AGRICULTURAL E-EXTENSION SERVICES: A HYBRID OF MULTILINGUAL TRANSLATION TEXT-TO-SPEECH - A FRAMEWORK

Article *in* i-manager's Journal on Pattern Recognition · January 2018

DOI: 10.26634/jpr.5.3.15679

Yahaya Sani

Federal University of Technology Minna

Ilyasu Anda

Federal University of Technology Minna

AGRICULTURAL E-EXTENSION SERVICES: A HYBRID OF MULTILINGUAL TRANSLATION TEXT-TO-SPEECH - A FRAMEWORK

By

YAHAYA MOHAMMED SANI * STELLA OLUYEMI ETUK ** ILYASU ANDA ***

MAMMAN ADAMU ****

** Assistant Lecturer, Department of Information and Media Technology, Federal University of Technology, Minna, Nigeria.*
*** Lecturer, Department of Information and Media Technology, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.*
**** Lecturer, Department of Library and Information Technology, Federal University of Technology, Minna, Nigeria.*
***** Research Scholar, Department of Computer Science, Federal University of Technology, Minna, Nigeria.*

Date Received: 11/01/2019 Date Revised: 12/02/2019 Date Accepted: 14/03/2019

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, and when agriculture research output is communicated in English language, it may be of little or no use to them, if delivered in a foreign language. Text-to-Speech Enabled Hybrid Multilingual Translation framework adopts a serial integration of NLP on one hand and 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.

INTRODUCTION

Agricultural extension services delivery is a system that facilitates farmers access, their groups, and other market players to information, knowledge and technologies (FAO, 2010) 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 (Kameswari, Kishore, & Gupta 2011). Today's role of extension services involve facilitating technology transfer, helping farmers to establish associations, overcoming market challenge, and partnering with service providers and other relevant agencies (Mokotjo & Kalusopa, 2010). 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 (Thomas & Laseinde, 2015). 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 (Sanga et al., 2014; Agbamu, 2000). Regrettably, in many sub-Saharan African countries, poverty and food insecurity have been identified as some of the underlying cause of under development (Derso & Ejiro, 2015). 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, inappropriate strategies for effective agricultural research and ineffective and inefficient extension practice 6. Partial reporting of extension services across rural regions and problem of embracing technology packages that are community-oriented have been described as major challenges in the effective coverage of extension services (Martin & Pound, 2014). 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 that 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 (Isewon, Oyelade, & Oladipupo, 2014). 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 do not require an internet connection (Angelov, Bringert, & Ranta, 2014). For example:

Query 1:

English: Are you a farmer?

Hausa: Shin kai manomi ne

Yoruba: Se o jẹ agbẹ

Ibo: ị bụ onye ọrụ ugbo

Arabic: هل انت مزارعة

Query 2:

English: What type of crop do you cultivate?

Hausa: wane irin amfanin gona kake noma

Yoruba: Iru irugbin wo ni ose

Ibo: kedu ụdị ihe ubi ị na-azụlite

Arabic: ما نوع المحاصيل التي تزرعها

There is yet no attempt in Nigeria in this area for a suitable approach; however, the authors 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. Therefore, 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.

1. Related Concepts

1.1 An Overview of Speech Synthesis

Speech synthesis can be termed as an artificial creation of human speech (Suendermann, Höge, & Black, 2010). 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 (Allen, Hunnicutt, Klatt, Armstrong, & Pisoni, 1987). Synthesized speech can be produced by concatenating bits of recorded speech that are warehoused in a databank (Figure 1). 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 (Rubin, Baer, & Mermelstein, 1981). 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 enables 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: (Santen, Olive, Sproat, & Hirschberg, 1997) a front-end and a back-end. The front-end has two 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

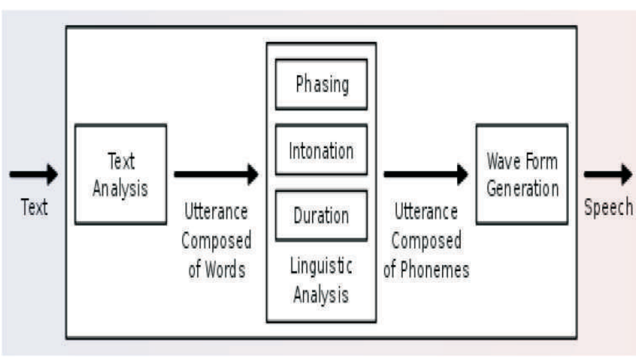

Figure 1. Overview of Text-to-Speech System (Davis, 2009)

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) (Santen, 1994), which is then imposed on the output speech.

1.2 Review of Existing e-Extension Services Delivery Systems

Final report : FGN-AETA (2009) 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.

Sharma, Murthy, and Attaluri (2016) noted the impact of ICT to extension services delivery for horticulture farmers in Uasin Gishu County in the other to ascertain the level of utilization of ICT by horticulture farmers. 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 Arokoyo (2005). 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 Abdullahi, Shenu, and Sani (2016). 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 dialectical problems in to account.

Also, a multilingual translation system to enhance agricultural e-extension services delivery was designed, developed, and implemented by Text-to-speech technology (2014) (Abdullahi, Shehu, & Sani, 2016). 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, the authors 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 into user-understandable native or indigenous languages. This is the gap that this research is set to achieve.

2. Research Methodology

2.1 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 analyzes and then converts 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 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 phonemes ("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 oftentimes 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) (Gilorien, 2004).

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).

2.2 Framework Representation

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

2.2.1 Presentation Tier

This layer is 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).

2.2.2 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 (Gilorien, 2000) Google Translate API for text-to-text component and Android Google Translates API text-to-speech recognizer and synthesizer for text-to-speech component for the following languages: English, Hausa, Yoruba, Ibo, and Arabic. Multilingual translation of text-to-text and text-to-speech equally takes place in this layer.

2.2.3 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 adopt 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.

2.3 Implementation Algorithm for the Framework Design

The proposed framework for agricultural e-extension services: a hybrid multilingual language translation text-to-speech is as shown in Figure 2. The framework links three 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) translates the feeds into English and forwards it to Agricultural Extension Worker (AEW) in text pattern.
- AEW then feeds the web server (Agricultural Knowledge-Based Server) in English.
- The server responds; while sending feed request back to AEW, if there is no response to the query/question from the server, the server then sends the query or question to the researcher for onward research and replies the research response back to the server-AEW-Farmers.
- AEW checks the response and forwards it to MLTE.
- MLTE then translates the feed responses into the farmers' respective registered languages using a

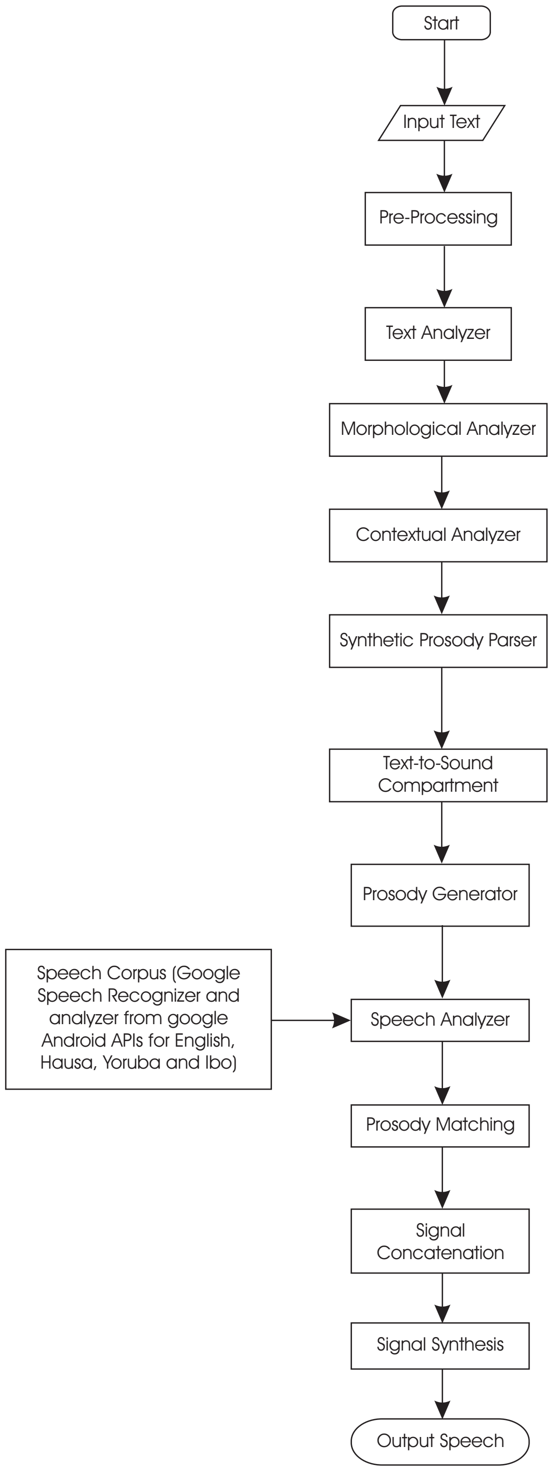

Figure 2. Operational Procedure of Natural Language Processing Component of TTS Synthesizer

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.

3. 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 (Figure 3). Java is used because it is robust and independent platform. The application is divided into two main components: the main application module which includes the basic GUI components that 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 is integrated into the main module 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.

Conclusion

This research work proposed a framework for agricultural

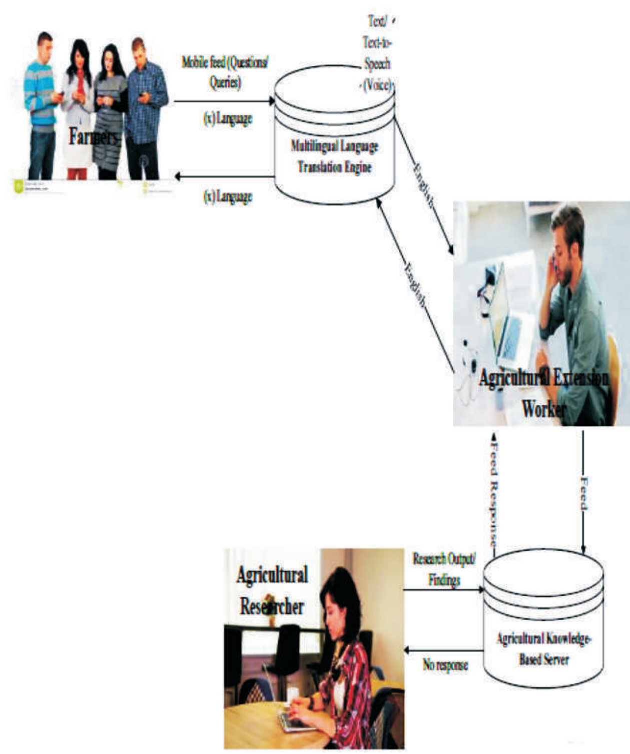

Figure 3. The Proposed Framework

e-extension services: an hybrid of multilingual language translation text-to-speech ensures by the second Agricultural Research Information (ARI) is delivered to agriculturalists regardless of their geographical boundaries and dialect. The proposed framework converts agricultural research finding from a source dialect (English) into four 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]. **Abdullahi, M. B., Shehu, I. S., & Sani, Y. M. (2016).** A Multilingual Translation System for Enhancing Agricultural e-Extension Services Delivery. *Int Conf. on Info and Com Tech and Its Applications (ICTA)* (pp. 62-68).

[2]. **Agbamu, J. U. (2000).** Agricultural research-extension linkage systems: An international perspective. *Network Paper-Agricultural Research and Extension Network,* (106), 1-7.

[3]. **Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., & Pisoni, D. B. (1987).** *From Text to Speech: The MITalk System.* Cambridge University Press.

[4]. **Angelov, K., Bringert, B., & Ranta, A. (2014).** Speech-enabled hybrid multilingual translation for mobile devices. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 41-44).

[5]. **Arokoyo, T. (2005).** *ICTs application in agricultural extension service delivery.* Agricultural Extension in Nigeria. Ilorin: Agricultural Extension Society of Nigeria, 245-251.

[6]. **Davis, K. E. (2009).** *Agriculture and climate change: An agenda for negotiation in Copenhagen,* The Important Role of Extension Systems, Version 2020 (pp.1-2). Food for Agriculture and Environment International Food Policy Research Institute, USA.

[7]. **Derso, D., & Ejiro, E. (2015).** The contribution of information and communication technologies to the Ethiopian agricultural extension system: A review of literature on agriculture knowledge management. *African Journal of Agricultural Science and Technology,* 3(9), 407-411.

[8]. **FAO. (2010).** *Agricultural Extension Services Delivery System in Nepal.* Food and Agriculture Organization of the United Nations UN Complex, Pulchowk, Nepal. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/AgExtServDelSysNepal.pdf

[9]. **Final Report: FGN-AETA. (2009).** *The Agricultural Extension Transformation Agenda.* Retrieved from http://unaab.edu.ng/wp-content/uploads/2012/10/Final%20Extension%20Report.pdf

[10]. **Gilorien. (2000).** *DHTML and JavaScript.* Upper Saddle River: Prentice Hall PTR.

[11]. **Isewon, I., Oyelade, O. J., & Oladipupo, O. O. (2014).** Design and implementation of text to speech conversion for visually impaired people. *International Journal of Applied Information Systems,* 7(2), 25-30.

[12]. **Kameswari, V., Kishore, D., & Gupta, V. (2011).** ICTs for Agricultural extension: A study in the Indian Himalayan region. *The Electronic Journal of Information Systems in Developing Countries,* 48, 1-12. doi:10.1002/j.1681-4835.2011.tb00339.x

[13]. **Martin, A., & Pound, B. (2014).** *Agricultural Extension, Advisory Services and Innovation.* Natural Resources Institute, University of Greenwich. Retrieved from https://www.nri.org/publications/thematic-papers/7-agricultural-extension-advisory-services-and-innovation/file

[14]. **Mokotjo, W., & Kalusopa, T. (2010).** Evaluation of the Agricultural Information Service (AIS) in Lesotho. *International Journal of Information Management,* 30(4),

350-356.

[15]. Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *The Journal of the Acoustical Society of America,* 70(2), 321-328.

[16]. Sanga, C., Mussa, M., Tumbo, S., Mlozi, M. R. S., Muhiche, L., & Haug, R. (2014). On the development of the mobile based agricultural extension system in tanzania: A technological perspective. *International Journal of Computing & ICT Research,* 8(1), 49-67.

[17]. Santen, J. P. H. V. (1994). Assignment of segmental duration in text-to-speech synthesis. *Computer Speech & Language,* 8(2), 95-128.

[18]. Santen, J. P. H. V., Olive, J. P., Sproat, R., W., & Hirschberg, J. (1997). *Progress in Speech Synthesis.* Springer.

[19]. Sharma, V. P., Murthy, L., & Attaluri, S. (2016). e-Extension: ICT Applications in Agricultural Extension Management. *USAID-INDIA-Afghanistan Feed The Future India Triangular Training (FTF ITT) Programme,* 2016. Retrieved from http://www.manage.gov.in/ftf-itt/prgReports/afgan.pdf

[20]. Suendermann, D. Höge, H., & Black, A. (2010). Challenges in Speech Synthesis. In Chen, F., Jokinen, K., (Eds.), *Speech Technology* (pp. 19-32). Springer Science + Business Media LLC.

[21]. Text-to-speech Technology. (2014). In Linguatec Language Technology Website. Retrieved from http://www.linguatec.net/products/tts/information/technology

[22]. Thomas, K. A., & Laseinde, A. A. (2015). Training needs assessment on the use of social media among extension agents in Oyo State, Nigeria. *Agrárinformatika/ Journal of Agricultural Informatics,* 6(1), 110-111.

ABOUT THE AUTHORS

Yahaya Mohammed Sani is currently an Assistant Lecturer in the Department of Information and Media Technology at the Federal University of Technology, Minna, Nigeria. He obtained his B.Tech Degree in Mathematics/Computer Science in 2007, MTech Degree in Computer Science in 2016 from the same University. His research interests are in Information System, Big Data Analytics and Neural Computing.

Stella Oluyemi Etuk is currently working as a Lecturer in the Department of Information and Media Technology at School of Information and Communication Technology, Federal University of Technology, Minna, Niger State, Nigeria. Her first degree (B.Tech) was in Mathematics and Computer Science in 2008 and her second degree (M.Tech) was in Applied Mathematics (Numerical Analysis) in 2014 from the same University. Her future research interests include System Modelling and Data Science. Her administrative responsibilities include student advising and membership of committee at faculty level in the University.

Ilyasu Anda is a Lecturer II in the Department of Library and Information Technology at the Federal University of Technology, Minna, Nigeria. He obtained B.Tech in Library and Information Technology from the same University in 2008, he was also awarded a Master of Science Degree in Information Technology from University of York, United Kingdom in 2015. His research interests are in Big Data Analytic. Ilyasu Anda is a Professional member of the Cyber Security Experts association of Nigeria.

Mamman Adamu holds a B.Tech in Mathematics/ Computer Science and M.Tech in Computer Science both from Federal University of Techology, Minna, Niger State of Nigeria. He is currently a Ph.D student in the Department of Computer Science, Federal University of Technology, Minna, Nigeria. His research interests include Wireless Sensor Network.