to-Speech Synthesis System

A. NNLP (Non-Natural Language Processing)

NNLP is a language i.e. spoken, written by the person for common communication. NNLP is a subfield of linguistic and artificial intelligence. It converts non-standard text or numeric style value for e.g. numbers, abbreviations currency and so on. Normally, the pronounceable words are called text normalization or pre-processing. The functionality of any TTS synthesis system is the conversion of input text into linguistic representation such as phrases, clauses and sentences [9]. It is usually called as text-to-phoneme or text-to-phonetic conversion. The process of conveying phonetic transcriptions to words is called as Phonetic transcriptions. The prosody information is an inner part of front-end which makes up the symbolic or linguistic representation with the help of phoneme [12].

B. DSP (Digital Signal Processing)

In the field of speech synthesis, DSP is a major component. DSP algorithms normally necessitate a mixture of mathematical expressions to be performed and constantly on a set of data samples. Speech signals are frequently translated from analog to digital, transformed back to analog type and manipulated digitally. All numeric units come from a prosodic way. Where a prosodic way is liable to find the best sequence of acoustic-parameters, there would be disposed of synthesized the phonemes or speech units [20]. The speech synthesizer translates the symbolic or linguistic representation into spoken form.

V. PROPOSED MODEL FOR NUMERICAL TEXT-TO-SPEECH SYNTHESIS SYSTEM IN THREE LANGUAGES

A text-to-speech technology is produced mankind syntactic voice form. The grapheme is to write in Roman or Devnagari script for Indian number system with various units. The phoneme performs a key role on the basis of Grapheme [7]. It produces the sound in Marathi, Hindi or English languages. In the process of sound generation, the speech signals are fetched from speech library and analyzed into prosody [12].

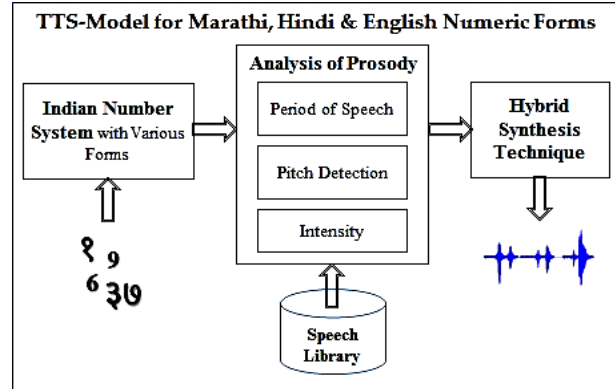


Fig.2. Architecture TTS-Model for Indian Number System

As depicted in Fig. 2, the hybrid technique was combined on the rule-based approach and the concatenation-based approach. In the rule-based approach, numerical patterns are disposed of the numerical rules for text form and the phonological rules for speech form. In the concatenation-based approach, all utterances of sound units have been used for concatenating and generation of sound signals. In order to generate the synthetic speech waveforms, this technique is used. The usage of a synthesis system is typically determining the specific approach [20].

A. Numeric Text Forms and Speech Selection based on Hybrid Technique

Numeric text form or speech form play a crucial role for NTTS system. NTTS system has two stages: text forms processing and concatenation of all given speech units. In the first stage, these rules were segmented into the units which recognized the units of numeric text. Text forms of numeric have utilized on three isolated languages [21]. Several dissimilar numeric phases for speech synthesizer includes numeric pre-processing; detection of the numeric unit; matching of numeric and speech unit; the concatenation of speech unit; and speech/waveform generation as revealed in Fig. 3.

For TTS-model, the following equations are assisted for parsing the Indian number system and symbolic notation.

$$script ::= devnagari | roman \quad (1)$$

$$devnagari ::= language \quad (2)$$

$$roman ::= language \quad (3)$$

$$language ::= Marathi | Hindi | English \quad (4)$$

$Marathi | Hindi ::= numeric forms$ (5)

$English ::= numeric forms$ (6)

$numeric forms ::= number$ (7)

$number ::= \{digit\} + [symbol]$ (8)

$digit ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9$ (9)

$symbol ::= , | ' | " | . | / -$ (10)

In order to analyze the numeric units, all equations are proposed in equation no. 1 to 10. The script is a part of Grapheme. Where there are different input forms of numeric, there is a barrier how to identify Devnagari or Roman script. Due to one of three languages is recognized at the stage of numeric processing level. For the system, three languages are used for implementation. Marathi or Hindi languages are based on Devnagari script and English is in Roman script. In event of recognizing, any language has command of writing. Natural and non-natural parts include in a specific language. The present work focuses on the non-natural region. Non-natural language can be a combination of the digits and the symbols. For the tags, the content of number is from 0 to 9. These numerals can be constructed in various units using the symbols. Several symbols are involved such as comma, single quote, double quote, decimal point or rupee for e.g. Rs. 1,20,000/-.

The step of the numeric analysis is rule-based. The rule-based synthesized speech relies on intelligibility. However, a level of naturalness is not always a goal of the rule-based synthesis system. Unit selection technique produces the synthetic speech as possible to close the natural sound. It is a sub-part of the concatenative synthesizer. Usually, it can be seen the voice characteristics of the person. The system does not need huge speech samples of different speakers. One speaker is sufficient either a female or a male. The extended version of the rule-based and the unit-selection synthesizer is a technique of the hybrid synthesis.

The required speech signals are sampled. The array of sampled data is converted back into the speech signal [8]. A process is to determine the pronunciation of a numeric form based on its spelling which is often called text-to-phoneme. Speech synthesis system for hybrid technique consists of two basic approaches. One is the dictionary-based approach and another is the rule-based approach [9]. The dictionary-based approach is the simplest approach for text-to-phoneme conversion which has advantages and disadvantages. The dictionary-based approach is quick, accurate and produces sound better in quality. It requires a large database as the dictionary [10].

On the other hand, the rule-based approach using the pronunciation of words is determined on the basis of some rules but the complexity grows as the system takes irregular inputs [11]. The last process is a speech generation before the concatenation of pre-recorded natural human speech. The difference between natural variations in speech signals and the nature of the automated techniques are to be naturalness and to be intelligible respectively. Both things rely on text analysis [12]. The next section describes a phase of the text processing in detail.

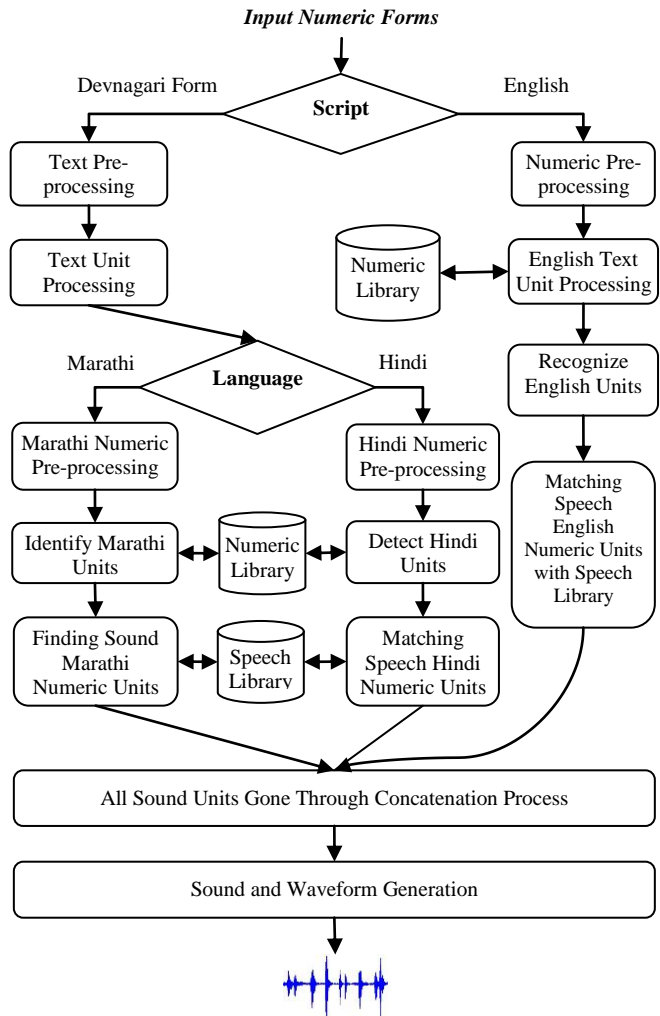


Fig.3. The Architecture of the Numeric TTS-System Algorithm for Marathi, Hindi and English Languages

a. Numerical Text Pre-Processing

Numerical text pre-processing is an initial step for NTTS system. Typically, text preprocessing is a complex task [1]. It includes several language dependent problems. The natural and non-natural language must be expanded into full words. The non-natural language text is to convert into the words [5]. This process is often called verbalization. Cardinal, ordinal and nominal parts of numerical are available. There would be able to focus on the cardinal form of numerical. Cardinal form is to write the digit into the word which represents the type of numeric units as '9' (Nine) in English, ' ९ ' in Marathi

and ‘९’ in Hindi. Some special numeric units are led in Table-2 with their formats: a digit; a variety of date formats; a pin-code; a telephone number; a mobile number; a measure and the rupees [20].

Table 2. Some Isolated Samples of Indian Numeric Units for Written Form in Marathi, Hindi and English Languages

Type of Numeric Units	Written Form in Devnagari and Roman Script		
	Marathi	Hindi	English
Digit	९	९	9
Date	१०-१०-२००६ किंवा १०/१०/२००६ किंवा १०- ऑक्टो-२००६ किंवा १०/ ऑक्टो/२००६ किंवा १०.१०.२००६	१०-१०-२००६ और १०/१०/२००६ और १०/ अक्टु /२००६ और १०.१०.२००६ और १०- अक्टु-२००६	10-10-2006 Or 10/10/2006 Or 10.10.2006 Or 10-Oct-2006 Or 10/Oct/2006
Pin Code	४२५००९	४२५००९	425001
Telephone No.	०२५७-२२५३४५७	०२५७-२२५३४५७	0257-2253457
Mobile No.	९९६०९५८९४८	९९६०९५८९४८	9960958948
Measure	२९९'५६"	२९९'५६"	299'56"
Rupees	₹ १२ किंवा १२ रुपये	₹ १२ और १२ रुपये	Rs. 12
	₹ ९८,७६,५४,३२,११०.९८ किंवा ₹ ९८,७६,५४,३२,११०.९८ रुपये	₹ ९८,७६,५४,३२,११०.९८ और ₹ ९८,७६,५४,३२,११०.९८ रुपये	Rs. 98765432110.98

Table 3. Group of Devnagari Form with ID

Sr. No.	Distinct Group of Devnagari and Roman Forms	Matchless ID
1	Devnagari Numbers (०-९,००,००,००,०००)	IN
2	English Numbers (0-1,00,00,00,00,000)	EN
3	Symbolic Notations	SN

Table 4. Parsed in Various Indian Number System

Sr. No.	Devnagari Numbers	Numbers in English Form	Unmatched ID	
			Mar or Hin	Eng
1	०	0	IN_0	EN_0
2	१	1	IN_1	EN_1
3	२	2	IN_2	EN_2
4	३	3	IN_3	EN_3
5	४	4	IN_4	EN_4
6	५	5	IN_5	EN_5
7	६	6	IN_6	EN_6
8	७	7	IN_7	EN_7
9	८	8	IN_8	EN_8
10	९	9	IN_9	EN_9
<i>Up to Crores</i>				
11	१,००,००,००,००,०००	1,00,00,00,00,000	IN_Crore	EN_Crore

Table 5. Parsed Isolated Symbolic Notation for Marathi, Hindi and English Languages

Sr. No.	Symbol	English Meaning	Unique ID	
			Mar or Hin	Eng
1	‘	Single Quote	SN_IN_SQ	SN_EN_SQ
2	,	Comma	SN_IN_CM	SN_EN_CM
3	“	Double Quote	SN_IN_DQ	SN_EN_DQ
4	.	Decimal Point	SN_IN_DP	SN_EN_DP
5	-	Hyphen	SN_IN_HN	SN_EN_HN
6	/	For date format	SN_IN_DT	SN_EN_DT
7	/-	For rupee notations	SN_IN_Rs1	SN_EN_Rs1
8	Rs.		-	SN_EN_Rs2
9	₹		SN_IN_Rs2	-

The given text recognizes the token of numeric units [9]. The text units are used in three unrelated languages. All numeric units should be able to normalize for matching with the content of a numeric library [10]. The positive rules are stress-free to be qualitative for parsing the numeric forms and symbolic notations as well. The numeric forms would be gone through the process for analyzing the string of symbols [11]. The symbols must be in Devnagari form for Marathi and Hindi languages or Roman script for English language. Table-3 shows the distinct groups of Devnagari and Roman forms: Devnagari numbers (IN), English Numbers (EN) and Symbolic notations (SN). The range of numbers is from 0 to 1,00,00,00,00,000 for analyzing Indian number system. The number is utilized for tracing it with the help of three distinct languages [12].

A number of text samples for appearance may be seen in Table-4 for parsing in various Indian number systems. Isolated symbolic notations are revealed how to parse syntactically for text analysis as per Table-5. Abbreviations may be expanded into full words or pronounced letter-by-letter [20]. Some contextual problems are shown for example ₹ or Rs. can be rupee or rupees for English and rupaye in Marathi and Hindi. The number in figure form is converted into a word or string representing e.g. Rs. 7,985/- (Seven thousand nine hundred eighty-five rupees only) in English [24].

b. Recognition Speech Units

The input for numerical text processing will be a number. Its output would be corresponding value of the number [15]. The objective is to analyze the number and to find it. All speech signals are to be used from the speech corpus [17]. Later, they are to be concatenated. A couple of processes are classified into some bunch of speech units. During classification of corpus, each recorded phone is divided into certain things: individual phones of numbers and symbolic notations in three individual languages. An index of the speech units is shaped on the basis of different acoustic attributes: duration, fundamental frequency (pitch), energy of sound and neighboring phones. The targets of desired utterances are created by determining the best chain from the corpus

(unit selection) [22]. Unit selection provides the greatest naturalness, because there is a small amount of digital signal processing (DSP) which is applied to the recorded phones. Few systems used a signal processing during concatenation process to generate the waveform smoothly [14, 23]. The output from unit-selection system is frequently identical a real human voice and become a new idea of the TTS system. Orthographic transcription in all kinds of units is shown in Table-6.

Table 6. Orthographic Transcription in Marathi, Hindi and English Languages Using Indian Numeric Units

Type of Numeric Units	Orthographic Transcription in All Kinds of Units		
	Marathi	Hindi	English
Digit	नऊ	नौ	Nine
Date	दहा ऑक्टोबर दोन हजार सहा	दस अक्टुबर दो हजार छः	Ten October Twenty Six
Pin Code	चार दोन पाच शून्य शून्य एक	चार दो पाँच शून्य शून्य एक	Four Two Five Zero Zero One
Telephone No.	शून्य दोन पाच सात दोन दोन पाच तीन चार पाच सात	शून्य दो पाँच सात दो दो पाँच तीन चार पाँच सात	Zero Two Five Seven Two Two Five Three Four Five Seven
Mobile No.	नऊ नऊ सहा शून्य नऊ पाच आठ नऊ चार आठ	नौ नौ छः शून्य नौ पाँच आठ नौ चार आठ	Nine Nine Six Zero Nine Five Eight Nine Four Eight
Measure	दोन हजार नव्याण्णव नऊशे आणि फुट इंच	दो हजार नित्यानबे फुट और छप्पन इंच	Two-Hundred Ninety-Nine Feet and Fifty-Six Inches
Rupees	एक हजार नऊशे पंचाऐशी रुपये	एक हजार नौसौ पचासी रुपये	One Thousand Nine Hundred Eighty Five Rupee
	अठठयाण्णव अरब शहात्तर कोड चौपन्न लाख बत्तीस हजार एकशे दहा रुपये आणि अठठयाण्णव पैसे	इटठानबे अरब छिहत्तर कोड चौपन्न लाख बत्तीस हजार एकसौ दस रुपये और इटठानबे पैसे	Ninety-Eight Arab Seventy-Six Crores Fifty-Four Lakhs Thirty-two Thousand One-Hundred Ten Rupees and Ninety- Eight Coins

Ordinal is converting a number into the position of something in a list. For example 1 is "first" in English, 'पहला' in Hindi and 'पहिला' in Marathi. Nominal is a type of number. It is used only as the name or to identify a postal code as '425001' is 'four two five zero zero one' in English, 'चार दोन पाच शून्य शून्य एक' in Marathi and 'चार दो पाँच शून्य शून्य एक' in Hindi. There are many numbers in Marathi which are pronounced differently based on context like the date. A normal programming is a challenge to translate a number into full words like '1985' becomes 'one thousand nine hundred eighty-five' in English. However, many different contexts occur the numbers; when a year or part of an address, '1985' should be read as 'nineteen eighty-five' in English, 'एकहजार नवशे पंच्याशी' in Marathi, 'एकहजार नौसौ पचास्सी' in Hindi; when part of any security number, as 'one nine eight five' in English, 'एकनऊ आठ पाच' in Marathi, 'एकनौ आठ पाँच' in Hindi. A TTS synthesis system can frequently gather how to expand a number based on surrounding numbers, punctuation and abbreviation. Sometimes, the system provides a line of attack to specify the framework if it is

ambiguous or unclear. The discussion is based on Table-6 of rupee unit how to speak of a given figure in three unrelated languages. For the instance Rs. 985.92 should be translated into a stream of phones using a grapheme-to-phoneme as 'nine hundred eighty-five and ninety-two coins' in English, 'नऊशे पंच्याशी रुपये आणि नव्यानऊ पैसे' in Marathi, 'नौसौ पचासी रुपये और नब्बे पैसे' in Hindi. The numerical parts of speech can be recognized but it needs speech library [3-6, 15-19]. The next section can be seen how to prepare speech library.

B. Preparation Speech Library and Enhanced Prosody

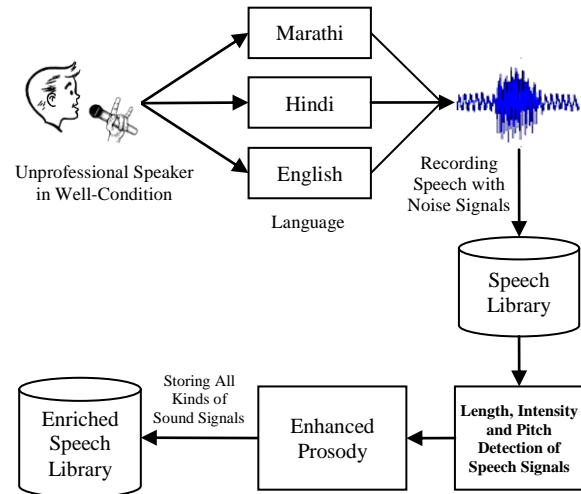


Fig.4. Main phases of Enhanced Prosody and a Speech Library Preparation in Three Languages

Finding correct sound is an uphill task for any speech synthesis system. The correct sound needs the collection of speech samples. This section focuses on how to prepare speech library and enhanced prosody. For preparation of speech storage, the various tools are available. The standardized database of numeric speech units in Marathi, Hindi or English language is unavailable [10, 20-23]. Thus, the speech corpus in three different languages is required. The miraculous thing is only one speaker needful for speech synthesizer. The attribute of speaker is non-mandatory, so it should be either unprofessional or professional in well-condition. The speaker is aware of three languages. The speech synthesized is able to make a speech unit which has been concatenated all pieces of recorded speech signals. All kinds of speech signals are stored in a speech library [12, 26]. The storage of phones is a huge amount of output range but may be lack clarity. The entire number permits to use for generating the high-quality speech output. In the earlier process of generating a voice, distinct sounds should be gone through the block of prosody [15]. Prosody includes length related to duration, stress related to the intensity or air pressure and pitch related to the frequency of vibration of the vocal cords as shown in Fig. 4 [27, 28]. The features of prosody are not essential for NTTTS, because the numerical part of each language is a non-natural language. On the basis of numerical part, the features of prosodic analysis can be extracted.

a. Length of the Speech Unit

The length of the signal speech unit is an actual duration of the sound signals. It is necessary two individual parameters: the sampled information and sampling frequency of sound signals for calculating the period of speech [20].

b. Modified Autocorrelation Pitch Detection Technique

Auto-correction technique is a popular method for detecting pitch in the domain of speech processing. In the section, the auto-correlation method has been modified for improving the pitch value of speech signals. The process of modified autocorrelation pitch detection technique is shown in Fig. 5.

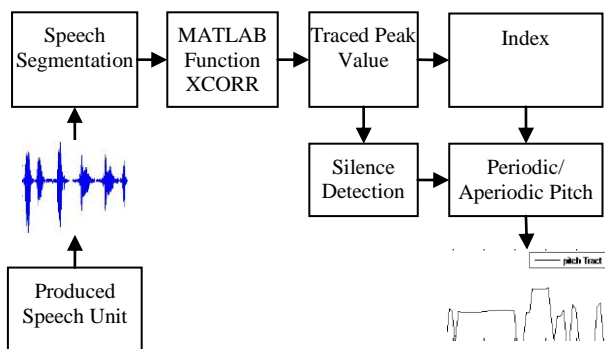


Fig.5. Process of Modified Autocorrelation Pitch Technique

The technique was applied on three dissimilar forms of speech which produces the speech units as time-domain layout. Speech segmentation is divided into frames. A number of frames are called as the window [16]. For speech signals, the mathematical definition is given as follows:

$$r(m) = \sum_{i=-\infty}^{\infty} x(i)x(i + m) \quad (11)$$

The equation is not required for implementation, because MATLAB tool is provided the function xcorr. The peak value of the frame is traced for recognizing the silence or index part of the speech [20]. The detection of the periodic or aperiodic pitch is based on the range of human speech. The range of sound for male and female or children is varied between 80-200 Hz and 150-350 Hz respectively.

c. Intensity of the Speech Unit

The intensity is a form of energy associated with the vibration of sound matter. Speech unit is a mechanical wave as in time-domain. Sound vibrations create as sound waves which move through mediums such air before reaching human ears. Usually, the intensity of sound signals is measured in decibels (dB) [4].

The following equations are given for estimating intensity of the speech units:

$$RMS = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}} \quad (12)$$

$$Intensity \text{ in dB} = 20 \log_{10} \frac{Peak \text{ value}}{RMS} \quad (13)$$

For speech frame i with N elements, a full form of RMS is root mean square, x is sampled information.

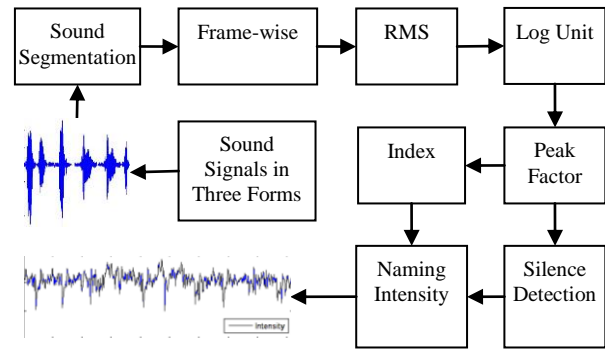


Fig.6. Architecture of Intensity for Each Speech Unit

Fig. 6 illustrates the process for designing of the intensity of each speech units. The initial step of the intensity is speech generation in three speech forms. The acoustic forms of a speech unit are revealed. The nature of the speech unit is continually changing. The segmentation of sound is a basic necessity for much speech processing. The speech signals are segmented into frames. The peak factor part of each frame is detected silence or index area of the sound signals. The segmented smaller units are often required in processing systems based on fixed size analysis frames. For speech analysis, the spectrogram is plotted a frequency domain against time with the intensity [23, 24]. It is an excellent method of visualizing sound structure. The identification of sound intensity is based on the human voice range in decibel. The human voice range for intensity is from 30 dB to 120 dB.

d. Enriched Speech Library

Table 7. Three Forms of Sounds for Three Languages

Three Forms of Speech	Language	Kinds of Phones	Number of Phones	Total Phonemes
Original	Marathi	Numbers	117	411
Noise		Symbolic Notation	20	
Quality				
Original	Hindi	Numbers	128	444
Noise		Symbolic Notation	20	
Quality				
Original	English	Numbers	57	231
Noise		Symbolic Notation	20	
Quality				
Total				1086

The speech library of Indian numbers with the visual figure in Marathi, Hindi and English is used [5]. Those speech samples are acquired through standard PRAAT tool. Its sampling frequency of speech units is 20 KHZ (22, 050 Hz). Two extra things of acoustic level are added such as one mono channel and wave file format for analyzed and synthesized [6]. One male or female speaker is enough for this model. Speaker should be aware of Marathi, Hindi and English languages for recording. All recorded speech signals are stored by a male speaker with the unwanted atmosphere [7]. All speech units are classified into three forms: original, noise and quality as shown in Table-7. The original sound is that speaker records the sound. Noise signals are added the unwanted or undesired or unexpected signals into original speech. The quality of speech is one type of clean sound for feeling pleasant while listening. All forms of speeches are enriched to the speech library [9-12].

C. Overall Quality Issue

The certain important factors are composed of awareness (intelligibility) and listening quality (naturalness). Naturalness explains how related the actual sound output such as to close human speech. Intelligibility is the ease with the output to be understood. Both natural & intelligible are ideal speech synthesizer for checking the quality [4].

Table 8. Value of Overall Quality Issue (OQI)

OOI Value	Awareness	Listening Quality
5	Much Better	Well-Quality
4	Better	Good-Quality
3	About the Same	Fair-Quality
2	Slightly Worse	Poor-Quality
1	Worse	Bad-Quality
0	Much Worse	Worse-Quality

Speech synthesis systems usually try to exploit the quality of synthesizer. In voice communication, quality usually dictates whether the experience is a good or bad on. A numerical method of expressing voice quality is called OQI (Overall Quality Issue). OQI is expressed in one number from 0 to 5 like a being the worse-quality and 5 the well-quality [19]. There are different criteria for awareness which is one type of understanding such as from much better to much worse for awareness as per Table-8. OQI is measured subjectively for quality sounds when a panel of listeners is involved [12]. The listener should be aware of Marathi, Hindi and English languages. The person will be able to give the judgmental scores based on OQI values after listening the syntactic generation of sound [13, 14]. For this experiment, various kinds of listeners who are well-known to three languages. They are too much familiar with the languages. There are 7 listeners who can judge and give the own score of OQI test. Seven listeners consist of 5-male and 2-female for

each language. All numeric units of Indian number system are 48 with their 7 numeric forms as shown Table-9. The total opinions are 1008 for numeric units in three individual languages [15, 16, 17]. Listeners has caliber that the model of speech generation is correct while receiving the text.

Table 9. Listeners and its Opinions for Indian Number Units

Language	Numeric Forms	Numeric Units	Listener (Male-ML/ Female-FL)	Total Opinions
Marathi	Date, Rupees, Measure	12*3=36 for 1 st Numeric Forms	ML – 5	240
			FL – 2	96
Hindi	Digit, Telephone No., PinCode, Mobile No.	4*3=12 for 2 nd Numeric Forms	ML – 5	240
			FL – 2	96
English			ML – 5	240
			FL – 2	96
Total		48	21	1008

VI. EXPERIMENTAL WORK

The numerical speech synthesizer is able to synthesize any number. The synthesized speech signal is sufficiently intelligible. Numbers can produce varying pronunciation depending on the way of various input text data. The speech synthesis systems are very difficult to evaluate because there is no standard of judgment and different speech samples uses from various organizations. Two different tests: Prosodic test and OQI test are conducted in three individual languages. Initially, the prosodic test has been computed of synthesized speech units. It included for duration, pitch detection and intensity of each speech unit for the evaluation of prosodic features.

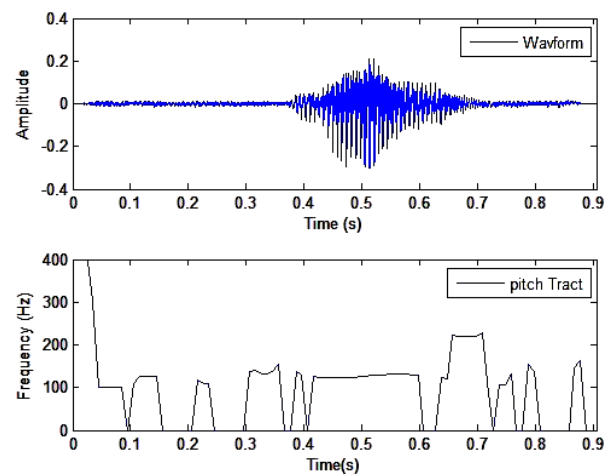


Fig.7. English Digit “9” Uttered by Male Voice a) Time Domain with Noise-Free form b) Pitch Tracking

Duration is the period of producing numerical speech units in seconds. The pitch detection and intensity of

three speech forms is demonstrated. A graphical demonstration is depicted the output in Fig. 7, Fig. 8 and Fig. 9 respectively. The Fig. 7 (a), Fig. 8 (a) and Fig. 8 (a) are generated in the time-domain waveform. The fundamental frequency has been determined by the modified autocorrelation pitch detection algorithm. The traced pitch of original, quality and noise signals on the basis of autocorrelation is exposed tracking in Figure 7 (b).

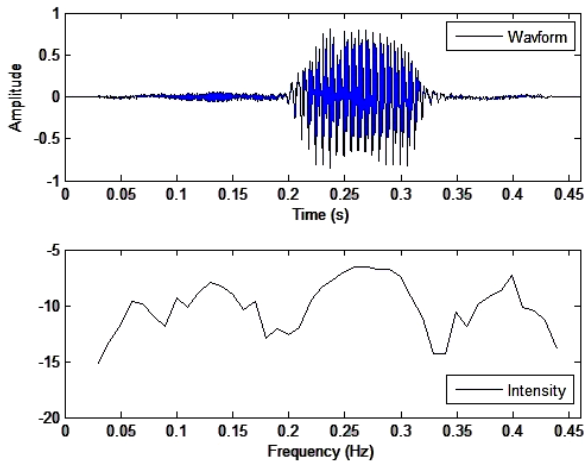


Fig.8. Speech Unit of Digit “7” in English Uttered by Male Voice a) Time Domain with Noise-Free form b) Intensity Tracking

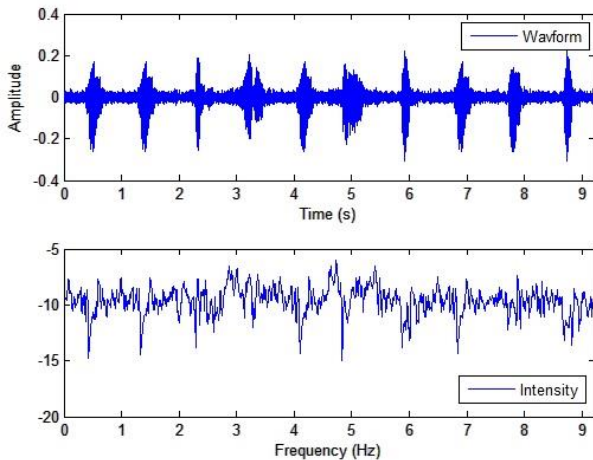


Fig.9. Speech Unit of Mobile Number “9960958948” in English Uttered by Male Voice a) Time Domain with Noise-Free form b) Intensity Tracking

There is considerable variation of pitch within voiced regions. The pitch detection and intensity of the each speech unit are applied on various spoken numeric units in three individual languages. The intensity of each speech units is depicted the output of intensity tracking as per Fig. 8 (b) and Fig. 9 (b). The prosodic and OQI tests would be discussed with the proposed TTS-system for a local language of Maharashtra, a national language of India and a world language.

VII. RESULT AND DISCUSSION

The intention of the result is to increase a level of intelligibility which would be for listening factor. The result of processing on each frame may be either a single number or a numeric unit. The simple time-domain processing technique should be capable of providing useful representations of such signal features as intensity, pitch. A number of schemes are estimated features of the speech units such as voiced/ unvoiced classification, pitch and intensity from the time-domain representations. Two bizarre tests are classified for the synthesized numeric unit. First is prosodic test which is used for detecting the pitch, intensity and length of each speech unit. Another is Overall Quality Issue (OQI) test which belongs on the human brain.

A. Prosodic test

The 0 to 9 in figure form have been used for calculating. Its categories were divided into 7 distinct units such as the digit, date, pin-code, telephone number, mobile number, measure and rupee in Table-10 and Table-11. Two isolated units out of 7-numeric units in three different languages were nominated for testing it. The numeric units of date and rupee had to be computed by modified autocorrelation pitch detection and intensity techniques.

Various speech units were collected from only one male voice and estimated by normal, quality and unclean speech signals. The synthetic voice was reflected to work out through pitch detection technique. The effect of each speech units has been publicized. The computed pitch detection can be revealed in mean and standard deviation (SD).

For instance, the digit (9 in English) is one of numeric units. According to Table-10, the statistical result of original speech form for digit unit is 176.99 Hz for mean and 147.61 Hz for SD. The mean and SD of noise speech form for same unit is 176.99 Hz and 173.99 Hz respectively. The results of quality form are 109.00 Hz for mean and 97.60 Hz for SD. As the results of pitch detection of all numeric units in Table-10, the statistical evaluation of quality speech form is to be satisfactory than other forms of speech. The figured pitch of seven numeric units is in the range of 90 Hz to 230 Hz. The results of estimated pitch detection of various speech forms are shown in Table-10. The overall range of mean and standard deviation for pitch detection is 105-240 Hz and 90-200 Hz respectively.

Similarly, for example, the date (१०-१०-२००६ in Hindi) is a type of various numeric units. In Table-11 for intensity tracking, the mean and SD of original speech signals are 70.29 dB and 10.37 dB respectively. The statistical result of noise speech signals is 70.32 dB for mean and 10.67 dB for SD. The results of noise-free speech signals are 67.35 dB for mean and 10.49 dB for SD. On the basis of results of intensity, the estimation of noise-free speech form for intensity tracking is to be acceptable than other forms of speech. The range of mean and SD for the intensity of all speech units is 8-14 dB and 65-72 dB respectively. The length of all speech units is

assisted to prosodic test which offered a clear measurement of how the voice has performed.

B. OQI test

OQI test is examined on synthesized sound on the basis of two dissimilar parameters: awareness and listening quality. For the test, few listeners are arranged. All listeners should be listened to 7 speech units. Seven speech units selected random form in three languages. Later, those candidates have given the rates which depend on the value of OQI. The average of OQI test has been computed. As shown in Table-12, for example, the pin code (४२५००१ in Marathi) is one of numeric unit. It calculated the score from 7 listeners. For original signals, the average of LQ (Listening Quality) and AR (Awareness Rate) was 4.25 (good-quality) and 4.75 (to close much better) respectively. The average for noise signals was 3.75 (to close good-quality) for LQ and 3.5

(between about the same to better) for AR For quality speech form, the average of LQ and AR was 4.25 (good-quality) and 4.25 (better) respectively. The average of all numeric units was between about the same and to close better for awareness and between fair-quality and good-quality for another parameter. The overall performance of the synthesized numeric unit for awareness was 3.6 for the original sound, 3.15 for noise sound and 3.92 for quality sound as per Table-12. As very few isolated numeric units in Marathi, Hindi and English were received the fewer score in OQI test, a number of numeric units have been achieved the high score. The nature of OQI test belonged on the two parameters. These parameters would be crucial elements of NTTS framework for increasing the level of intelligibility and to close human voice.

Table 10. Pitch Detection Algorithm for Indian Number System in Marathi, Hindi and English Languages

Numeric Units	Input Numeric Form	Language	Period of Speech Signals in Seconds	Original, Noise and Quality Forms of Speech Signals Using Standard Deviation (SD) and Mean (MN) in Hz					
				Original		Noise		Quality	
				MN	SD	MN	SD	MN	SD
Digit	9	English	0.90	176.99	147.61	176.99	173.79	109.00	97.60
	९	Hindi	0.67	186.17	184.62	225.81	197.30	121.83	115.68
		Marathi	0.44	166.12	152.12	174.57	158.63	162.65	138.18
Date	10-10-2006	English	3.98	190.79	172.44	190.79	174.25	135.05	134.45
	१०-१०-२००६	Hindi	6.14	180.67	173.73	198.49	191.88	150.09	136.44
		Marathi	5.07	184.60	180.36	213.13	192.56	146.28	127.98
Pin Code	425001	English	5.32	188.99	162.23	188.99	171.82	139.09	143.47
	४२५००१	Hindi	6.97	182.13	177.46	190.43	187.78	143.47	139.09
		Marathi	4.98	166.32	152.92	193.33	185.38	166.32	152.92
Telephone No.	0257-2253457	English	6.42	189.40	161.48	189.40	171.26	159.59	151.28
	०२५७-२२५३४५७	Hindi	6.29	181.76	173.86	195.60	191.30	135.05	134.45
		Marathi	9.10	167.04	148.69	196.01	195.19	167.04	148.69
Mobile No.	9960958948	English	4.61	179.80	150.18	179.80	149.35	140.63	137.95
	९९६०९५८९४८	Hindi	12.80	181.99	168.35	191.37	178.39	140.63	137.95
		Marathi	9.36	194.93	186.10	192.86	186.83	184.38	150.51
Measure	299' 56"	English	8.26	172.29	141.49	172.29	148.45	123.71	129.97
	२९९'५६"	Hindi	6.83	145.61	131.86	156.32	132.59	123.71	129.97
		Marathi	6.33	175.51	171.37	180.30	172.69	163.03	144.80
Rupee	12	English	0.85	169.16	151.45	169.16	151.63	150.50	148.29
	१२	Hindi	2.33	173.71	162.17	180.77	164.54	158.48	147.48
		Marathi	1.05	182.36	173.45	196.46	186.37	142.13	138.87
	98765432110.98	English	9.38	175.38	150.36	175.38	155.87	132.71	136.15
	९८.७६.५४.३२.११०.९८	Hindi	17.07	176.09	155.30	190.77	174.68	138.42	122.80
Marathi		15.36	176.46	171.20	176.68	172.97	149.16	137.63	

Table 11. Intensity of All Speech forms for Indian Number System in Marathi, Hindi and English Languages

Numeric Units	Input Numeric Form	Language	Length of Speech Signals in Seconds	Three Forms of Speech Signals in Hz					
				Original		Noise		Quality	
				Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Digit	9	English	0.90	69.62	9.54	69.82	10.03	67.89	9.68
	९	Hindi	0.67	70.85	9.74	70.18	10.76	65.73	10.23
		Marathi	0.44	68.94	10.50	71.92	9.98	67.39	8.02
Date	10-10-2006	English	3.98	68.45	9.19	70.57	11.28	67.35	10.49
	१०-१०-२००६	Hindi	6.14	70.29	10.37	70.32	10.67	66.56	11.18
		Marathi	5.07	70.13	10.43	71.27	11.19	67.66	11.41
Pin Code	425001	English	5.32	70.70	11.15	71.02	11.24	67.85	10.43
	४२५००१	Hindi	6.97	69.85	10.90	71.35	11.46	67.76	10.69
		Marathi	4.98	70.33	10.87	71.48	11.10	68.34	10.77
Telephone No.	0257-2253457	English	6.42	69.90	10.25	70.72	11.35	68.43	10.81
	०२५७-२२५३४५७	Hindi	6.29	70.36	10.54	71.26	10.98	66.97	10.76
		Marathi	9.10	70.33	10.87	71.48	11.10	67.87	10.95
Mobile No.	9960958948	English	4.61	70.21	10.42	70.74	11.10	68.67	10.64
	९९६०९५८९४८	Hindi	12.80	70.11	10.10	71.63	11.11	65.65	11.13
		Marathi	9.36	70.09	10.46	71.07	10.65	68.22	10.47
Measure	299' 56"	English	8.26	69.04	10.37	69.96	10.83	68.16	10.99
	२९९'५६"	Hindi	6.83	69.36	10.64	70.28	11.34	69.36	10.64
		Marathi	6.33	69.79	10.53	70.99	11.15	67.32	11.27
Rupee	12	English	0.85	69.49	9.66	70.78	13.18	67.81	9.10
	१२	Hindi	2.33	70.82	11.35	71.52	11.91	68.42	12.85
		Marathi	1.05	69.16	11.53	70.40	11.17	70.01	11.34
	98765432110.98	English	9.38	69.92	10.62	70.78	11.47	67.98	11.21
	९८.७६.५४.३२.१०.९८	Hindi	17.07	69.93	10.10	70.92	10.79	66.76	11.41
		Marathi	15.36	69.23	10.84	71.06	11.35	67.34	11.84

Table 12. Overall Quality Issue test for Indian Number System in Marathi, Hindi and English Languages

Numeric Units	Input Numeric Form	Language	Duration of Speech Signals in Seconds	Average Overall Quality of LQ (Listening Quality), AR (Awareness Rate) by 7 Listeners					
				Original		Noise		Quality	
				LQ	AR	LQ	AR	LQ	AR
Digit	9	English	0.90	2.25	2.5	3.25	3	2.5	2.75
	९	Hindi	0.67	2.5	2.75	3	3.25	2.75	2.25
		Marathi	0.44	3.5	3.25	2.75	3.5	4.5	4.5
Date	10-10-2006	English	3.98	4	4	3.25	3.25	4.25	3.75
	१०-१०-२००६	Hindi	6.14	3.75	3.5	3	2.5	4.25	4
		Marathi	5.07	4.25	4.25	3.75	3.75	4.25	4.25
Pin Code	425001	English	5.32	4	3.25	2.5	2.5	4.5	4.5
	४२५००१	Hindi	6.97	4	3.75	3.75	3.75	4.25	4
		Marathi	4.98	4.25	4.75	3.75	3.5	4.5	4.25
Telephone No.	0257-2253457	English	6.42	3.5	3.25	2.75	3	4.5	4.25
	०२५७-२२५३४५७	Hindi	6.29	3.5	3.25	3.25	3.25	4	4.25
		Marathi	9.10	4	4	3.25	3.5	4.25	4.25

Mobile No.	9960958948	English	4.61	4	3.75	2.75	2.5	4.5	4.25
	९९६०९५८९४८	Hindi	12.80	3.5	3.5	3.25	3.25	4	3.5
		Marathi	9.36	4	4.25	3.75	3.75	4	4.25
Measure	299' 56"	English	8.26	3.75	3.25	2.5	3	4	4
	२९९'५६"	Hindi	6.83	3	3.5	3.25	3	3.5	4
		Marathi	6.33	4	4.5	3.25	3.25	4	4
Rupee	12	English	0.85	4	3.25	2.75	2.75	4	3.75
	१२	Hindi	2.33	3.25	3.5	3.25	3.5	3.5	4.25
		Marathi	1.05	4	4.25	3	3	4.75	4.75
	98765432110.98	English	9.38	3	3.25	2.75	2.75	3.5	2.5
	९८७६५४३२११०.९८	Hindi	17.07	3.75	3.25	2.75	2.75	3.5	3.75
		Marathi	15.36	3.25	3.75	3.25	3.25	4.25	4
Average of Overall Quality Score				3.63	3.6	3.11	3.15	4	3.92

VIII. CONCLUSION

The efficient model of an Indian numerical TTS-system is proposed using the hybrid technique. The model has examined the performance of converting input numeric unit into synthesized spoken form. A variety of numeric units would be written in Devnagari or Roman script and to generate the waveform form. A text-to-speech (TTS) system has been demonstrated using few patterns of numerical. Prosodic and OQI tests have to be motivated on synthesized numeric units. The prosodic test was found out the duration, pitch and intensity of each numeric speech unit. The pitch has been computed between 105 and 240 Hz for mean and 90-200 Hz for standard deviation by the modified pitch detection algorithm. Likewise, the range of mean and standard deviation for the intensity of all speech units is 8-14 dB and 65-72 dB. The extracted feature was assisted to find out length of signals too. For OQI test, the implemented numeric synthesizer was tested on few listeners who are familiar with Marathi, Hindi and English languages. The OQI test for awareness parameter is between about the same and to close better. Another parameter of OQI is between fair-quality and good-quality for three languages. The application is very cost effective as compared to existing technology but not available in Marathi, Hindi and English languages. The achievement of a present model is understandable and to reach a humankind voice.

ACKNOWLEDGMENT

The work is sponsored by different organizations: Rajiv Gandhi Science and Technology Commission, North Maharashtra University Centre, Govt. of Maharashtra, India for funding the project (Code No. 7-II-DP/2014); a G. H. Raisonni Doctoral fellowship, North Maharashtra University, Jalgoan (MH-India); and SAP (DRS-I), UGC New Delhi, India.

REFERENCES

- [1] William A. Ainsworth, "A System for Converting English Text into Speech", *IEEE Transactions on Audio and*
- [2] Katsanobu Fushikida, Yukio Mitome and Yuji Inoue, "A Text To Speech Synthesizer for the Personal Computer", *IEEE Transactions on Consumer Electronics*, Vol. CE-28, No. 3., pp. 250-256, August 1982.
- [3] Fu-Chiang Chou, Chiu-Yu Tseng and Lin-Shan Lee, "A Set of Corpus-Based Text-to-Speech Synthesis Technologies for Mandarin Chinese", *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 7, pp. 481 - 494, 2002.
- [4] Marc Schroder, Jurgen Trouvain, "The German Text-to-Speech synthesis System MARY: A tool for Research Development and Teaching", *International Journal of Speech Technology*, 6, pp.365-377, 2003.
- [5] Bhuvana Narasimhan, Richard Sproat and George Kiraz, "Schwa-Deletion in Hindi Text-to-Speech Synthesis", *International Journal of Speech Technology*, Vol. 7, pp. 319-333, 2004.
- [6] Diemo Schwarz, "Concatenative sound synthesis: The early years", *Journal of New Music Research*, Vol. 35, No. 1, pp. 3-22, 2006.
- [7] Jerome R. Bellegarda, "Unit-Centric Feature Mapping for Inventory Pruning in Unit Selection Text-to-Speech Synthesis", *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 16, No. 01, PP. 74-82, Jan-2008.
- [8] S. D. Shirbahadurkar, D. S. Bormane, "Speech Synthesizer Using Concatenative Synthesis Strategy for Marathi language (Spoken in Maharashtra, India)", *International Journal of Recent Trends in Engineering*, Vol. 2, No. 4, pp. 80-82, 2009.
- [9] Pamela Chaudhury, Madhuri Rao, KVinod Kumar, "Symbol Based Concatenation Approach for Text to Speech System for Hindi using Vowel Classification Technique", *IEEE, 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC 2009)*, pp. 1082-1087, 2009.
- [10] Naim R. Tyson and Ila Nagar, "Prosodic rules for schwa-deletion in hindi text-to-speech synthesis", *International Journal Speech Technology*, Vol. 12, pp. 12-25, 2009.
- [11] Junichi Yamagishi and Keiichi Tokuda, "Robust Speaker-Adaptive HMM-Based Text-to-Speech Synthesis", *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 17, No. 6, pp. 1208-1230, August 2009.
- [12] Aimilios Chalamandaris, Sotiris Karabetos, Pirros Tsiakoulis and Spyros Raptis, "A Unit Selection Text-to-

- Speech Synthesis System Optimized For Use with Screen Readers”, *IEEE Transactions on Consumer Electronics*, Vol. 56, No. 3, 1890-1897, Aug. 2010.
- [13] Muhammad Masud Rashid, Md. Akter Hussian, M. Shahidur Rahman, “Text Normalization and Diphone Preparation for Bangla Speech Synthesis”, *Journal of Multimedia*, Vol. 5, No. 6, pp. 551-559, 2010.
- [14] Gerry Kennedy, “Benefits of Text to Speech Software”, *Australian Journal of Learning Disabilities*, Vol. 3, No. 3, pp. 31-34, 2010.
- [15] D. J. Ravi and Sudarshan Patilkulkarni, “A Novel Approach to Develop Speech Database for Kannada Text-to-Speech System”, *International Journal on Recent Trends in Engineering & Technology*, Vol. 05, No. 01, pp. 119-122, 2011.
- [16] S. S. Nimbhore, G. D. Ramteke and R. J. Ramteke, “Pitch Estimation of Devnagari Vowels using Cepstral and Autocorrelation Techniques for Original Speech Signal”, *International Journal of Computer Applications (0975-8887)*, Vol. 55, No. 17, pp. 38-43, October 2012.
- [17] Lakshmi Sahu and Avinash Dhole, “Hindi & Telugu Text-to-Speech Synthesis (TTS) and inter-language text Conversion”, *International Journal of Scientific and Research Publications*, Vol. 2, Issue 4, pp. 1-5, April 2012.
- [18] Catalin Ungurean, Dragos Burileanu and Mihai Surmei, “Statistically Augmented Preprocessing/ Normalization Module for a Romanian Text-to-Speech System”, *IEEE 7th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pp. 1-6, 16-19th Oct 2013.
- [19] Mukta Gahlawat, Amita Malik and Poonam Bansal, “Natural Speech Synthesizer for Blind Persons Using Hybrid Approach”, *5th Annual International Conference on Biologically Inspired Cognitive Architectures*, Vol. 41, pp. 83-88, 2014.
- [20] G. D. Ramteke and R. J. Ramteke, “Text-To-Speech Synthesis of Marathi Numerals”, *International Journal of Engineering and Technical Research (IJETR)*, Vol 7, Issue 7, pp. 360-367, 2015.
- [21] Sunil S. Nimbhore, Ghanshyam D. Ramteke and Rakesh J. Ramteke, “Implementation of English-Text to Marathi-Speech (ETMS) Synthesizer”, at *IOSR Journal of Computer Engineering (IOSR-JCE)*, Vol. 17, No. 1, pp. 34-43, 2015.
- [22] Soumya Priyadarsini Panda and Ajit Kumar Nayak, “An efficient model for text-to-speech synthesis in Indian languages”, *International Journal Speech Technology*, Vol. 18, No. 3, pp. 305-315, January 2015.
- [23] Maile Timm and Krista Uibu, “Development of Student Text Comprehension and Language Semantics in Primary School”, *Elsevier Proceeding Social and Behavioral Sciences*, Vol. 191, pp. 793-800, 2015.
- [24] Sumit Soman and B. K. Murthy, “Using Brain Computer Interface for Synthesized Speech Communication for the Physically Disabled”, *Elsevier International Conference on Information and Communication Technologies (ICICT 2014)*, 46, pp. 292-298, 2015.
- [25] G. D. Ramteke and R. J. Ramteke, “Text-To-Speech Synthesizer for English, Hindi and Marathi Spoken Signals”, at *British Journal of Applied Science & Technology*, ISSN: 2231-0843, Vol-15, Issue-3, pp. 1-16, Mar 2016.
- [26] Saleh M. Abu-Soud, “ILA Talk: A New Multilingual Text-to-Speech Synthesizer with Machine Learning”, *International Journal Speech Technology*, vol. 19, no.1, pp. 55-64, Mar-2016.
- [27] J. Sirisha Devi, Dr. Srinivas Yarramalle and Siva Prasad Nandyala, “Speaker Emotion Recognition Based on Speech Features and Classification Techniques”, *International Journal Image, Graphics and Signal Processing*, 7, pp. 61-77, Jun-2014.
- [28] G. D. Ramteke and R. J. Ramteke, “Text-To-Speech Synthesizer based on Phonetic and Voice Processing in Marathi and Hindi Language”, *Asian Journal of Mathematics and Computer Research*, ISSN: 2395-4205 (P), ISSN: 2395-4213 (O), vol. 14, issue 3, pp. 247-266, Dec-2016.

Authors' Profiles



G. D. Ramteke, male, is a research fellow at School of Computer Sciences in North Maharashtra University, Jalgaon. He received the MCA degree from North Maharashtra University, Jalgaon in 2010. He has been presently working toward the Ph.D. degree (Computer Science) since May, 2014. His research interests include Text-to-Speech Synthesis, Speech Enhancement, Speech Recognition and Speech Cryptography.



R. J. Ramteke, male, is a Professor and Head of Department of Information Technology at School of Computer Sciences in North Maharashtra University, Jalgaon. He received the degree of Ph. D. (Comp. Sci.), M. Tech. (Comp. Engg.), M.Sc. (Comp. Sci.) from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India. He awarded UGC teacher research fellowship during 2004-2006. He achieved the best student paper award at IEEE International Conference on Cybernetics and Intelligent System (CIS-2006), Bangkok, Thailand. His areas of research interests are Image Processing, Pattern Recognition, Document Image Analysis and so on. He has published more than 87 research papers in journals and conferences. His professional and research experiences are above 20 years and 12 years respectively.

How to cite this paper: G. D. Ramteke, R. J. Ramteke, "Efficient Model for Numerical Text-To-Speech Synthesis System in Marathi, Hindi and English Languages", *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, Vol.9, No.3, pp.1-13, 2017. DOI: 10.5815/ijigsp.2017.03.01