Implementation of Yoruba Text-To-Speech E-Learning System

Afolabi A.O & Wahab A.S

Ladoke Akintola University of Technology, Ogbomoso. Nigeria

ABSTRACT

Speech synthesis is the artificial procedure of human speech. A computer system for this purpose is called a speech synthesizer and can be implemented in software and hardware. A text-to-speech (TTS) system converts normal language text into speech. In the last few years, this technology has been widely available for several language for different platform ranging from personal computer to smart devices, but for Yoruba language which is been spoken by over 30 million people out of 150 million Nigerian populace have not such system therefore, there is need to develop such system for Yoruba language, Concatenative system based on the concatenation (stringing together) of segment of recorded part of speech. Concatenation method was used to develop this TTS system using syllables ad the basic unit of concatenation written in C# programming language. The Yoruba data syllable was matched with corresponding recorded sound syllable database so that there can be meaning sound pronunciation. The pronounced word where evaluated based on the perception of some native speakers of the language.

The acceptability of the text-to speech was high likewise the naturalness of the output, though the pronunciation was not perfectly smoothened but the system still provide the user with good capabilities of Yoruba word pronunciation. This work has created audio database for Yoruba consonant -vowels syllable, Yoruba alphabet and Text To Speech system was also developed.

1 INTRODUCTION

Computer assisted language learning (CALL) provides the basic technology for assisting learners to acquire important communication skill in a given language. Recent advances in Computer Mediated Communication (CMC), CACL and World Wide Web (www) facilitates the integration of these technologies in the development of powerful language education systems. We present a general framework. Underlying a pioneering research work to the best of our knowledge, this is such project on Yoruba language focused on the development of a web-based intelligent system for Yoruba language.

It is important that a CALL system possess the ability to adapt it behavior to the goals, tasks, interests, and specific needs of individual users or groups of users [1]. The central goal of modern approached to language learning and teaching goal-oriented learning and process approach to writing. Basically, language learning strategies seeks to enhance student autonomy and control over the learning process [2]. Since speech and writing are the basic media of human communication, a CALL system that exploits them would a better language learning environment. This paper provides an overview of ongoing research into the development of an intelligent web-based ICALL for Yoruba. Yoruba is one of the four major languages systems in Africa, other languages in these category include; Arabic, Hausa, and Swahili.

In Nigeria, Yoruba is one of the three major native languages (Hausa and Igbo) systems alongside English, which is the official language. In Nigeria, the home land of Yoruba lies between longitude 230 and 6 30 east of the meridian and latitude 6 and 9 north of the equator [3]. Yoruba is the native language of people in Lagos, Oyo, Osun, Ondo, Ekiti and Ogun states of Nigeria. It is also system in some part of Edo and Kogi states of Nigeria as well as Togo. East central part of republic of Benin and in Sierraleon(where it is called Aku).

Additionally, there are 25 letters in the Yoruba language alphabet. This is made up of 18 consonants (b d e f g gb h j k l m n p r ss t u w y) and seven vowels (a e e i o o u). there are five nasalized vowels in the language with three contrastive tone and 2 allotone. There are about 30 million people speaks Yoruba in the south western part of Nigeria. Student cite many reason for studying Yoruba including personal interest in at least African cultivate, research interest and fulfillment of foreign language requirements [3] African American students

often study Yoruba out of interest in their own heritage since many of the slaves brought to North America during 18th and 19th centuries came from Yoruba. Speaking area [4].

2. METHODOLOGY

The focus of this research work is to make use of E-learning text to speech to teach Yoruba language on line and enable people that are not able to learn and differentiate between Yoruba language synonyms, antonyms and homonyms (tonation).

Development of a Yoruba language text to speed is a powerful technique for solving problems between people of another ethnic group that want to associate with Yoruba people such as NYSC program that post people from Hausa, Igbo and Efik land to Yoruba land and vice versa.

2.1 SPEECH SYNTHESIS

Speech is the act of producing voice via variation of the air that is emitted by the articulate system [5] Whilst, speech synthesizer is the artificial production of human speech where a text-to-speech synthesizer should be able to automatically convert any text into signal carrying linguistic information before it is converted into an acoustic waveform using machine. Major purpose of TTS synthesis System is to transform a given linguistic representation, say a chain of phonetic symbols into artificial, machine-generated speech with information on phrasing, intonation and stress by means of an appropriate synthesis method.

The modern TTS system converts text into 'synthetic speech 'sound in a two-stage process [6] the first stages i.e. High Level Synthesis (HLS) reads the input text and generates a representation of how the text will be pronounced. The HLS stage is implemented using two modules, the first module, i.e. Text-analysis module, analyses the input text to identify its basic elements and the context in which they are used. The results of the text- analysis module is fed into the second module i.e. prosody module, which generate a linguistic description of how the text will be pronounced. It also integrates timing and rhyme information into the generated representation. All the processing involved in this stage are together called high level synthesis (HLS) and the technology for implementing them is dram from the domain of Natural Language Processing (NLP) and computational Linguistic [8].

The second stage, called Low Level Synthesis (LLS), takes the linguistic description outputted from the HLS stage input and generates the corresponding speech wave form. The ultimate goal of this stage is to generate speech signal which has as much as possible mimics the acoustic behavior of the speech produced by the native speaker reading the text aloud. There are three methods used to realize the LLS modules, these are Articulatory Synthesis Methods, Formats Synthesis Method and Concatenative Synthesis Method. All these methods will be fully discussed in the next section

A TTS system is composed of two parts a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviation into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion [7]. Phonetic transcription and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end often referred to as the synthesizer which converts the symbolic linguistic representation into sound.

2.2 SYNTHESIZER TECHNOLOGIES

An ideal speech synthesizer must be natural and intelligent, any speech synthesis system usually try to maximize both characteristics. The choice of technique generally depends on the language, platform used and the system itself. Although it is almost impossible to approximate a human natural speech, it is important to make sure that the synthesis speech is of sufficient quality so that an adequate and understandable reading culture can be achieved [8]. There are three major methods which can be used to generating synthetic speech waveforms; they are articulatory synthesis, formats synthesis and concatenative synthesis and lot of technique are available for determining the control parameters (duration, pitch, gain and fundamental frequency) for speech synthesis.

2.2.1 CONCATENATIVE SYNTHESIS

Concatenated synthesis is based on the concatenation (stringing together) of segments of recorded speech. In this method synthesis is done by using natural speech, this methodology has advantage because of it simplicity i.e. there is no mathematical

model involved [9]. Speech is produced out of natural human speech. Generally, concatenative synthesis produces the most natural sounding synthesized speech.

A recent trend in concanative synthesis approach is to use large databases of phonetically and prosodically varied speech, thereby minimizing the degradation caused by pitch and time scale modifications to match the target specification and to minimize the concatenation discontinuities at segment boundaries [9]. There are three main sub types of concatenative synthesis with some model that can enhance their performances they are unit selection synthesis, Diphone synthesis, and Domain-specific synthesis, this is the smallest synthesis method in producing the most natural sounding speech. Concatenative synthesizers which uses certain length of pre-recorded samples from speech database (or also known as lexicon), is the most commonly used technique nowadays since it produces acceptable quality of speech and perhaps is the simplest way in producing intelligible and natural sounding synthetic speech [8]. The main part for this type of synthesis is choosing a unit for concatenation purpose. This selection will affect the overall performance and quality of the output of speech where the longer length of a segmental unit implies a higher naturalness, less concatenation point and better control of co articulate on parameter.

However,the amount of required units and memories will increase dramatically as the number of units needed to be concatenated and stored will increase(Lemmety, 1999).While on the other hand,the selection of a shorter segmental unit might effectively overcome these problem yet the sample collecting and labeling procedure are more complex with higher distortion at the concatenation points. Nowadays, the units used in a system might consist of either syllables, demisyllables, phonemes, diphones or even triphones and word.

2.3. TEXT NORMALIZATION

The process of normalizing text is not straight forward; Texts are full of heteronyms, numbers and abbreviation that all require expansion into a phonetic representation. There are many spelling in English which are pronounced differently based on context. For example, “My latest project is to learn how to better project my voice” contains two pronunciations of “project”. Most text-to-speech (TTS) systems do not generate semantic representations of their input texts, as processes for doing so are not reliable, well understood or computationally effective. As a result, various heuristic technique are used to guess the proper way to disambiguate homographs, like examining neighbouring words and using statics about frequency of occurrence. Text normalization typically involves tokenization of input text into tokens.

Token can be standard and non-standard worlds (NSW). Standard words are those whose entry can be found in pronunciation dictionary while Non-Standard are words that their pronunciations are not found in the pronunciation dictionary. Text can tokenized based on whitespace[11]. Once tokenization is done, each token has to be identified for its corresponding meaning. After identifying the category, expansion of NSW can be accomplished by a combination of some step, such as expanding numbers, currency, dates and look-up tables for abbreviations, acronyms, e.t.c. work on text normalization in TTS systems mostly involves a set of rules namely duration, fundamental and intonation rules and these rules are language depended [10]. In homograph disambiguation, (word with the same spelling but with a different pronunciation), identification of token category involves a high degree of ambiguity, for example, ‘1974’ could refer to nineteen seventy four as a year, or one thousand nine hundred seventy four as a cardinal number. Disambiguation generally handles manually using some rules. However, such rules are very difficult to write maintain and adapt to new domains. Yoruba language does have much more homograph ambiguity but with these set of rules and being a tonal language one can easily cater for this. The digit “1997” might be spoken “nineteen ninety-seven” if someone is talking about the year, or “one thousand nine hundred and ninety seven” if someone is talking about the number. Likewise in Yoruba the word ‘Igba’ have up to four different meanings with different pronunciation but Yoruba being a toner language distinguishes each pronunciation from one another by categorizing them to different tone. There are three basic tones of different pitch levels in Yoruba namely, High, Mid and Low. In the writing system, the high and Low are marked with (‘) and (’) respectively, over the vowel. The mid tone is generally unmarked except where there might be ambiguity or confusion. In which it is marked with an over-bar sign (Olanike 2006).

3 LIMITATIONS OF DIFFERENT TECHNIQUES

It is quite certain that TTS technology will create new speech output applications which are associated with the improvement of speech quality. But the important part of this system is to make more natural conversion, which means reading a text like a human with stressing, intonations and durations. The technique use in analyzing the speech also matter has each technique has its own limitation and weakness and these are discussed section below. In formants synthesizer, the maximizing of naturalness is not the ultimate, the quality of the output speech is low since the sound are rather synthetic robotic sounding, therefore it is below the level of human hearing acceptability [8]

Articulatory Synthesizer gives the most satisfying natural sounding speech theoretically, but it however the most complex technique because of highest computational load due to the fact that human anatomy is very complex and flexible. In HMM, there is limitation in its ability to remove discontinuities' at the concatenation points at different phone boundary [12]. In concatenative synthesis, synthesis is done by using natural speech and this serves as an advantages because there is no mathematical model involvement, speech is produced out of natural human speech and it produces an intelligible and natural sounding synthetic speech better and preferable than others [9]

Summarily, the choice of the technique generally depends on the language, platform used and the purpose of the system itself. Although it is almost impossible to approximate a human natural speech, it is important to make sure that the synthesized speech is of sufficient quality. In this research concatenating method is chosen based on the issues addressed in the previous paragraph.

4. TEXT-TO-SPEECH IN E-LEARNING SYSTEM

E-learning services have evolved since computers were first used in education. There is a trend to move towards blended learning services, where computer-based activities are integrated with practical or classroom-based situation.

[13] and the OECD (2005) suggest that different types or forms of e-learning can be considered as a continuum, from no e-learning, i.e. no use of computers and /or the internet for teaching and learning, through classroom aids, such as making classroom lecture power point slides available to students through a course web site or learning management system to laptop programs, where students are required to bring laptops to class and used them as part of a face to face to hybrid learning, where classroom time is reduced but not eliminated, with more time devoted to online learning, through to fully online learning, which is a form of distance education. This classification is somewhat similar to that of the Sloan Commission report on the status of e-learning which refer to web enhanced, web supplemented and web dependent to reflect increasing intensify of technology use. In the Bates and Poole continuum, blended learning can cover classroom aids, laptop and hybrid learning while distributed learning can incorporate either hybrid or fully online learning.

It can be seen then that E-learning can describe a wide range of applications, and it is often by no means clear even in peer reviewed research publication which form of E-Learning is being discussed however. However, bates and Poole argue that when instructors say they are using e-learning, this most often refers to the use of technology as classroom.

5. DESIGN AND METHODOLOGY

5.1 TEXT TO SPEECH SYNTHESIZER MODELS

The components of a Test To Speech consist of two major components namely Natural Language Processing (NLP) and Digital Signal Processing (DSP). Natural language Processing component performs the task of decomposing a sentence to its sequence of the parts of speech such as noun, verb, adverb e.t.c. It consists of Text Analyzers which deduce the syllable to be use for specific word. The block of text will be inputted and it will be processed by passing through the text after this, block of text will be tokenized based on the tones of the syllables. Digital Signal Processing (DSP) is the computer analogy of the dynamically controlled the articulatory muscles and the vibratory frequency of the vocal so that the output signal matches the input signal matches the input requirements. It consist of the speech processing and sound processing. The speech processing lookup for syllables matching them to strings(concatenate) and smoothening them while sound processing (speech signals) process the audibly, this illustrated in the Figure 3. In this research, concatenative synthesis method was used, this was based on concatenation (stringing together)of segments of recorded speech and unit selection synthesis will be adopted. A database of recorded sound was created that contains vocal Yoruba syllables and synthesis of the test supplied was performed by speech synthesizer developed.

5.2 SYLLABLE IDENTIFICATION

NLP will be perform by synthesizer, it will brake every block of text to syllable and identify the vowel and the consonant vowel (V,CV) and recognizes the tone bearing vowel, it also recognize ẹ, ọ, ṣ consonants and differentiate them from the similar consonant. This process serves as normalization of the block of text. The algorithm and flow chart for this is shown in Table1 C# programming language was used for coding and was implemented on Microsoft visual basic studio environment.

TABLE 1 SYLLABLE IDENTIFICATION ALGORITHMS

STEP	DESCRIPTION
1:	Start
2:	Declare Word (n) character as array
3:	Supply Character into word
4:	Check if character is valid IF Yes GOTO STEP 5 IF No GO To STEP 3
5:	CHECK if Character is 1 or 2 If YES GOTO STEP 6 If NO GOTO STEP 7
6:	CHECK IF Character is Vowel or Consonant If YES GOTO STEP 9 IF NO GOTO STEP 3
7:	CHECK if Character is Vowel or Consonant If YES GOTO STEP 8 IF NO GOTO STEP 3
8:	CHECK FOR END of string If No GOTO STEP 3
9:	Compute Character
10:	STOP

5.3 DIGITAL SPEECH PRONUNCIATION

Digital Processing Signal aspect of this work was in two stages creation of sound database for Yoruba syllable and letter syllable database. A database was created for the recording of all the syllables in three tones of Yoruba language: High, Mid and Low tones. The recording facilities and ahead phone with Condenser Microphone (SONY ECM_44S Electrets) was connected to it. Finally the data was edited by sound editing software (Audacity) where noise and other unwanted pitch were removed. The recording was done by a female as the native speaker of the language.

Integrating the recorded syllable sound to match the block of supplied text so as to have correct pronunciation was achieved with C# programming language and implemented on Microsoft visual basic environment, the algorithm and flow chart were shown on Table 2 and Figure 3 respectively. The sound code for this project can be found on appendix A.

TABLE 2 SPEECH PRONUNCIATION ALGORITHMS

STEP	DESCRIPTION
1:	START
2:	SET COUNTER=1
3:	DECLARE VALUE for n

4: Check if Char[count]

5: If No GOTO STEP 8

6: BREAK Char is space

7:GOTO STEP 10

8: Check if the Char is space

If No GOTO STEP 10

9: BREAK SPACE //{Pronounce the word}

10: SET COUNTER= COUNTER+1

11: STOP

In this part of the world are mostly run on window environment that makes C# with Visual studio.NET preferred for this design.

6 YORUBA PHONOLOGY

The Yoruba alphabet consist of 25 letters which were derived from Latin characters.

The Yoruba language learner is fortunate for two reasons. First, with the exception of a few segments, the writing system closely matches the sounds system of the language. Secondly, with the exception of almost the same set of unique sound, the Yoruba segments in many other languages [14] .Note that consonant "gb" is a diagraph i.e a consonant written in two letters Yoruba alphabet is shown in Table 3, it consist of 25 characters.

TABLE 3: THE UPPER CASE AND LOWER CASE REPRESENTATION OF YORUBA ALPHABET

Aa Bb Dd Ee Ee Ff Gg GB gb Hh iI
Jj Kk Ll Mm Nn Oo Oọ Pp Rr Ss
ṢṢ Tt Uu Ww Yy

6.1 YORUBA CONSONANT

Yoruba consonant is made up of 18 alphabet and will be shown in the Table 3.5. This consonant letters were part of the Yoruba alphabet but 18 alphabet made up Yoruba consonant while the remaining letters were Yoruba vowel letters.

Orthography Representation of Yoruba consonant are shown in Table 4

Bb Dd Ff Gg GB gb Hh Jj Kk Nn Pp Rr Ss Ṣṣ Tt Ww Yy

6.2 SYLLABIC NASAL

There also exists in the language a syllabic nasal phoneme. This has six homorganic allophones that are phonologically consonants but share the characteristics of being syllabic and tone bearing with vowels in the language (/n/,/an/,/en/,/in/,/on/,/un/). They occur before other consonants in syllable junctions [16]. The syllabic nasal phonemes is here represented with an m. the six homorganic allophones (dependent on the type of consonant they occur before)of the syllabic nasal phoneme are:/m,/M/./n/,ñ/,/Ñ/,/n/, and/Ñm/ but this is not in the scope of this study.

6.3 SYLLABLE STRUCTURE

The function of syllable is to regulate the structure of complex segments. The syllable serves as a building block for higher-level phonological and morphological processes, Yoruba syllable is the smallest tone bearing unit. The three basic syllable types in Yoruba are V, CV and N. The first type of syllable involves only a single vowel and this is often the shape of lexical items such as pronouns. The second syllable type in Yoruba is a consonant and a vowel, this is the basic shape of simple verbs in the language. The third and final syllable type in Yoruba is the syllabic nasal. Due to the shape of the syllable types in Yoruba, there are no consonant-final words and therefore, there are no closed syllables in the language. All the three syllable types have either combinations of first and third type or combinations of first and second and it can be vowel only (Nucleus). Moreover, consonant clusters are not allowed to occur in Yoruba syllables, further work done on Yoruba syllable with the help of ONSET-RHYME theory shows that Yoruba syllable can be specify to five syllable configuration which are CV,

CVn, V,Vn and N [17] . The three syllable types are illustrated in the Table 3. These are Vowel (V), Consonant Vowel (CV) and Nasal Syllables respectively.

TABLE 5 YORUBA SYLLABLE TYPES

SYLLABLE	EXAMPLE MEANING
VOWEL (V)	a‘we’ ó ‘he’
CONSONANT VOWEL	rán “to sew’
(CV)	t à to sell’
NASAL VOWEL [N]	òron bó ‘lemon’
	dù n dú fried yam’

Yoruba is very strict in regards to its prohibition of closed syllables. In terms of word structure, nouns often begin with vowels and verbs with consonants. There are however no fixed rules in the language as to the number of possible syllables within a word.

6.4 THE IMPORTANT OF SYLLABLE STRUCTURE

The syllable structure is used in the development of the system (synthesizer) and it helps in following areas:-

- The structure helps in the recording of the possible syllables and stores them in the database.
- It will also help in the design of the parsing algorithms which is the rule use to derive the syllable of a given word.
- It also helps in the designing of a storage structure for easy and efficient retrieval of the audio files.
- The structure helps as a control to confirm the validity of a word with respect to its contextualizes.

SUPRA SEGMENTAL ELEMENTS.

Yoruba is a tonal language. It has three surface tones of different pitch levels. The syllable is the tone bearing unit in the language but orthographically, tones are marked on vowels and syllabic nasals [15] The tones and their orthographic representations are shown in Table 3.6

TABLE.6 YORUBA TONES AND THEIR ORTHOGRAPHIC REPRESENTATION

High		as Wálé ‘a male name’
Mid	unmarked	as in aago ‘watch’
Low		as in dòdò ‘fried plantation’

The mid tones can sometimes be marked with an over-bar in order to remove ambiguity or confusion. Lexically, Yoruba tones are significant because a change in tones will completely change the meaning of a word. Additionally, tone markings allow for improved Yoruba writing and reading. One way to consider the three level Yoruba tones is to think of the music notes to which they correlate. These correlations are shown in.

TABLE 7 YORUBA TONE REPRESENTATION AT DIFFERENT LEVELS

Tone	Musical notes	Orthography
HIGH as in j}	‘to be’	mi
MID as in jẹ	‘to eat’	re
LOW as in j]	‘to wonder around’	do

Identification of the syllables of a word with the right tone is important because it is use to determine the stress and intonation of that particular word. Therefore, Yoruba syllable has three tones for a single syllable and the total number of combine consonant vowel syllable (CV) in Yoruba is illustrated in the Table 6 where the numbers of CV combination were shown.

7. IMPLEMENTATION AND DISCUSSION

7.1 RECORDING OF YORUBA SYLLABLE

A database was created for the recorded syllables in the three tone of Yoruba language. The recording was done in sound proof environment with a PC that has voice recording facilities and ahead phone with condenser Microphone SONY ECM-44S Electrets connected toit. The data was edited by sound editing software Audacity where noise and

other unwanted pitch were removed. Figure 4.1 shows the snap shot of the created database for the recorded syllables. Furthermore database for Yoruba records was also created.

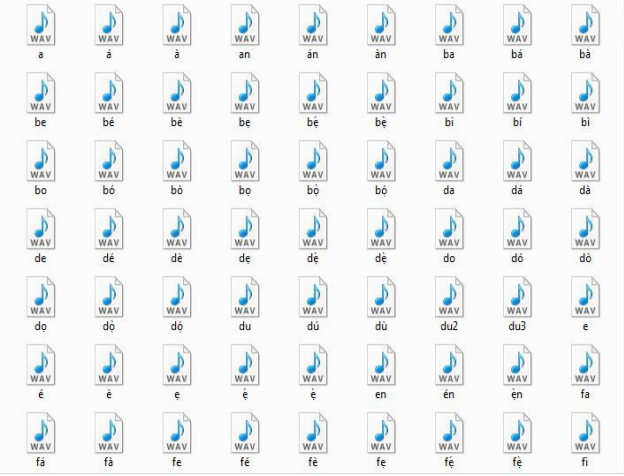

Figure1 :Database created from recorded Yoruba syllable

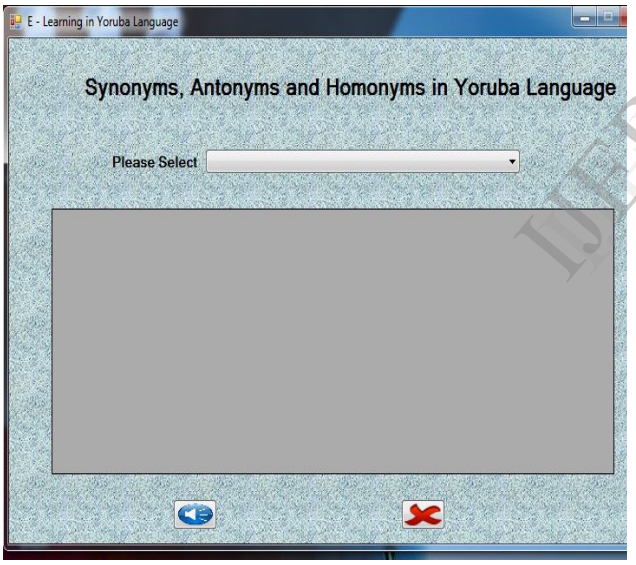

Figure 2 The Interface Of Synonyms, Antonyms and Homonyms in Yoruba Language

Figure 2 enable the user to select learning option which can either be: Start learning, for general users where user choses either synonyms, antonyms or homonyms in other to start learning and Input words for authorized users only.

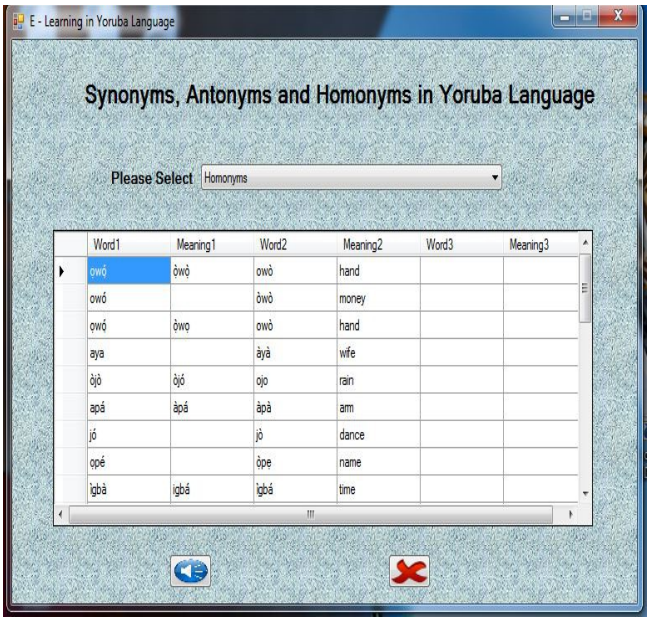

Figure 3

Figure 3 teaches learner the homonyms in Yoruba language and their corresponding meaning in English language to enable effective communication and user friendly interface.

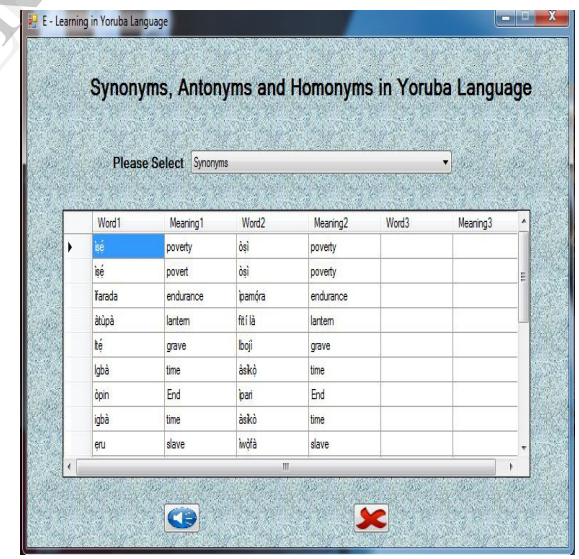

Figure.4

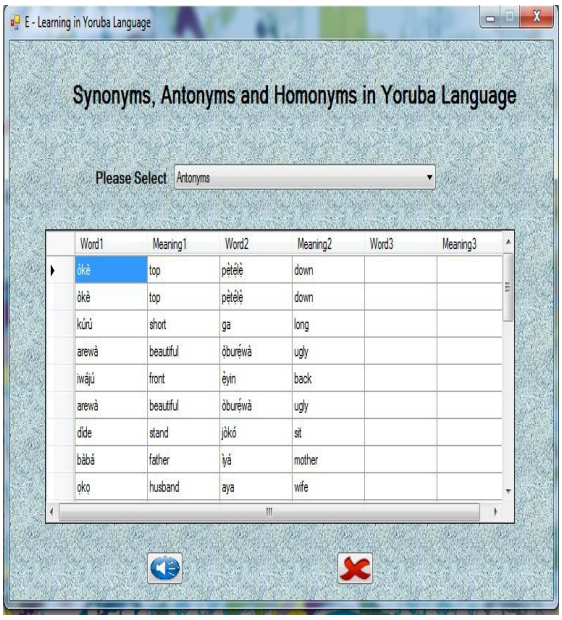

Figure 5

Both figure 4 and figure 5. are the synonyms and antonyms database that were assembled using some of the Yoruba language vocabulary with the aid of concatenation of both consonants and vowels inculcating intonation signs to differentiate each words from each other.

Figure 6

The figure above shows the INPUT interface which is solely used by the authorized user to input words in to the database depending on the scope of the work and the range of its use. However, the case study of our project is on Homonyms, Antonyms and Synonyms which have being already discussed above.

5.1 CONCLUSION

With this project, Yoruba TTS E-learning system was developed and the aim and objectives were achieved. Yoruba language structure was likewise the text to phonetics analysis were carried out which made it possible to develop easy and fascinating E-learning interface environment.

Although, the pronunciation was not perfectly smoothened but the system still provide the learners with the capability of pronouncing some of the concatenated words such as the Antonyms, Homonyms and Synonyms.

REFERENCES

[1] Bozkurt B,.Dutoit T.,(2001): "Animplementationand Evaluation of two Diphone Based Synthesiszers for Turkies" Procceeding of 4th ISCA Tutorial and Research work shop on speech synthesis. Pp.247-250.

[2] Alan W.,(2002): "Perfect synthesis for all the people of the time".keynote address at the IEEE Workshop on Text to speech on 30th of September, 2002. View the demonstration at :http://www.cs.cmu.edu/-awb/papers/IEEE2002/allthetime.html.

[3] Conkie A., Beutnagel M., SyrdalA.K.,BrownP.e.,(2000): "Preselection of candidate units in a UnitSelecction-Based Text-to speech synthesis system" ICSLP2000.

[4] Delanọ, Oloye Isaac (1958).Atúmọ̀ ede Yoruba [short dictionary and grammar of the Yoruba language]. London: OxfordUniversity Press.

[5] Dutoit T. and Leich H., (1993): "Text-to-speech synthesis based on a MBE. Re-synthesis of segments Database" Speech commtn vol., 13 pp. 435-440.

[6] Allen J.,Hunnicut S., Klatt D., (1987): " From Text to Speech, the MITTALK system". Cambrige University Press, USA.

[7] Bernd Mobius, Richard Sproat Jan P.H., Van Santen, Joseph P. Olive (1997): "The bell Lab German Text-to-speech system over view". Bell Labs-Lucent Technologies 600 Mountain Avenue, Murrey Hills NJ07974 USA.

[8] Huang, A.J. and Black, A.W. (1996):"Unit selection in a concatinative speech synthesis system using a large speech Database" CASSP, pp 373-376.

[9] Aida-Zade K.R. and sharifova A.M., (2010): "The main principle of text-to-speech synthesis system".international journal of information and communication Engineering.

[10] John K. and Alan W.B., (2003): "CMU ARCTIC database for speech synthesis".CMU-LTI vol. 3 pp.177. Language Technologies Institute, School of computer science, Cranegie Mellon University

[11] Husni M.,Moustafa E., Mansour A.,(2001): "Technique for high quality Arabic Speech synthesis". An international journal on information Sciences. June 28, 2001. Views the demonstration at www.elseview.com/locate/ins

[12] Ken Fujisawa and Nick Campbell (1998): "Prosody _Based unit selection for Japanese Speech synthesis" The third ESCA/COCOSDA workshop (ETRW) on speech synthesis Jenolaon cave House, Blue Mountain NSW, Australia,

[12] Nicolas, pierre L., Mathiew A., "Towards improved HMM-Based Speech Synthesis using high level syntactical features"http://alpage.iniria.fr/alpc.en.html.

[13] Odejobi O.A., Beaumont A.J.,Wong. (2006): "intonation contour realization for standard Yoruba text to speech synthesis A Fuzzy computational approach, computer speech and language, Volume 20,pp 953-956.

[14] Adéwọlé, L.O. (2000). Beginning Yorùbá (Part I).Monograph Series no. 9.Cape Town: CASAS.

[15] Odejobi O.A., Beaumont A.J.,Wong. (2006): "intonation contour realization for standard Yoruba text to speech synthesis A Fuzzy computational approach, computer speech and language, Volume 20,pp 953-956.

[16] Ngogi K., Okelo-Odongo W., Wagacha P.M., (2005): "Swahili Text-to-speech system". African Journal of Science and Technology (AJST) vol.6,No.1 pp.88-89

[17] Goldsmith J., (1990): Auto segmental and Metrical phonology, Blackwell Oxford. http://www.cs.cmu.eduS/~awb/papers/IEEE2002/allthetime.html

.