MACHINE TO MAN COMMUNICATION IN YORUBA LANGUAGE

Abimbola Akintola[1], Tunji Ibiyemi [2]

[1] Department of Computer Science, University of Ilorin, Nigeria
[2] Department of Electrical Engineering, University of Ilorin, Nigeria

Corresponding Author: Abimbola Akintola, abimbolar@yahoo.com

ABSTRACT: Man communicates with man by natural language, sign language, or/and gesture but communicates with machine via electromechanical devices such as mouse, and keyboard. The media of effecting machine to man be enhanced using language technology. This paper proposed a man to machine communication through Text-To-Speech system for Yoruba language. The process used involves text analysis, natural language processing and digital signal processing. The developed system shows an exhaustive representation of Yoruba language through concatenation of possible syllables in the language. The proposed model was evaluated using mean opinion score. Results showed a significant performance with an average of 92% for similarity and naturalness in the samples.
KEYWORDS: Tex-To-Speech, Yoruba Language, Syllable Structure, Similarity, Naturalness, Communication.

1. INTRODUCTION

Speech is the vocalized form of human communication. It is based upon the syntactic combination of lexical and names that are drawn from very large vocabularies. Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units ([BK13]).
A Text-To-Speech (TTS) synthesizer is a machine that takes text from either keyboard, output of a speech recognition system, or scan text document and then convert the text to corresponding speech sound.
The process of TTS can be divided into two:

1. Text and Linguistic Analysis: This is the conversion of text (an imperfect representation of language) into some form of linguistic representation which contains information on the phonemes to be produced, duration, locations silence, and the F0 contour to be used.
2. Speech Synthesis (conversion to speech waveform)

The text and linguistic analysis encompasses:

I. Accent assignment: the assignment of levels of prominence to various words in the sentence.
II. Text preprocessing: to detect word, grammatical part-of-speech assignment, and text normalization
III. Segmental durations: Appropriate duration of phonemes inputs is determined on the basis of linguistics information computed.
IV. Word pronunciation: this is to remove the disambiguation of homographs and pronunciation of names.
V. Intonational phrasing: the breaking of broadened text into one or more intonational units.
VI. *F0 contour computation.*

Speech synthesis is broken down into two parts:

I. The selection and concatenation of appropriate concatenative units given the phoneme string.
II. The synthesis of a speech waveform given the units, plus a model of the glottal source

Yorùbá language is native to Nigeria, Togo and Benin. It is spoken by about 50 million people in south west Nigeria, Togo, Benin, Brazil, UK and USA. It is one of the three official languages of Nigeria and also a member of the Niger-Congo language family ([A+14]). Yorùbá alphabet can be classified into two major kinds, namely: consonants-alphabet without tone accent; and vowels- with tonal accent. The upper and lower Yorùbá alphabets which comprises of both the consonants and vowels are;

A	B	D	E	Ẹ	F	G	Gb	H	I	J	K	L
a	b	d	e	ẹ	f	g	gb	h	i	j	k	l

M	N	O	Ọ	P	R	S	Ṣ	T	U	W	Y
m	n	o	ọ	p	r	s	ṣ	t	u	w	y

The Yorùbá consonants are 18 in number and are drawn from the 25 letters of the Yorùbá alphabets. The consonants are: B, D, F, G, GB, H, J, K, L, M, N, P, R, S, Ṣ, T, W, and Y. The Yorùbá vowels are 7 in number and are also drawn from the 25 letters of the Yorùbá alphabets. The vowels are: A, E, Ẹ, I, O, Ọ, and U.

2. LITERATURE REVIEW

Speech is the vocalized form of human communication. It is based upon the syntactic combination of lexical and names that are drawn from very large vocabularies. Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units ([BK13]).

Text-To-Speech, also known as Speech Synthesis, is the computer production of human speech. It is the process of generating spoken words by machine from written input. Speech is often based on concatenation of natural speech i.e units that are taken from natural speech put together to form a word or sentence. Concatenative speech synthesis, according to Sproat and Olive ([SO99]), has become very popular in recent years due to its improved sensitivity to unit context.

Rhythm also is an important factor that makes the synthesized speech of a TTS system more natural and understandable. The prosodic structure provides important information for the prosody generation model to produce effects in synthesized speech ([SC12]).

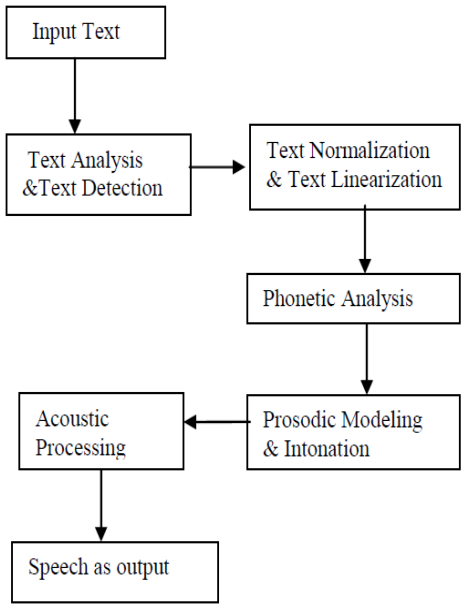

Figure 1: Overview of TTS (Source: [SC12])

Text-To-Speech TTS is still very much at infancy as researchers are working round the clock to have a better algorithm. A TTS system developed through the establishment of corpus-based synthesis unit database that includes nasals, tones, stops and sadhi rules ([S+10]), subsystems of the system includes text-input system, text-to-sound convert system, training of basic synthesis units, and the acoustic wave play system. The system has a multiple accent corpus-based database which was developed using combination of basic phonemes of vowels, consonants and tones from MLT (Modern Literal Taiwanese) books. It has limited speech input but uses large database to develop the MLT. A concatenative synthesis and bell lab approach (combination of phonetics and linguistic structure) to speech synthesis relies on designing and creating the acoustic inventory of the language by taking real recorded speech, cutting it into segments and concatenating these segments back together during synthesis ([C+00]). The synthesizer then produces a concatenative system, based on a set of prerecorded acoustic inventory elements that represent all the possible phone-to-phone transitions of the language. An Arabic system that uses a rule-based hybrid system, which is a combination of formant and concatenative speech techniques reduces the vocabulary independence and can handle all types of input text ([Z+10]). The system omits some vowels of the language in use and also does not take intonation into consideration.

The use of concatenative synthesis bypasses most of the problems encountered by articulatory and formant synthesis techniques ([S+10]). Most developed systems make use of very large database that can slow the system down and also require lots of memory space. The issue of incorrect labeling due the large database can also lead to poor quality of the system.

In ([SGS06]) the system contains front-end which comprised of text analysis and phonetic analysis. The unit selection algorithm is based on Viterbi decoding algorithm of the best-path in the network of the speech units using spectral discontinuity and prosodic mismatch objective cost measures in place of HMM. The back-end is the speech waveform generation based on the harmonic coding of speech. The Harmonic coding enabled the system to compress the unit inventory size by a factor of three. Though, the system used transplanted prosody which does not take intonation into consideration, where generated prosody would have been more effective for the same purpose.

([S+03]) presents techniques for speech-to-text and speech-to-speech automatic summarization. It uses a two-stage summarization method consisting of important sentence extraction and word-based sentence compaction. Sentence and word units which maximize the weighted sum of linguistic likelihood, amount of information, confidence measure, and grammatical likelihood of concatenated units, were extracted from the speech recognition results. For speech-to-speech, sentences, words and between-filler units are investigated as units to be extracted from original speech and concatenated for producing summaries.

The proposed system is a concatenative speech synthesizer and combines real recorded speech

sounds. It is based on prerecorded speech inputs which represents Yoruba language exhaustively by using all possible forms of syllable in the language, the syllabic structure is generated using vowels (v) only and consonant + vowels (cv). Each word is recognized if it exists in the library or broken down into syllables.

3. METHODOLOGY

Text-To-Speech module converts text to speech through text analysis, natural language processing and digital signal processing. Vector Quantitization, and Mel-Frequency Cepstral Coefficient are applied to have better results. Concatenative Synthesis approach of TTS is used to form words by combining syllables.

3.1 Data collection

Speech data (Yorùbá speech corpus): The data collection at this stage involves adequate training and testing data of Yorùbá speech samples. Samples from Male and female Yoruba speakers were captured using continuous and isolated speech sounds. The samples were repeated 5 times each and recorded using 8 kHz, 16 bit, Mono.

Yorùbá Character Generation: This is a distinct catalog of characters (Yorùbá Alphabet, counting numbers and special symbols) recognized by the computer hardware and software.

The character set defines 105 characters, the characters are:

I. 24 uppercase Yoruba alphabet (without GB)
II. 24 lowercase Yoruba alphabet (without gb)
III. 7 uppercase vowel with high tone (`)
IV. 7 lowercase vowel with high tone (`)
V. 7 uppercase vowel with low tone (´)
VI. 7 lowercase vowel with low tone (´)
VII. Digit 0 – 9
VIII. 19 special characters

Syllable: Yorùbá syllable is a unit of pronunciation having one vowel sound, with or without surrounding consonants, forming the whole or a part of a word. Table 1, 2 and 3 shows all possible forms of Yorùbá language syllables.

Tables 1, 2, and 3 show the four hundred and ninety-seven (467) possible syllables in Yoruba language. It includes the vowels alone, the vowels alone with a high tone, vowels alone with low tone, the concatenation of consonant + vowels in high, mid, and low tone.

Phonemes: Yorùbá phonemes are the perceptually distinct units of sound that distinguish a word from another. Table 4 shows the phonemes and pronunciation for Yorùbá alphabets.

Homographs: Yorùbá homographic words are two or more Yorùbá words spelt the same way but not pronounced the same and have different meanings. Yorùbá language makes use of tones to differentiate these words. Table 5 shows some of the homographic words in Yorùbá with corresponding syllable, English meaning and phoneme pronunciation.

Table 1: 201 (Two hundred and one) Yorùbá Syllables with mid tone

A	E	Ẹ	I	O	Ọ	U	Ba	Da	Fa
Ga	Gba	Ha	Ja	Ka	La	Ma	Na	Pa	Ra
Sa	ṣa	Ta	Wa	Ya	Be	De	Fe	Ge	Gbe
He	Je	Ke	Le	Me	Ne	Pe	Re	Se	ṣe
Te	We	Ye	Bẹ	Dẹ	Fẹ	Gẹ	Gbẹ	Hẹ	Jẹ
Kẹ	Lẹ	Mẹ	Nẹ	Pẹ	Rẹ	Sẹ	ṣẹ	Tẹ	Wẹ
Yẹ	Bi	Di	Fi	Gi	Gbi	Hi	Ji	Ki	Li
Mi	Ni	Pi	Ri	Si	ṣi	Ti	Wi	Yi	Bo
Do	Fo	Go	Gbo	Ho	Jo	Ko	Lo	Mo	No
Po	Ro	So	ṣo	To	Wo	Yo	Bọ	Dọ	Fọ
Gọ	Gbọ	Họ	Jọ	Kọ	Lọ	Mọ	Nọ	Pọ	Rọ
Sọ	ṣọ	Tọ	Wọ	Yọ	Bu	Du	Fu	Gu	Gbu
Hu	Ju	Ku	Lu	Mu	Nu	Pu	Ru	Su	ṣu
Tu	Wu	Yu	N	M	An	Ẹn	Ọn	Un	Ban
Dan	Fan	Gan	Gban	Jan	Kan	Lan	Han	Yan	Pan
Ran	San	ṣan	Tan	Wan	Bẹn	Dẹn	Fẹn	Gẹn	Gbẹn
Hẹn	Jẹn	Lẹn	Pẹn	Rẹn	Sẹn	ṣẹn	Tẹn	Wẹn	Yẹn
Bọn	Dọn	Fọn	Gọn	Gbọn	Họn	Jọn	Kọn	Lọn	Pọn
Rọn	Sọn	ṣọn	Tọn	Wọn	Yọn	Bun	Dun	Fun	Gun
Gbun	Hun	Jun	Lun	Pun	Run	Sun	ṣun	Tun	Wun
Yun									

Table 2: 133 Upper Tone of All Possible Form of Yoruba Syllables

A	E	Ẹ	I	O	Ọ	U	Bà	Dà	Fà
Gà	Gbà	Hà	Jà	Kà	Là	Mà	Nà	Pà	Rà
Sà	Ṣà	Tà	Wà	Yà	Bè	Dè	Fè	Gè	Gbè
Hè	Jè	Kè	Lè	Mè	Nè	Pè	Rè	Sè	Ṣè
Tè	Wè	Yè	Bẹ̀	Dẹ̀	Fẹ̀	Gẹ̀	Gbẹ̀	Hẹ̀	Jẹ̀
Kẹ̀	Lẹ̀	Mẹ̀	Nẹ̀	Pẹ̀	Rẹ̀	Sẹ̀	ṣẹ̀	Tẹ̀	Wẹ̀
Yẹ̀	Bì	Dì	Fì	Gì	Gbì	Hì	Jì	Kì	Lì
Mì	Nì	Pì	Rì	Sì	Ṣì	Tì	Wì	Yì	Bò
Dò	Fò	Gò	Gbò	Hò	Jò	Kò	Lò	Mò	Nò
Pò	Rò	Sò	ṣò	Tò	Wò	Yò	Bọ̀	Dọ̀	Fọ̀
Gọ̀	Gbọ̀	Họ̀	Jọ̀	Kọ̀	Lọ̀	Mọ̀	Nọ̀	Pọ̀	Rọ̀
Sọ̀	ṣọ̀	Tọ̀	Wọ̀	Yọ̀	Bù	Dù	Fù	Gù	Gbù
Hù	Jù	Kù	Lù	Mù	Nù	Pù	Rù	Sù	ṣù
Tù	Wù	Yù							

Table 3: 133 Lower Tone of All Possible Form of Yoruba Syllables

A	E	Ẹ	I	O	Ọ	U	Bá	Dá	Fá
Gá	Gbá	Há	Já	Ká	Lá	Má	Ná	Pá	Rá
Sá	Ṣá	Tá	Wá	Yá	Bé	Dé	Fé	Gé	Gbé
Hé	Jé	Ké	Lé	Mé	Né	Pé	Ré	Sé	Ṣé
Té	Wé	Yé	Bẹ́	Dẹ́	Fẹ́	Gẹ́	Gbẹ́	Hẹ́	Jẹ́
Kẹ́	Lẹ́	Mẹ́	Nẹ́	Pẹ́	Rẹ́	Sẹ́	ṣẹ́	Tẹ́	Wẹ́
Yẹ́	Bí	Dí	Fí	Gí	Gbí	Hí	Jí	Kí	Lí
Mí	Ní	Pí	Rí	Sí	Ṣí	Tí	Wí	Yí	Bó
Dó	Fó	Gó	Gbó	Hó	Jó	Kó	Ló	Mó	Nó
Pó	Ró	Só	ṣó	Tó	Wó	Yó	Bọ́	Dọ́	Fọ́
Gọ́	Gbọ́	Họ́	Jọ́	Kọ́	Lọ́	Mọ́	Nọ́	Pọ́	Rọ́
Sọ́	ṣọ́	Tọ́	Wọ́	Yọ́	Bú	Dú	Fú	Gú	Gbú
Hú	Jú	Kú	Lú	Mú	Nú	Pú	Rú	Sú	ṣú
Tú	Wú	Yú							

Table 4: Thirty (30) Yorùbá phonemes

S/No.	Phoneme	Pronunciation
1	/b/	B
2	/d/	D
3	/f/	F
4	/g/	G
5	/g͡b/	Gb
6	/h/	H
7	/dʒ/ or /ɟ/	J
8	/k/	K
9	/l/	L
10	/m/	M
11	/n/	N
12	/k͡p/	P
13	/r/	R
14	/s/	S
15	/ʃ/	ṣ

S/No.	Phoneme	Pronunciation
16	/t/	T
17	/w/	W
18	/j/	Y
19	/a/	A
20	/e/	E
21	/ɛ/	ẹ
22	/i/	I
23	/o/	o
24	/ɔ/	ọ
25	/u/	U
26	/ã/	An
27	/ɛ̃/	ẹn
28	/ĩ/	In
29	/ɔ̃/	ọn
30	/ũ/	Un

Table 5: Yorùbá Homographic Words

S/NO	Word	Homographs	Syllable	Meaning	Pronunciation
1	Aba	Abá	A/bá	Attempt	Abá
		Àbá	À/bá	Mat	Àbá
		Abà	A/bà	Barn	Abá
		Aba	A/ba	Staple, Incubation	Aba
2	Abe	Abẹ́	A/bẹ́	Bottom	abɛ́
		Abẹ	A/bẹ	Razor	abɛ
3	Abo	Abo	A/bo	Female	Abo
		Àbò	À/bò	Refuge	Àbò
4	Aja	Ajá	A/já	Dog	adʒá
		Ajà	A/jà	Attic	adʒà
5	Ajẹ	Àjẹ́	À/jẹ́	Sorcerer	àdʒɛ́
		Àjẹ̀	À/jẹ̀	Oar, Paddle	àdʒɛ̀
6	Ala	Àlá	À/lá	Dream	Àlá
		Àlà	À/là	Boundary	Àlà
7	Apa	Apà	A/pà	Arm	a͡kpà
		Àpa	À/pa	Prodigal	àkpa
		Apá	A/pá	Mark, Sign	a͡kpá
8	Ara	Ara	A/ra	Body	Ara
		Ará	A/rá	Relative	Ará
		Àrá	À/rá	Thunder	Àrá
		Àrà	À/rà	Fashion	Àrà
9	Baba	Baba	Ba/ba	Father	Baba
		Bàbà	Bà/bà	Guinea Corn	Bàbà
		Bàbá	Bà/bá	Great thing	Bàbá
10	Dana	Dáná	Dá/ná	Make fire	Dáná
		Dánà	Dá/nà	Robbery	Dánà
		Dána	Dá/na	Pay dowry	Dána
11	Ede	Èdè	È/dè	Dialect	Èdè
		Edé	E/dé	Lobster	edé
		Èdé	È/dé	Buffalo	Èdé
12	Ere	Ère	È/re	Gain	Ère
		Eré	E/ré	Game	Eré
		Èrè	È/rè	Statue	Èrè
		Erè	E/rè	Snake	Erè
13	Ewu	Èwú	È/wú	A day pounded yam	Èwú
		Ewu	E/wu	Danger	ewu
		Ewú	E/wú	Grey hair	ewú
14	Ẹfon	Ẹfọn	Ẹ/fọn	Mosquito	èfɔn
		Ẹfọ́n	Ẹ/fọ́n	Arrow	ɛfɔ́n
		Ẹfọn	Ẹ/fọn	Buffalo	ɛfɔn
15	Ẹgba	Ẹgba	Ẹ/gba	Whip	ɛg͡ba
		Ẹgbà	Ẹ/gbà	Two thousand	ɛg͡bà
		Ẹgbà	Ẹ/gbà	Bracelet	èg͡bà
		Ẹgbá	Ẹ/gbá	Yorùbá Tribe	èg͡bá
16	Ẹrin	Ẹrín	Ẹ/rín	Laughter	èrĩ
		Ẹrin	Ẹ/rin	Four	èrĩ
17	Ẹtu	Ẹtù	Ẹ/tù	Guinea Fowl	ɛtù
		Ẹtù	Ẹ/tù	Gun Powder	ètù
		Ẹtu	Ẹ/tu	Antelope	ɛtu
18	Ẹwa	Ẹwa	Ẹ/wa	Ten	ɛwa
		Ẹwà	Ẹ/wà	Beauty	ɛwà
		Ẹ̀wà	Ẹ̀/wà	Beans	èwà
19	Giri	Gìrì	Gì/rì	Convulsion	gìrì
		Gírí	Gí/rí	Promptly	gírí
		Girì	Gi/rì	Suddenly	girì
20	Gba	Gbà	Gbà/	Receive	g͡bà
		Gbá	Gbá/	Sweep	g͡bá

(Table 5)

21	Gbo	Gbo᷄	Gbo᷄/	Bark, Ripen	g͡bó
		Gbò	Gbò/	To affect	g͡bò
22	Iba	Ìba	Ì/ba	Few	Ìba
		Ìbà	Ì/bà	Respect	Ìbà
		Ibà	I/bà	Fever	Ibà
23	Ibo	Ìbò	Ì/bò	Plant	Ìbò
		Ibo	I/bo	Where	Ibo
24	Idi	ÌdÌ	Ì/dÌ	Bundle	ÌdÌ
		Idi	i/di	Bud	Idi
		Ìdí	Ì/dí	Waist, Reason	Ìdí
25	Igba	Ìgbà	Ì/gbà	Time	Ìg͡bà
		Igba	I/gba	Two thousand	Ig͡ba
		Igbá	i/gbá	Calabash	ig͡bá
		Ìgbá	Ì/gbá	Locust beans	Ìg͡bá
		Igbà	i/gbà	Rope for climbing	ig͡bà
26	Ika	Ìkà	Ì/kà	Cruelty	Ìkà
		Ìka	Ì/ka	Finger	Ìka
27	Iko	Ìkó	Ì/kó	Hook	Ìkó
		Ìkò	Ì/kò	Delegate	ìkò
		Ikó	I/kó	Cough	ikó
28	Obi	Òbí	Ò/bí	Parent	Òbí
		Obì	O/bì	Kolanut	obì
29	Ogun	Ogún	O/gún	Inheritance	ogṹ
		Ògún	Ò/gún	God of iron	ògṹ
		Ógún	Ó/gún	Medicine	ógṹ
		Ogùn	O/gùn	Twenty	ogũ̀
		Ogun	O/gun	War	ogũ
30	Ojo	Òjò	Ò/jò	Rain	òjò
		Ojo	O/jo	Fear	ojo
		Òjó	Ò/jó	Name	Òjó
31	Okun	Òkun	Ò/kun	Sea	òkũ
		Okùn	O/kùn	Rope	okũ̀
		Okun	O/kun	Strength	okũ
32	Orun	Orùn	O/rùn	Sun	orũ̀
		Orun	O/run	Sleep	orũ
		Orún	O/rún	Scent	orṹ
33	Ọka	Ọkà	Ọ/kà	Corn	ɔkà
		Ọkà	Ọ̀/kà	Child's disease	ɔ̀kà
		Ọká	Ọ/ká	Snake	ɔká
34	Ọkọ	Ọkọ̀	Ọ/kọ̀	Canoe	ɔkɔ̀
		Ọ̀kọ̀	Ọ̀/kọ̀	Spear	ɔ̀kɔ̀
		Ọkọ́	Ọ/kọ́	Hoe	ɔkɔ́
		Ọkọ	Ọ/kọ	Husband	ɔkɔ
35	Ọrun	Ọ́rún	Ọ́/rún	Bow	ɔ́rṹ
		Ọrùn	Ọ/rùn	Neck	ɔrũ̀
		Ọrún	Ọ/rún	Hundred	ɔrṹ
		Ọrun	Ọ/run	Heaven	ɔrũ
36	Ọwọ	Ọwọ̀	Ọ/wọ̀	Honour	ɔwɔ̀
		Ọwọ́	Ọ/wọ́	Flock of birds	ɔwɔ́
		Ọwọ̀	Ọ/wọ̀	Broom	ɔwɔ̀
		Ọwọ́	Ọ/wọ́	Hand	ɔwɔ́

3.2 Sound library

The sound library houses the recorded words, continuous speech, vowels, phonemes, syllables and homographs pronunciation. The total number of sounds in the library is 31750.

The phonemes and their pronunciations are thirty (30) as shown in Table 4. All forms of syllables as derived from Table 1, 2, 3 are four hundred and ninety seven (467), digit 0 – 9, 20 continuous speech sounds. This comprises of vowels (V), consonant vowel (CV) nasal stops (M and N). The Thirty-Six (36) lexis which gave rise to (108)

homographic words were also included in the library. All samples were pronounced five (5) times each by five (5) male and five (5) female Yoruba speakers.

Pseudo Code for Converting Text-To-Speech

```
Repeat
{  For each word
   {
      If word is in library generate
      speech else
      {For each syllable
         {
            If syllable in
            library generate
            speech else
               {
                  For each letter
                  generate speech
                  }
            }
      }
   }
} until end of text
```

Text-To-Speech interface: This comprises of text input from the speech to text implementation, input from the Yoruba keyboard and input from optical character recognition. Each of the inputs is to be performed independently. The TTS breaks the text into possible Yoruba Syllables and concatenate it to give corresponding utterances. Figure 2 below shows the test to speech interface.

.

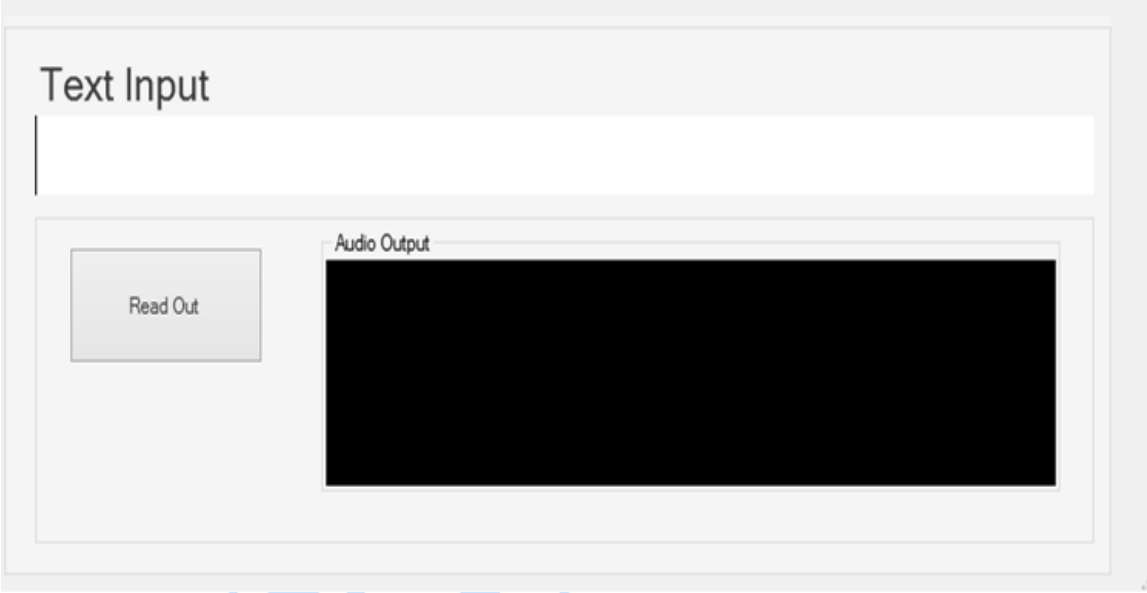

Figure 2: Text to Speech Interface

Input from Keyboard: The input from keyboard is generated from a virtual Yoruba keyboard that helps to type Yoruba characters on the Text Input tab. Figure 3 shows the keyboard used.

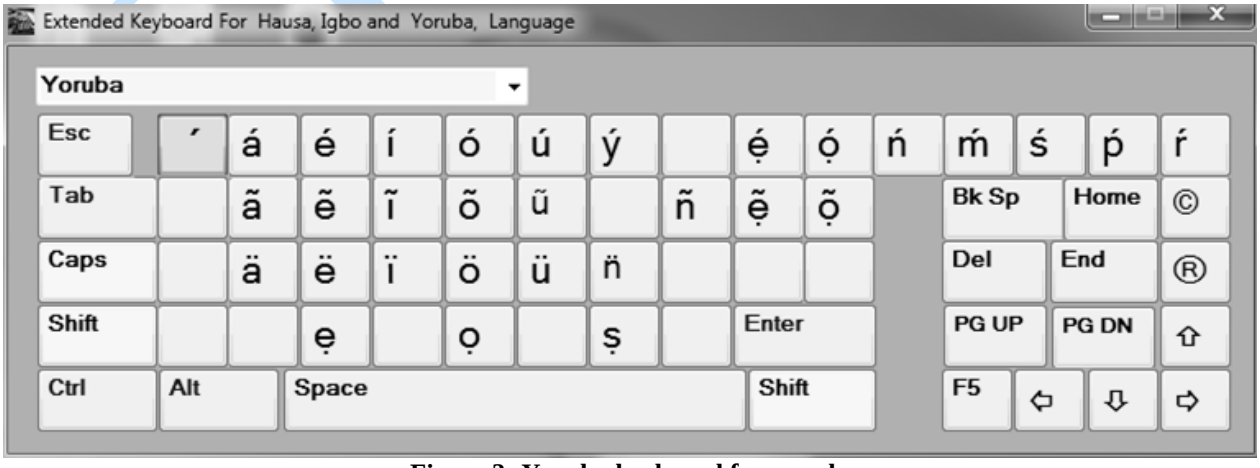

Figure 3: Yoruba keyboard for vowels

The inputs from the keyboard ensure that the text is in machine readable format before text to speech is performed. The figure 4 shows the text input box and figure 5 shows the waveform for the conversion to speech of the text input.

Figure 4: Text Input Interface

Figure 5: Speech waveform

4. RESULTS

The evaluation of Text-To-Speech was done using Listening Test conducted through Mean Opinion Score. The scoring methods of the listening tests include:

1. Similarity: The listener plays a few samples of the original speech and one synthetic sample. The listener then chooses a response that represented how similar the synthetic voice sounded as compared to the original speaker's voice.
2. Naturalness : The listener listens to a sample of synthetic speech and chooses a score which represents how natural or unnatural the sentence sounded on a scale of 1-5 ([K+14], [SMC15])

The listeners are to score from 1 to 5 (Excellent – 5 Very good – 4 Good – 3 Satisfactory – 2 Not understandable-1) for understandable contents.

Table 6 below shows the Mean Opinion Score based on similarity and naturalness from the listener perspective. Five sentences were used in the test. Each sentence consists of at least 5 words and 13 syllables.

Table 6: Scores given by each listener to each sentence

Sentence	A	B	C	D	E
Sentence 1	5	5	5	5	5
Sentence 2	5	5	4	3	3
Sentence 3	5	5	5	5	5
Sentence 4	5	4	4	5	5
Sentence 5	5	3	5	5	4

Performance Average = 92%

The performance showed average of 92% closeness to human voice.

5. CONCLUSION

The Machine To Man though Text-To-Speech system represent Yoruba language exhaustively. Hence, the following recommendations are made:

1. For deployment to grass-root level so that people who do not know how to speak English language can benefit from technology advancement;
2. For the aged and illiterate that may not be able to read Yoruba text and literature (bible, Qur'an, Yoruba newspapers and so on), it can be fed into

the system to read out in Yoruba language to them; and

3. The disabled and unskilled people can also benefit from the system in the sense that they may not be able to use electromechanical devices like keyboard and joystick but can speak to and hear from the system.

REFERENCES

[A+14] **O. Abiola, A. Adetunmbi, A. Fasiku, K. Olatunji** - *Web-based English to Yoruba noun-phrases machine translation system.* International Journal of English and Literature. Vol 5(3), 71 – 78, 2014.

[BK13] **C. Bhusan, B. Krishna** - *Nepali Text to Speech Synthesis System using ESNOLA Method of Concatenation*, International Journal of Computer Application, vol. 62: 24-28, 2013.

[C+00] **C. Christogiannis, T. Varvarigou, A. Zappa, Y. Vamvakoulas** - *Construction of the acoustic inventory for Greek Text-To-Speech Synthesis system.* In Proceedings of International Conference on Spoken Language Processing, 267-270, 2000.

[K+14] **P. Kishore, V. Anandaswarup, K. Santosh, A. Hema, P. Swaran, S. Mahadeva, P. Hemant, K. Anil, K. Simon, W. Alan, T. Keiichi** - *The Blizzard Challenge 2014.* http://www.festvox.org/blizzard/ [March 16, 2015]

[SC12] **D. Sasirekha, E. Chandra** - *Text-to-speech: a simple tutorial.* International Journal of Soft Computing and Engineering, vol 2, 275-279, 2012.

[SO99] **R. Sproat, J. Olive** - *Text-to-Speech Synthesis* In K. Madisetti & B. Douglas (Eds.) Digital Signal Processing Handbook, Boca Raton: CRC Press LLC, 1999.

[SGS06] **H. Sak, H., T. Gungor, Y. Safkar** - *A Corpus-Based concatenative speech synthesis system for Turkish*. Turkish Journal of Electrical Engineering and Computer Science, vol. 14: 209 – 223, 2006.

[SMC15] **K. Sangrramsing, M. Monica, K. Charansing** - *Performance Evaluation of Speech Techniques for Marathi Language*. International Journal of Computer Applications, vol 130:3, 45-50, 2015.

[S+03] **F. Sadaoki, K. Tomonori, S. Yousuke, H. Chiori** - *Speech-to-Speech and Speech-to-Text Summarization*. First International Workshop on Language Understanding and Agents for Real World Interaction. 2003.

[S+10] **Y.-J. Sher, Y.-H. Chiu, M.-C. Hsu, K.-C. Chung** - *Development of Hmm Based Taiwanese Text-To-Speech system.* In International Conference on Software Techniques and Engineering, 330-333, 2010.

[Z+10] **M. Zeki, O. Othman, A. Khalifa, W. Naji** - *Development of an Arabic Text-To-Speech System.* In International Conference on Computer and Communication Engineering (ICCCE), 11-14, 2010.