



The Acceptability and Comparative Study of DTMF and Speech for Banking Transactions in Nigeria

Oyelami Olufemi Moses^{1*} and Akinyemi Ibidapo Olawole²

¹Department of Computer and Information Sciences, Covenant University, Ota, Nigeria.

²Department of Computer and Information Sciences, Covenant University, Ota, Nigeria.

Original Research Article

Received: 27 December 2013

Accepted: 01 March 2014

Published: 13 March 2014

Abstract

Aims: This article reports the development and the experimental results of the acceptability testing and comparison of speech and dual tone multi frequency (DTMF) for interacting with a speech application for carrying out banking transactions.

Study Design: A within-subjects experiment where each of the participants tested both the DTMF only and speech only modalities was carried out. Each one filled a questionnaire after testing each modality

Place and Duration of Study: Department of Computer and Information Sciences, Covenant University, Ota, Nigeria, between May 2013 and June 2013.

Methodology: Voice Objects 11 was used as the development platform. Voxeo Prophecy 9 was used as the implementation platform while MySQL 5.5 was used as the database management system for storing each customer's account details. X-Lite soft phone 3.3 was used for testing the system. 50 undergraduates of Covenant University, Ota, Nigeria participated in the evaluation. Each one tested the system using DTMF and speech separately. After each round of the tests, each one filled a questionnaire for the modality tested in a bid to measure the acceptability and user satisfaction with the modality.

Results: The entire system's satisfaction ratings for DTMF were significantly higher ($M=36.18$) than for speech ($M=34.76$), $t(49)=-1.46$, $t_{crit}=2.0$, $P=.05$. For modality evaluation, speech was more satisfying ($M=12.22$) than DTMF ($M=12.18$), $t(49)=0.14$, $t_{crit}=2.0$, $P=.05$. For modality entertainment, both speech and DTMF were rated equal ($M=13.94$), $t(49)=0$, $t_{crit}=2.0$, $P=.05$. For modality naturalness, speech was more natural ($M=10.2$) than DTMF ($M=9.44$), $t(49)=-2.41$, $t_{crit}=2.0$, $P=.05$. 52% of the subjects chose DTMF modality, whereas 48% chose speech ($p=.05$).

Conclusion: Dialogue systems are a widely acceptable technology for carrying out banking transactions in Nigeria. It will provide a cost-effective and easily accessible means of carrying out banking transactions.

Keywords: Speech, DTMF, banking transaction, satisfaction, naturalness, entertaining, automated speech recognition, text-to-speech.

*Corresponding author: olufemioyelami@gmail.com, femi.oyelami@covenantuniversity.edu.ng;

1 Introduction

Speech technology incorporating Automated Speech Recognition (ASR) and Text-to-Speech (TTS) enables humans to interact with electronic devices through human language. It is the most human and benign of all technologies. ASR, also referred to as Speech Recognition, is a technology that enables computers, mobile phones and other software-enabled electronic devices to recognize and act upon spoken input in human (natural) language. TTS technology gives these devices the capacity to output their responses to human users in natural language [1]. Speech technology can be used to bridge the digital divide [1] and is also capable of providing widespread access to services by the people, by exploiting the ubiquitous and widespread availability of mobile phones. It also has the advantage of being usable by the non-literates in Africa and suitable for the visually impaired. Furthermore, the dual tone multi frequency (DTMF) part of the technology is useful for people with speech disorder since it gives them the option of responding to speech applications by pressing the keys on the telephone keypad instead of a spoken response. It can be implemented in the various indigenous languages of Africa and can provide eyes-free interaction. It can also serve as a great aid for the physically challenged [2].

Banking transactions in Nigeria have become electronic to the extent that customers do not have to enter the banking halls to make withdrawals, make deposits and check their accounts' balances, since these can be done by making use of the Automatic Teller Machines (ATMs), computers (using the Internet) and mobile phones (using short message service (SMS)). However, there is no single solution for carrying out banking transactions that is speech technology-based in spite of the great benefits like cost effectiveness, widespread accessibility, narrowing of the digital divide for the illiterates and being usable to provide services in different languages in a country like Nigeria with diverse tongues. There is therefore a need to know the suitability of this technology for transacting with banks and its acceptability. This paper reports the development and experimental evaluation of a speech-based system for carrying out basic banking transactions using both DTMF and speech modalities.

2. Related Work

DTMF and speech was compared in [3] for interacting with health dialogue expert system for diagnosing diseases among participants in some selected places in Nigeria. The authors found that DTMF was more satisfying than speech for system satisfaction. Modality wise, speech was more satisfying than DTMF. Speech was also more natural than DTMF and the two modalities were equally entertaining. DTMF was preferred by the majority and was more effective and efficient than speech. For diseases diagnosis expert health dialogue systems in Nigeria, DTMF was recommended for effectiveness and efficiency. It was also recommended for satisfaction. Speech was recommended for modality satisfaction while both modalities were recommended for entertainment purpose. Speech was advocated for modality naturalness. They concluded, however, that a platform that incorporates the two modalities would provide the benefits of the two, and allow the users varieties of choices that best suit their needs [3].

The acceptability and user satisfaction of using dialogue systems for detecting counterfeit drugs in Nigeria was developed and tested in [4]. The results showed that the users were satisfied with the system and the system was widely accepted as a means of fighting counterfeited drugs.

Similarly, touchtone (DTMF) and speech input modalities in the context of low literacy users and a health information service for the provision of health information for caregivers of HIV positive children in Botswana, Southern Africa was compared in [5]. Their hypotheses were: a. DTMF is likely to be more acceptable in the developing world because general numeracy is less common; and b. DTMF-based systems are much easier to develop than natural language systems and are consequently more attainable in the resource-constrained environments that typically characterize the developing world. The experiment was a within-subjects and a post-study questionnaire was administered verbally to each of the subjects. The results obtained showed that there were no significant differences between task completion rates even though speech performed a bit better than DTMF. 59% of the users preferred DTMF while 19% preferred speech. This contrasted the studies carried out in the developed world where users preferred speech. However, the results correlate with the fact that simple DTMF is generally not viewed as favourable. The participants that preferred speech did not comment as in the studies in the developed world that speech is more entertaining or enjoyable, rather, they commented on the utility of speech as being more accessible for older people or faster [5].

Kwan and Jennifer reported “an experiment that critically tested user’s preference for an input modality (speech vs. DTMF) in a phone-based message retrieval system using a fully functioning natural language system. The experiment was a within-subjects design. Sixteen subjects (8 women and 8 men) from the IBM T. J. Watson Research Center in New York tested the system. Their results indicated that “(a) DTMF was more effective and efficient for linear tasks, whereas speech was better for nonlinear tasks; (b) speech was preferred to DTMF by a majority of users; (c) speech was judged as being more satisfying, more entertaining and easier to use than DTMF; and (d) user preference for a particular modality was better predicted by user performance in nonlinear tasks rather than linear ones” [6].

Patel et al. compared speech and DTMF for farmers in Gujarat, India in a non-within subjects experiment. They “designed Avaaj Otalo (“voice-based community forum”), a Gujarati language application allowing farmers to access agricultural information over the phone. 45 participants, mainly farmers, recruited from ten districts in rural parts of Gujarat tested the system. “Input modality (speech vs. DTMF) was randomly assigned to each user, but was anonymously corrected to maintain balance across age, education and gender. A post-test questionnaire with Likert scales was administered to measure user satisfaction, ease of use and learnability” [4]. The results showed that “task completion rate with DTMF was significantly higher than with speech. No significant difference in user satisfaction. Over 80% of users reported that they found it easy to access information from the system in both groups. Over 75% also of both groups reported that they would “definitely” use such an application if it was made available. Across all tasks, the percentage of responses that rated the difficulty level of a particular task as either “difficult” or “very difficult” on a five-point Likert scale was 49% for speech and 30% for DTMF. For difficulty faced when providing input to the system, 81% of DTMF users answered “no” or “definitely no”, compared to 38% for speech users ($P < 0.01$)” [7].

3. Methodology

The prototype system was developed using Voice Objects Desktop for Eclipse 11 as the development platform. Voxeo Prophecy 9 was used as the implementation platform and MySQL 5.5 was used as the relational database management system for storing each customer's account details. X-Lite soft phone 3.3 was used to call the system for testing.

50 participants, mainly undergraduates of Covenant University, Ota, Nigeria participated in the within-subjects study. The participants are from different ethnic groups in Nigeria. Each one tested the system using DTMF and speech separately. After each round of test, each one was given a questionnaire to fill for the modality tested in a bid to measure the acceptability and user satisfaction with the modality. The questionnaire contained questions on modality evaluation and the system evaluation. They were extracted from similar studies by Kwan and Jennifer, and Marilyn et al. [6, 8, 9, 10]. “The measures used in the questionnaire have both face and content validities. In terms of face validity, all the measures were constructed by experts with over 10 years of experience in usability tests of mobile and speech user interface (SUI) applications. In terms of content validity, the measures cover all dimensions of usability in telephony applications as defined by European Telecommunications Standard Institute (ETSI)” [6]. The questionnaire was based on Likert scales of 1 (strongly disagree) to 5 (strongly agree). The system was tested on a laptop PC running Windows operating system as the application could not be hosted since companies that host this kind of application are non-existent in Nigeria. However, the tools used for development allow for testing locally.

3.1 Functionalities Provided by the Prototype System

The prototype system provides the functionalities below for users:

- i. Logging in using account number and a PIN that serves as a password.
- ii. Checking of the balance in an account.
- iii. Transferring of fund from one account to another.
- iv. Creating of a new PIN.
- v. Changing of PIN.
- vi. Speaking to an agent.

3.2 Snapshots of the Prototype System

Fig. 1 below shows the dialogue flow for a user to log in, using his account number and a PIN. Once the user has been authenticated, the menu options are presented for him to choose which transaction to be carried out.

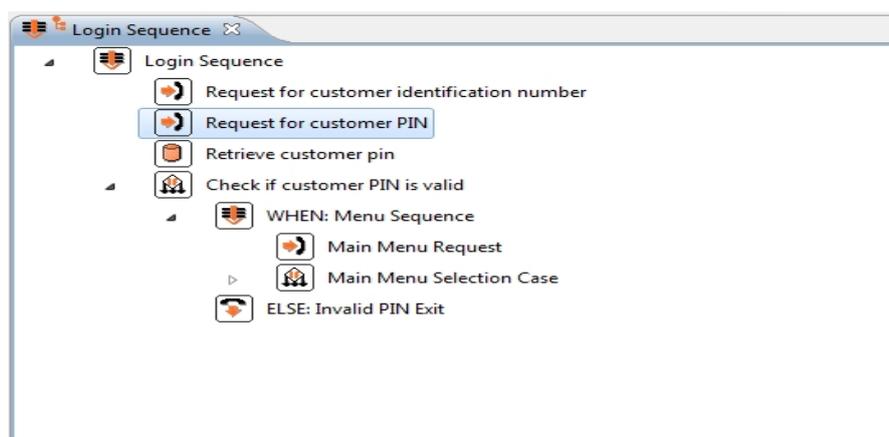


Fig. 1. Dialogue Flow for Logging in Into the System

Fig. 2 below shows the dialogue flow for checking the account balance. Once a customer has been authenticated to use the system and chooses to check the balance in his account from the menu options, the system goes into the database and retrieves the balance in the account and relays it to the customer.

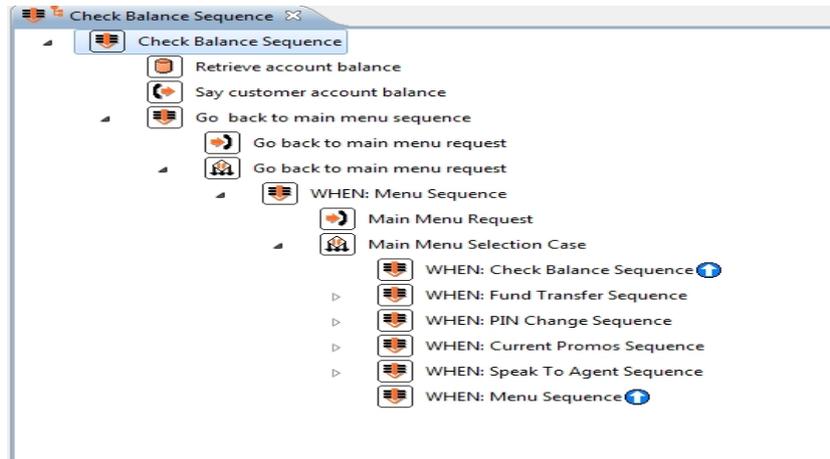


Fig. 2. Dialogue Flow for Checking an Account Balance

Fig. 3 shows the dialogue flow for the customer to listen to the current promotion in the bank. Once a customer picks this option, the current promotion information is relayed to him.

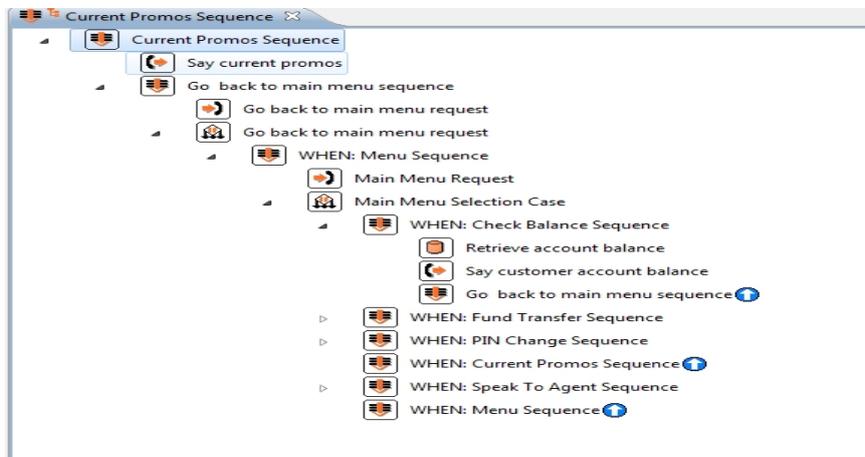


Fig. 3. Dialogue Flow for Current Promotion

Fig. 4 shows the dialogue flow for fund transfer menu option. The customer is expected to supply the account number to which fund is to be transferred. Once this is done, the system checks the database to confirm the number. If the number is valid, the user is prompted to supply the amount to be transferred. If the account has enough funds, the transfer is carried out.

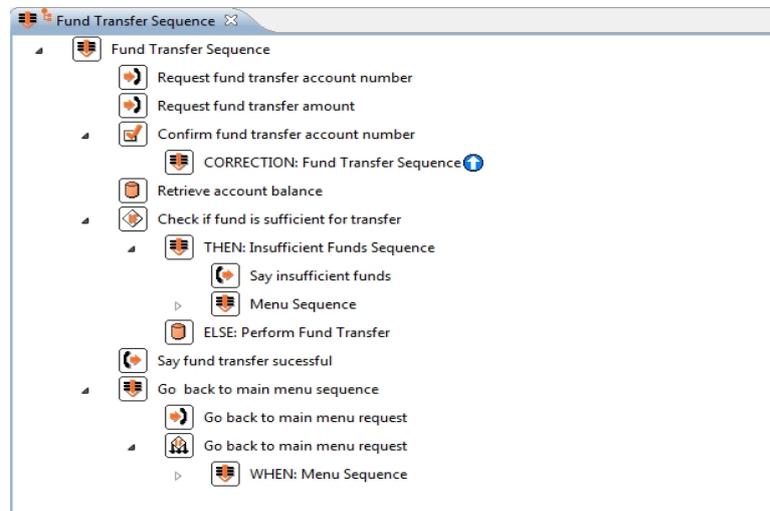


Fig. 4. Fund Transfer Dialogue Flow

Fig. 5 shows the dialogue flow for changing the PIN. The system will request the customer to enter a new PIN. Once this has been successfully confirmed, the PIN entered is played back to the customer. The new PIN entered may also be denied.

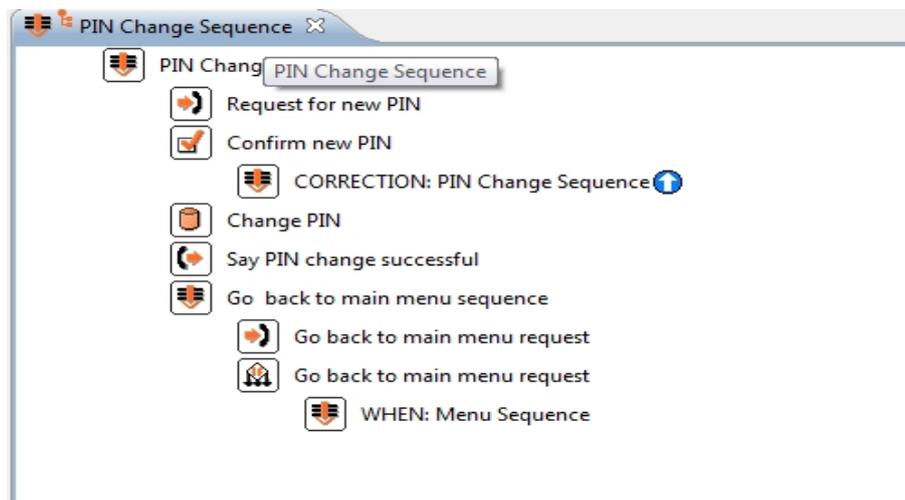


Fig. 5. Dialogue Flow for Changing PIN

Fig. 6 shows the dialogue flow that transfers call to a customer care representative. Once a caller chooses this option, the system checks to see if the representative is available or not. If available, call is transferred to him. However, because the system was tested as a standalone application, this option was not tested.

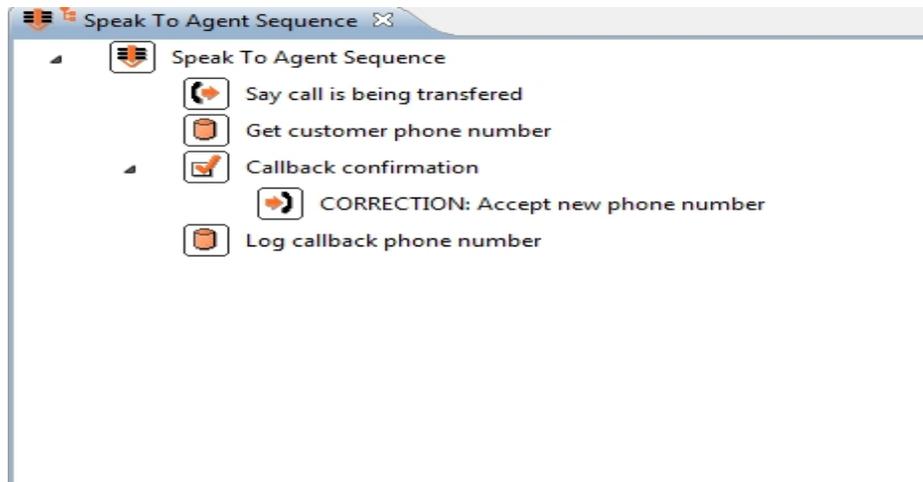


Fig. 6. Dialogue Flow for Call Transfer

Fig. 7 shows how the system was called with X-Lite soft phone for testing.



Fig. 7. Testing the Application with X-Lite

4. Results and Discussion

4.1 System Evaluation

The null hypothesis is that the mean difference between DTMF and speech user satisfaction is zero. The alternative hypothesis is that there is a mean difference between the two input modalities.

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

Satisfaction ratings for DTMF were significantly higher ($M=36.18$) than for speech ($M=34.76$) as indicated by a significant t-test, $t(49) = -1.46$, $t_{crit} = 2.0$, $P = .05$. The null hypothesis is thus rejected, and the finding indicates that DTMF was more satisfying than speech. This is in tandem with the results obtained in [3]. The implication of this is that among the user population - Nigerians, DTMF seems to always produce a better user satisfaction than speech. This could be as a result of inaccurate recognition of different accents with which the people speak by the implementation platform - Prophecy.

4.2 Modality Evaluation

For modality satisfaction, the null hypothesis is that the mean difference between DTMF and speech modalities satisfaction is zero. The alternative hypothesis is that there is a mean difference between the two input modalities.

$$\begin{aligned} H_0: \mu_d &= 0 \\ H_1: \mu_d &\neq 0 \end{aligned}$$

The subjects evaluated their interaction with speech as more satisfying ($M=12.22$) than DTMF ($M=12.18$) as indicated by a t-test, $t(49) = 0.14$, $t_{crit} = 2.0$, $P = .05$, though the two systems did not differ significantly. The alternative hypothesis is rejected, thus, the two modalities are equally satisfying. This correlates with the results obtained in [3].

For modality entertainment, the null hypothesis is that the mean difference between DTMF and speech modalities entertainment is zero. The alternative hypothesis is that there is a mean difference between the two input modalities.

$$\begin{aligned} H_0: \mu_d &= 0 \\ H_1: \mu_d &\neq 0 \end{aligned}$$

For modality entertainment, the results were interesting; both speech and DTMF were rated equal ($M=13.94$) as indicated by a t-test, $t(49) = 0$, $t_{crit} = 2.0$, $P = .05$. The alternative hypothesis is thus rejected. Hence, the results show that both speech and DTMF are equally entertaining. Thus there is a correlation with the results obtained in [3]. The implication of the result is that any of the two modalities could be used by the user population to command the best entertaining result for dialogue systems meant for carrying out banking transactions.

Finally, for modality naturalness, the null hypothesis is that the mean difference between DTMF and speech modalities naturalness is zero. The alternative hypothesis is that there is a mean difference between the two input modalities.

$$\begin{aligned} H_0: \mu_d &= 0 \\ H_1: \mu_d &\neq 0 \end{aligned}$$

The subjects evaluated their interaction with speech as more natural ($M=10.2$) than DTMF ($M=9.44$) as indicated by a t-test, $t(49) = -2.41$, $t_{crit} = 2.0$, $P = .05$. The null hypothesis is therefore rejected. Thus, speech is more natural than DTMF. This correlates with the results obtained in [3]. Also, this implies that the user population believes that speech is more natural for them to interact with dialogue systems. This therefore means that implementation platforms that specifically take

into consideration, accents of the user population will be of tremendous benefits in Nigeria. It also means that the users may probably prefer to interact with dialogue systems via speech if recognition of their accents is high.

4.3 Modality Preference

In response to the question **which of speech and DTMF do you prefer to interact with the system?** 52 % of the subjects chose DTMF modality, whereas 48 % chose speech ($P = .05$). Thus, there is a correlation with the results obtained in [3]. This generally implies that majority of the user population in Nigeria will prefer DTMF to speech as a means of interacting with dialogue systems.

4.3.1 Experiences with mobile and computing devices

0% of the subjects rated themselves as **novice** in the use of computer software. 12% rated themselves as **expert**. 68% rated their skill as being **good**, while 20% rated their skill as being **average**. These results show that those who are already familiar with the use of software will find the system usable. In the use of devices for enhancing work, 0% replied that they do not use technology to enhance their work, 2% use Desktop PC, 2% use Tablet PC, 48% use laptop/notebook to enhance their work, while 48% use personal digital assistant (PDA)/cell phone. These statistics show that since the majority of the respondents make use of laptops/notebooks and PDAs/cell phones, these categories of people will also find it easy to adopt dialogue systems for carrying out banking transactions and that those who are already using electronic devices to enhance their work will generally not have difficulty in adopting dialogue systems for their banking transactions. All the participants reported that they owned a mobile phone or a PDA. 56% have owned a mobile phone/a PDA for more than 2 years and 44% for 2 years. In response to the number of times they make or receive calls a week, 0% does not make and receive calls. 0% makes or receives calls 1-2 times a week, 38% 3-4 times a week, 54% 5-6 times a week and 6% more than 7 times a week. 2% did not specify the number of times they make or receive calls. Lastly, 100% support the use of mobile devices for banking. These results also imply that

5. Conclusion

This article has reported the development of a speech-based system for carrying out banking transactions and a comparative study of interacting with the system using both DTMF and speech. The results indicate that the system will be found useful when deployed in real life in the country (Nigeria). Prospective users of systems of this nature who already have been using electronic devices to enhance their work, and who are familiar with the use of software will have less difficulty in adopting this technology. For modality satisfaction, any of DTMF and speech could be used to achieve the best result. The same applies to entertainment purpose, but speech modality is recommended for the best result for naturalness. For systems of this nature, DTMF is recommended for best overall user satisfaction. The technology could be exploited to provide widespread, cost-effective, anywhere and anytime access to banking transactions in Nigeria. The cost of using this system is comparable with the cost of making calls or sending an SMS which is always affordable by the general populace. It can also be implemented in the various indigenous languages to mitigate the effects of illiteracy on technology usage. Because banking transactions are targets for hackers, the use of voice biometrics and authentication via password are recommended for systems of this nature for security purpose.

Acknowledgements

The authors appreciate Damilola Oluwafayokemi Idowu for her contributions towards the development of this system, as well as helping in administering the questionnaires used.

Competing Interests

Authors have declared that no competing interests exist.

Authors' contributions

Author OOM conceived and designed the study, carried out the literature review, wrote the first draft of the manuscript and interpreted the data. He also carried out data analysis. Author AIO revised the manuscript and provided contributions in the final write-up. All the authors have read and approved the final manuscript.

References

- [1] Rondel S, Pattabhiraman PT, Ganapathiraju A., Khademi P, Rondel J. Strategic Importance of Speech Technology for NGNs. White Paper, Conversay 15375 NE 90th Street, Redmond, WA, 98052, USA.
- [2] Oyelami MO. Dialogue Systems as a Means of Detecting Counterfeited Pharmaceuticals/Drugs by Consumers. In the 7th International Conference on ICT Applications (AICTTRA 2012), National Defence College, Abuja, Nigeria, September 2012.
- [3] Oyelami OM, Uwadia CO, Akinwale AT, Akinyemi IO. An Experimental Comparison of Speech and DTMF for Voice XML-Based Expert Systems. *Afr J. of Comp & ICTs*. 2013;6(1):87-94.
- [4] Oyelami OM. User Satisfaction and Acceptability of Dialogue System for Detecting Counterfeit Drugs. In International Conference on Human-Computer Interaction Amsterdam, The Netherlands May 15-16, 2013.
- [5] Aditi SG, Madelaine P, Etienne B, Christiaan K. HIV Health Information Access Using Spoken Dialogue Systems: Touchtone vs Speech. In 3rd International Conference on Information and Communication Technologies and Development, Doha, Qatar; 2009.
- [6] Kwan ML, Jennifer L. Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation. *International Journal of Human-Computer Interaction*. 2005;19;343-360.

- [7] Patel N, Agarwal S, Rajput N, Nanavati A, Dave P, Parikh TS. A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India. In CHI 2009, Boston, MA, USA, April 2009.
- [8] Walker M, Litman D, Kamm C. Evaluating spoken language systems. In American Voice Input/Output Society (AVIOS), May 1999.
- [9] Marilyn W, Diane L, Candace K, Alicia A. PARADISE: A framework for evaluating spoken dialogue agents. In 35th Annual Meeting of the Association of Computational Linguistics (ACL 97), July 1997.
- [10] Walker MA, Kamm CK, Litman D. Towards Developing General Models of Usability With PARADISE. *Natural Language Engineering*. 1998;1:1-13.

© 2014 Moses and Olawole; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

www.sciencedomain.org/review-history.php?iid=448&id=6&aid=3972