



## Agricultural e-Extension Services: A Hybrid of Multilingual Translation Text-to-Speech-A Framework

Yahaya Mohammed Sani<sup>1</sup>, Stella Oluyemi Etuk<sup>2</sup>, Ilyasu Anda<sup>3</sup>, and Mamman Adamu<sup>4</sup>

<sup>1,2</sup>Department of Information and Media Technology, Federal University of Technology, Minna, Nigeria

<sup>3</sup>Department of Library and Information Technology, Federal University of Technology, Minna, Nigeria

<sup>4</sup>Department of Computer Science, Federal University of Technology, Minna, Nigeria

{<sup>1</sup>yahayasani, <sup>2</sup>abiolastella, <sup>3</sup>ilyasu.anda} @futminna.edu.ng, <sup>4</sup>bejian2004@gmail.com

**Abstract**—This paper presents a framework for a text-to-speech translation on Android Devices based on Natural Language Processing (NLP) and text-to-speech synthesizer (TTS) to deliver real-time agricultural update to farmers by agricultural extension service workers (AEW) as speech is the most used and natural way for people to communicate with one another. In order to increase the naturalness of oral communications between Agricultural Extension Service workers and farmers, speech aspects must be involved. This is because most local farmers have good understanding of their local language and have strong preferences for it over any other language. Since, majority of farmers are in rural areas, they have little or no understanding of English language, agriculture research output communicated in English language, may be of little or no use to them, if they are delivered in a foreign language. Text-to-Speech Enabled Hybrid Multilingual Translation framework adopts a serial integration of Natural Language Processing (NLP) on one hand and text-to-speech synthesizer (TTS) interpretation technique using android google translate API text-to-speech synthesizer and recognizer to translate English, Hausa, Yoruba, Ibo and Arabic texts in to speech(es) respectively in accordance with farmers registered dialect on the other hand.

**Keywords**-multilingual translation; Natural language Processing (NLP); Text-to-Speech Synthesizer (TTS); e-Extension

### I. INTRODUCTION

Agricultural extension services delivery is a system that facilitates farmers access, their groups and other market players to information, knowledge and technologies [1] as well as the process of bringing extension workers, researchers and farmers together in order to bring suitable technology transformation in agricultural production and thereby raise living standards of rural folks [2]. Today's role of extension services involves facilitating technology transfer, helping farmers to establish associations, overcome market challenge and partner with service providers and other relevant agencies [3]. In the last few decades, agricultural information dissemination has been conquered by both print and electronic media; newspapers, television, magazines and the like. However, social media as a form of

ICT method for harmonizing sustainable agriculture and natural resources provides wider agricultural community that eliminates physical distance [4]. In the light of the foregoing, an efficient and effective agricultural extension is measured by extension messages (information) reaching farmers on one hand and farmers challenges reaching extension agents promptly and on timely basis on the other hand with a view to proffer lasting solutions to their challenges [5, 6]. Regrettably, in many sub-Saharan African countries, poverty and food insecurity have been identified as some of the underlying cause of under development [7]. This is in view of numerous challenges encountered in delivering extension and consultative services (EAS) to farmers. These ranges from inadequate funds poor extension infrastructure, non-inclusion of rural farmers in extension framework development, to inappropriate strategies for effective agricultural research and ineffective and inefficient extension practice. Partial reporting of extension services across rural regions and problem of embracing technology packages that are community-oriented have been describe as major challenges in the effective coverage of extension services [8]. Moreover, the traditional technique of relaying agricultural research update to farmers through physical contact has various bottlenecks due to many constraints such as geographical location of farmers to extension agents, poor communication infrastructure, poor carriage services, deplorable condition of roads, insufficient funds and language barrier, has been identified as some of the challenges to effective and timely dissemination of real-time research update to agriculturalists. This technique cannot sufficiently cater for real-time research up-to-date information should reach the rural farming community. In view of the foregoing constraints, a dynamic ICT mediated knowledge management model should be embraced to agricultural researches carried out at any university/research organization anywhere in the world and this research can be disseminated in a matter of seconds amongst scientists, faculty, private industry, students, extension workers and smallholder farmers.

Text-to-speech (TTS) is the generation of synthesized speech from text. During synthesis very small segments of recorded human speech are concatenated together to produce

the synthesized speech. The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. A text-to-speech synthesizer allows illiterate farmers with reading and writing disabilities to listen to real-time agricultural update on Android devices. Many computer operating systems have included speech synthesizers since the early 1990s [9]. The speech recognition and text-to-speech (TTS) is done using public Android APIs. On most devices, these make use of Google's speech recognizer and synthesizer, which are available in both online and offline versions. The offline engines tend to have a reduced choice of languages and reduced quality compared to the online engines, but don't require an internet connection [10]. For example:

Query 1:

English: Are you a farmer?

Hausa: Shin kai manomi ne

Yoruba: Še o je agbe

Ibo: i bu onye oru ugbo

Arabic: هل انت مزارع

Query 2:

English: What type of crop do you cultivate?

Hausa: wane irin amfanin gona kake noma

Yoruba: Iru irugbin wo ni o se

Ibo: kedu udj ihe ubi i na-azulite

Arabic: ما نوع المحاصيل التي تزرعها

There is yet no attempt in Nigeria in this area for a suitable approach; however, we hypothesize in this paper that setting a translation of text-to-speech to indigenous language of a farmer from the context of agricultural e-extension services delivery research output as a practical use case can help achieve the main goal of agricultural extension services delivery. Therefore, it is an essential technique to translate agricultural research output and translate them into user-understandable native or indigenous languages. Since the goal of this paper is to make synthesized speech as intelligible, natural and pleasant to listen, as human speech. Speech is the primary means of communication between agricultural extension workers and farmers.

## II. RELATED CONCEPTS

### A. An Overview of Speech Synthesis

Speech synthesis can be termed as artificial creation of human speech [11]. A computer system used for this motive is termed a speech synthesizer, and can be executed in either software or hardware. A text-to-speech (TTS) system translates standard language text into speech [12]. Synthesized speech can be produced by concatenating bits of recorded speech that are warehoused in a databank. Systems vary in the size of the warehoused speech components; a system that stocks phones or diphones provides the largest range of output, with little or no clarity. For particular usage domains, the storing of whole words or sentences allows for high-quality output. On the other hand, a synthesizer can integrate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output [13]. The quality of a speech synthesizer is adjudged

by its likeness to the human voice and by extension its ability to be comprehended. An intelligible text-to-speech programme enable people with visual impairments or illiterate farmers with reading disabilities to listen to written works on an Android mobile device. A text-to-speech system (or "engine") is composed of two parts: [14] a front-end and a back-end. The front-end has two (2) major tasks. First, it converts raw text comprising of symbols such as numbers and abbreviations into the corresponding 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. Phonetic transcriptions 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 then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody (pitch contour, phoneme durations), [15] which is then imposed on the output speech.

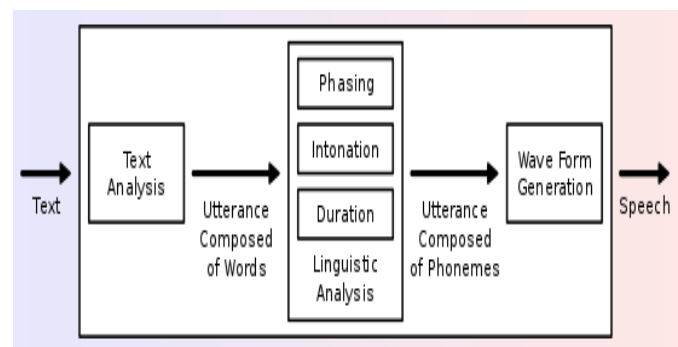


Figure 1. Overview of Text-to-Speech System [16]

### B. Review of Existing e-Extension Services Delivery Systems

[17] Reported on the readiness, usage and information pursuing conduct of a farming locality with particular emphasis to ICTs. The findings therefore show that research gap was filled by observing how farmers respond information medium when is accessible to them, instead of focusing on obstacles that envelope the use of ICTs and digital divide matters arising as a result of varying access and capabilities.

[18] Noted the impact of ICT to extension services delivery for horticulture farmers in Uasin Gishu County in order to ascertain the level of utilization of ICT by horticulture farmers was studied by. The challenges and opportunities to effective use of ICT were carried out. The study was however guided by correlation research design. The findings revealed that utilization of ICT by horticulture farmers is abysmally low. It further shows that majority of the farmers still rely on traditional ICTs to get agricultural information.

The role of ICT to facilitate the Ethiopian agricultural extension delivery system was examined by [19]. Farmers were given different sources of options in using ICT tools in collecting information on suitable technologies required to meet their particular farming needs and those that encircle their local, social, cultural, economic and political atmosphere. The results of this study however, revealed that different infrastructure of ICT have been utilized in extension services delivery and were more in tune with the situations and desires of small scale farmers with language barrier being a major issue.

A mobile-based farmers' advisory information system (M-FAIS) aimed at improving the coverage of extension services in Tanzania was presented by [20]. The study contributes to the approach for testing complex system in agricultural informatics. The approach integrates the conventional system testing approaches (technology-centric approach) and extends it to include consideration of system usability (user-centric testing approach) prior to system release. In view of the foregoing, the major setback of this system is that it does not take dialectal problems in to account.

Also, a multilingual translation system to enhance agricultural e-extension services delivery was designed, developed and implemented by [21]. The implemented system adopts an hybrid machine translation techniques to translate agricultural information from English (source text) into farmer's chosen registered local dialect. The system integration test shows 65% accuracy in translation capability of the research outputs from English-to-farmer's registered local dialect. However, the major limitation of the system is that it cannot translate the agricultural text messages from English to the native languages in a text-speech fashion; its capability is only text-text translation.

Arising from the above literature review on related work, there is yet no attempt in Nigeria in this area for a suitable approach; however, we hypothesize that setting a translation of text-to-speech to indigenous language of a farmer from the context of agricultural e-extension services delivery research output as a practical use case can help achieve the main goal of agricultural extension services. Therefore, it is an essential technique to translate AES research output and translate them into user-understandable native or indigenous languages. This is the gap that this research is set to achieve.

### III. RESEARCH METHODOLOGY

#### A. Multilingual Language Translation Text-to-Speech Framework

Text-to-speech synthesis takes place in several steps. The TTS systems get a text as input, which it first analyze and then convert into a phonetic form. Then in subsequent step it produces the prosody. From the available information, it can generate a speech signal.

The configuration of the text-to-speech synthesizer can be fragmented into two (2) major components:

- Natural Language Processing (NLP) Component: This module produces a phonetic transcription of the text read, together with prosody.

- Android Google Speech Synthesizer (AGSS) Component: This component transforms the symbolic information it receives from NLP into audible and intelligible speech.

The operational procedure of the NLP component is as follows:

1) Text Analysis: First the text is segmented into tokens. The token-to-word conversion creates the orthographic form of the token. For the token "Ms" the orthographic form "Miss" is formed by expansion, the token "10" gets the orthographic form "ten" and "2018" is transformed to "two thousand and eighteen".

2) Application of Pronunciation Rules: After the text analysis has been accomplished, the application of pronunciation rules takes place. Letters cannot be converted 1:1 into phonemes due to information mismatch. In certain atmospheres, a single letter can match to either no phoneme (for instance, "h" in "laugh") or several phoneme ("n" in "next"). In addition, several letters can correspond to a single phoneme ("re" in "store"). There are two approaches to define pronunciation:

- In dictionary-based solution with morphological modules, as many morphemes (words) as possible are warehoused in a dictionary. Full forms are generated by means of inflection, derivation and composition rules. On the other hand, a full form dictionary is used in which all conceivable word forms are warehoused. Pronunciation rules define the pronunciation of words not found in the dictionary of any of this particular language from google translate API (from English to Hausa, Yoruba and Ibo) respectively).
- In a rule based solution, pronunciation rules are produced from the phonological information dictionaries. Only words whose pronunciation is a complete exception are included in the dictionary of any of this particular language English, Hausa, Yoruba and Ibo).

The two applications vary significantly in the size of their dictionaries. The dictionary-based solution is often times larger than the rules-based solution's dictionary of exception. Nevertheless, dictionary-based solutions can be more thorough than rule-based solution if they have a large enough phonetic dictionary available.

3) Prosody Generation: after the pronunciation has been ascertained, the prosody is produced. The extent of naturalness of a TTS system is dependent on prosodic factors like intonation modeling (phrasing and accentuation), amplitude modeling and duration modeling (including the duration of sound and the duration of pauses, which determines the length of the syllable and the tempos of the speech) [22].

The output of the NLP component is passed to the Android google speech recognizer and synthesizer (AGSS) component. This is where the actual synthesis of the speech signal occurs. In concatenative synthesis the selection and

linking of speech segments take place. For individual sounds the best option (where several appropriate options are available) are selected from a database (Android google text-to-speech speech analyzer and synthesizer for English, Hausa, Yoruba and Ibo) respectively and concatenate it).

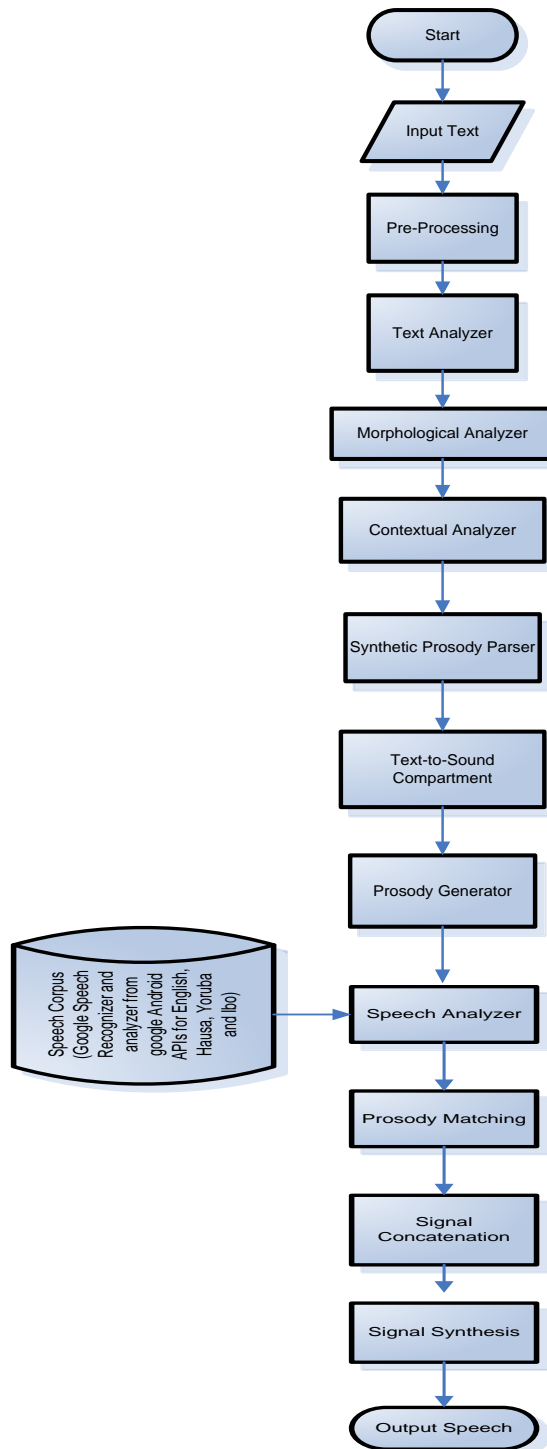


Figure 2. Operational procedure of natural language processing component of TTS synthesizer

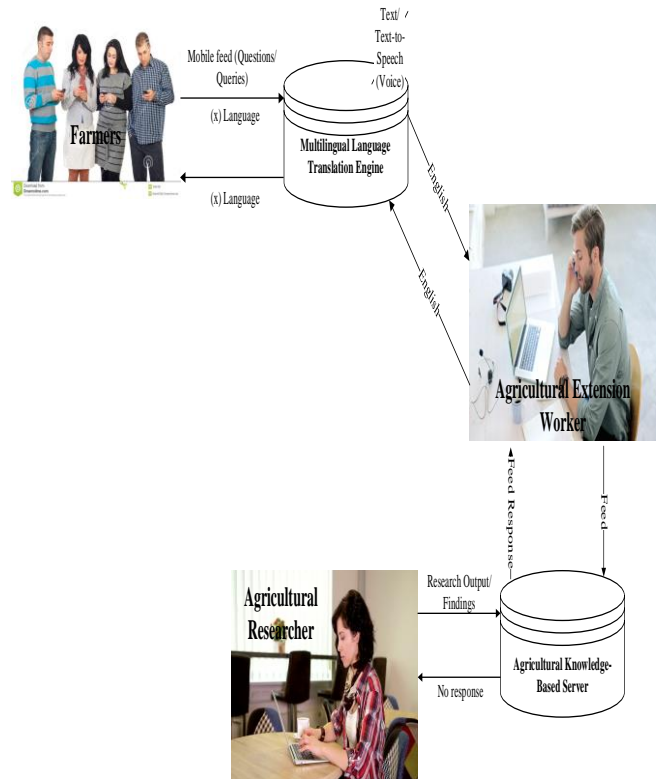


Figure 3. The Proposed Framework

### B. Framework Representation

In this research work, the proposed multilingual translation text-to-speech framework deploys three (3) tiers architecture namely: the front-end (presentation tier), the middle tier and the back-end (data tier) as described underneath:

1) *Presentation Tier*: This layer also referred to as the front end; information is made available to client (user interface) at this phase via the browsers (like Mozilla Firefox or Internet explorer).

2) *Middle Tier*: Also called server side. This layer was developed with the use of the following tools: Hyper Text Markup language (HTML), Cascading Style Sheet (CSS), Java programming language [23] google translate API for text-to-text component and Android google translate API text-to-speech recognizer and synthesizer for text-to-speech component for the following languages: English, Hausa, Yoruba, Ibo and Arabic respectively. Multilingual translation of text-to-text and text-to-speech equally takes place in this layer.

3) *Data Tier*: This layer is referred to as the back-end. This layer will provide the data warehouse for the designed framework. This framework proposed to adopts google API for text –to-text component and Android google API Text-to-speech recognizer and synthesizer for text-to-speech component for the following languages: English, Hausa, Yoruba, Ibo and arabic respectively.

### C. Implementation Algorithm for the Framework Design

The proposed framework for agricultural e-extension services: a hybrid multilingual language translation text-to-speech as shown in Fig. 2. The framework links three (3) major actors together (farmer, researcher/expert and agricultural extension workers). The implementation algorithm is explained as follows:

- Farmers initiate feeds (that is questions/queries) using their mobile phones in any of their registered/choosing language(s) (that is Arabic, Hausa, Yoruba and Ibo) tagged (x) language to the multilingual language translation engine (MLTE) in either text-text or text-to-speech fashion.
- Multilingual Language Translation Engine (MLTE); translate the feeds into English and forward it to Agricultural Extension Worker (AEW) in text pattern.
- AEW then feeds the web server (Agricultural Knowledge-Based Server) in English.
- The server respond; while sending feed request back to AEW, if there is no response to the query/question from the server, the server then send the query or question to the researcher for onward research and reply the research response back to the server-AEW-Farmers.
- AEW checks the response and forward it to multilingual language translation engine (MLTE).
- MLTE; then translate the feeds response to the farmers registered languages respectively using a unique id (farmers language identity) in order to forestall any possible misplacement of translated responses to farmers in either text-text or text-to-speech format.

### IV. FRAMEWORK EVALUATION

The application named the Text-to-Speech Intelligible System is a simple application with the text to speech functionality. The system was proposed to be developed and implemented using Java programming language. Java is used because it is robust and independent platform. The application is divided into two (2) main components: the main application module which includes the basic GUI components which handles the basic operations of the application such as input of parameters for conversion either via file or direct keyboard input or the browser. This would make use of the open source API called google translate API (for English, Hausa, Yoruba Ibo and Arabic) from google library respectively for text-to-text module. The second module, the main conversion engine which integrated into the main module is for the acceptance of data hence the conversion. This would implement the API called freeTTS for Android google text-to-speech synthesizer (for English, Hausa, Yoruba, Ibo and Arabic) respectively.

### V. CONCLUSION

This research work proposed a framework for agricultural e-extension services: an hybrid of multilingual

language translation text-to-speech that ensures by the second agricultural research information (ARI) is delivered to agriculturalists regardless of their geographical boundaries and dialect. The proposed framework convert agricultural research finding from a source dialect (English) into four (4) identified local target dialects (Hausa, Arabic, Yoruba, Ibo and Arabic) based on selected local language understood by the farmer. To this end, the framework also links the key players (agricultural expert, extension workers and farmers) in agricultural research information generation, processing and broadcasting under one umbrella. This is accomplished by allowing the extension workers to send farmers queries to experts and receive feedback via the mobile application and the web-based system as well. Moreover, farmers take delivery of timely text-to-text and text-to-speech agricultural research communications from the extension via their Android cellphones based on requests, queries forwarded to research for them on their area of interest or preference.

### REFERENCES

- [1] FAO, Agricultural Extension Services Delivery System in Nepal Food and Agriculture Organization of the United Nations UN Complex, Pulchowk, Nepal, PP 1-63, 2010.
- [2] D. Kishore and V. Gupt, ICT for agricultural extension: a study in the Indian Himalayan region, the electronic J on info Sys in developing countries, vol. 48 No.3, pp. 1-12, EJSDC 2011.
- [3] M.W. Trywell. Evaluation of the Agricultural Information Service (AIS) in Lesotho, Int J of Info Mgt, pp. 4-11, 2010.
- [4] K. A. Thomas and A. A. Laseinde, Training needs assessment on the use of social media among extension agents in Oyo State, Nigeria, journal of agricultural informatics, Hungarian Association of agricultural informatics, European federation for information technology in agriculture, food and the environment, (ISSN 2061-862X) vol. 6, No. 1, pp. 100-111, 14th January, 2015.
- [5] C. Sanga, M.Mussa, S. Tumbo, M.R.S. Mlozi, L. Muhiche, and R. Haug, Development of the mobile based agricultural extension system in Tanzania: a technological perspective, Int. J of Computing and ICT Research, vol.8 No.1, pp. 49-67, June, 2014.
- [6] J. U. Agbamu, Agricultural research-extension linkage systems: An international perspective. Odi Agricultural Research and Extension Network, Network paper No.106, 2010.
- [7] D. Derso and E. Ejiro, The contribution of information and communication technologies to the Ethiopian agricultural extension system: a review of literature on agriculture knowledge management, African J of Agri. Sci. and Tech. (AJAST), vol. 3, No. 9, pp. 407-411, September, 2015.
- [8] A. Martin, B. Pound, Agricultural Extension, Advisory Services and Innovation, Natural Resources Institute, University of Greenwich, pp.1-12, 14th June, 2014.
- [9] I. Itunuoluwa, O. Jelili and O. Olufunke, Design and Implementation of Text To Speech Conversion for Visually Impaired People, International Journal of Applied Information Systems (IJ AIS), Foundation of Computer Science FCS, New York, USA vol (7) 2, pp 25-30, April 2014, www.ijais.org.
- [10] A. Krasimir, B. Bjorn B and R. Aarne, Speech-Enabled Hybrid Multilingual Translation for Mobile Devices. Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics, pp 41-44, Gothenburg, Sweden, April 26-30 2014.
- [11] D. Suendermann, H. Höge, and A. Black, Challenges in Speech Synthesis. Chen, F., Jokinen, K., (eds.), Speech Technology, Springer Science + Business Media LLC. 2010.
- [12] J. Allen, M. S Hunnicutt and D. Klatt D, From Text to Speech: The MITalk system. Cambridge University Press, 1987

- [13] P. Rubin, T. Baer and P. Mermelstein, An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America* vol. 70, pp 321-328, 1981.
- [14] J.P.H.V. Santen, J.P.H., Sproat, R. W., Olive, J.P., and J. Hirschberg, *Progress in Speech Synthesis*, Springer, 1997.
- [15] J.P.H.V. Santen, Assignment of segmental duration in text-to-speech synthesis. *Computer Speech & Language*, vol. (8) 2, pp 95-128, 1994.
- [16] J. Van Santen, "Assignment of segmental duration in text-to-speech synthesis", *Computer Speech and Language*, 8 (2), pp 95-128, April 1994.
- [17] E. D. Kristin. Agriculture and Climate Change: An Agenda for Negotiation in Copenhagen, The Important Role of Extension Systems, Vision 2020, Food for Agriculture and Environment, International Food Policy Research Institute, USA, pp. 1-2, May, 2009.
- [18] USAID-INDIA-Afghanistan Feed The Future India Triangular Training (FTF ITT) Programme on "e-Extension: ICT applications in agricultural extension management" Ministry of Agriculture, Irrigation and Livestock (MAIL) Kabul, Islamic Republic of Afghanistan, pp. 1-34, 10th–19th October 2016.
- [19] Final Report: FGN-AETA, Agricultural Extension Transformation Agenda, pp. 1-62, 2009.
- [20] T. Arokoyo, ICTs application in agricultural extension services delivery, Agric Ext. in Nigeria. Ilorin: Agricultural Extension Society of Nigeria. pp. 245-25, 2005.
- [21] M. B. Abdullahi, I. S. Shehu, and Y. M. Sani, Multilingual Translation System for Enhancing Agricultural e-Extension Services Delivery, Int Conf. on Info and Com Tech and Its Applications (ICTA) Federal University of Technology, Minna, Nigeria, pp. 62-68, November 28th – 30th 2016.
- [22] Text-to-speech technology: In *Linguatec Language Technology Website*. Retrieved February 21, 2014, from <http://www.linguatec.net/products/ts/information/technology>
- [23] Gilorien. DHTML and JavaScript, Prentice Hall PTR, Upper Saddle River, NJ 07458; pp 1-1164, [www.phptr.com](http://www.phptr.com).