USING HIDDEN-MARKOV MODEL IN SPEECH-BASED EDUCATION SYSTEM FOR THE VISUALLY IMPAIRED LEARNER

Ambrose A. Azeta, Charles K. Ayo, Nicholas A. Omoregbe, Victor C. Osamor

Department of Computer and Information Sciences, Covenant University (NIGERIA)

Abstract

Speech-based e-Education technology allows users to access learning content on the web by dialing a telephone number. Speech-enabled applications, particularly in the domain of education are primarily implemented to cater for the plight of the visually impaired towards addressing the shortcomings of user interface (UI) design of a mobile learner. However, with the increase in learning resources on the web, using telephone to find suitable content has become a painstaking task for speech-based online users to achieve improved performance. The problem of finding suitable content with speech e-learning applications is more difficult when the sight impaired learner is involved. It is convenient to use existing mobile speech-enabled e-learning applications, but it can be exceptionally time-consuming when the user is forced to navigate through several levels of options before finding exact content. The adoption of Hidden markov model (HMM) for interface and interaction design is required to provide easy navigation and adaptation in speech-enabled learning towards solving the problem of finding suitable content. The objective of this study is to provide a design and implementation of a HMM-based speech interactive education. The system will be useful especially for the physically challenged students such as the visually impaired. It also offers an alternative platform of learning for the able-bodied learners.

Keywords: adaptation, case-based reasoning, e-Education, e-learning, hidden markov model and VoiceXML.

1 INTRODUCTION

The increasing number of Learning Management Systems (LMS) for online teaching, quiz, assignment delivery, discussion forum, email, chat, et cetera, means that dynamic educational online services will be needed for efficient management of all educational resources on the web. The process of selection may be easier with the normal users, but for certain category of learners with visual impairment, navigating a Voice User Interface (VUI) for the desired learning content is a strenuous task.

From the foregoing, it becomes obvious that there exists a need to provide a model for improving the navigational ability and content retriever technique of speech-based e-Education applications. This study aims at addressing the needs as identified, and employing the model so obtained for the development of a prototype HMM-based speech interactive education system. The prototype application developed was tested in a school for the blind, and the result of evaluation reported in [1]. The HMM suffices as a reference model for implementing speech applications in the domain of e-Education. The type of e-Education application that was provided through this research included that of teaching, learning and multiple choice examination. The application is helpful for people with physical access difficulties engendered by their visual impairment.

With the increase of web-based learning technologies, using telephone to navigate and find content of web resources has become a painstaking job for speech-based learners to achieve high performance. The biggest disadvantage of speech-based technologies is the rigid structure that they impose on the end user. While it is convenient to use mobile telephony application, it can be exceptionally slow when the user is forced to drill through several layers of options before getting to the destination content [2]. The problem with navigation and content suitability with speech e-learning applications is more difficult for the visually impaired learner to utilize.

The objective of this study is to provide a HMM-based speech interactive education application that will improve the level of system navigational abilities and finding content suitability for the sight impaired learners with the capability to enhance learning using telephone and web interfaces. Micro soft Vosio was used to design the HMM. The tools used for the implementation of the prototype speech application are: VoiceXML, Hypertext Preprocesor (PHP), Apache and MySQL for Voice User Interface (VUI), Web User Interface (WUI), middle-ware and database respectively. The Voxeo speech engine used was provided by [3] for speech application development and testing. Case-based reasoning (CBR) was used to achieve the adaptation component service such as the ability of the system to reason from experience [4].

This paper is organized as follows. In Section 2, below, we provide the literature review. In section 3, we described ‘HMM-based speech interactive education’, a HMM framework and architectural design of the system, as well as implementation results summary. Finally, section 4 concludes the paper.

2 LITERATURE REVIEW

Speech Gateway provides middleware functionality between the users and application/database content. In the speech business value chain, some vendors that provide speech server services include: Voxeo, BeVocal, VoiceGenie, Audium, Clarity, Entervoice, HeyAnita, Microsoft, Plum Voice, Tellme, VoiceObjects, Voxpilot, VoiceShot, Avaya, Cisco Systems, Envox Worldwide, Convergys, Dimension Data, Fluency Voice Technology, Genesys, HP, ICT Group, Intervoice, Nortel, Nuance, RightNow, Tuvox and West Corporation ([5], [6]).

A speech-enabled web-based absentee system was developed in [7]. The application was tested by some students in software engineering; students who intended to miss a class. Users called the VoiceXML telephone number and were led through an automated dialog to record their details, the date and time of the call, the courseID, and the date they would miss class in a database. The system provides record keeping of absentee calls from students, faculty and university staff. Likewise in [2], a V-HELP system was develope. VoiceXML was used to voice enable a section of the website for Computer Science and Engineering (CSE) department to allow the visually impaired students have access to the information on it. The system also allows on the move access.

An m-learning and speech technology was integrated together in [8] to reduce access barriers to a large population of users. An adapted interactive voice response (IVR) and text-to-speech (TTS) platform to allow users to interact with the forum using voice interface was developed. This speech-enabled discussion forum application does not only help normal users to avoid the cumbersome task of typing using small keypads, but also enables people with visual and mobility disabilities to engage in online education. The prototype forum was tested in two different vision impaired schools in Massachusetts with ten users.

A Voice-enabled Interactive Service (VoIS) was proposed in [9] for e-learning. The goal of VoIS platform for an e-learning system is to better meet the needs of students and mobile learning communities via speech recognition technology. The pilot study explored this goal by demonstrating how an interaction tools like discussion forums can be voice-enabled using VoiceXML. The author designed, developed and evaluated a prototype application with blind and visually impaired learners from different organizations. The study in [10] introduced ‘‘Voice Interactive Classroom’’, a software solution that put forward a middleware approach to provide access through visual and auditory interfaces to web-based e-learning.

All the existing approaches discussed above have successfully implemented the integration of visual interfaces to voice dialogues and to provide auditory access to functionalities already present in various learning management systems such as Segui, Moodle and Sakai. The HMM-based interactive education provided in this study, is met to use the HMM technique to enhance navigational abilities and ease search of suitable learning content for the visually impaired in an e-Education platform.

3 HMM-based speech interactive education

3.1 The Hidden Markov Model (HMM) Framework

Hidden Markov Model (HMM) is sometimes referred to as Finite State Automata (FSA). It is a mathematical apparatus for modeling sequence analysis.. A HMM consists of a set of states connected by a number of possible transition paths (see Fig. 1) for a three-state HMM. Each transition path has an associated transition probability labeled by a_{ij}. A symbol is produced on arrival into a state, assuming for now we are working with a discrete observation HMM. Each state in the HMM has associated with it a set of output symbols with its own probability distribution. The notation $b_i(k)$ indicates the observation probability that symbol k is produced given state i.

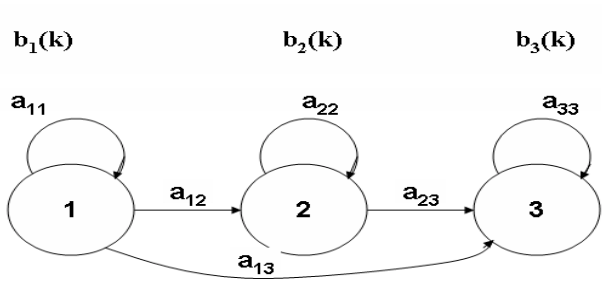

Fig. 1. A three-state HMM. (Source: Burke, 2007)

The following elements are required to fully define HMM [11]: (i) The number of states of the model, *N*, (ii) The number of observation symbols in the a[lphabet, *M*, and (iii) A set of state transition probabilities, satisfying the constraints:

$$\sum_{j=1}^{N} ai_j = 1, \quad 1 \le i \le N$$

where $aij \ge 0, \quad 1 \le i, j \le N$

- A probability distribution in each of the states, $b_j(k)$ satisfying the constraints:

$$\sum_{1}^{m} b_j(k) = 1, \quad 1 \le j \le N$$

where, $B_j(k) \ge 0 \quad 1 \le j \le N, \quad 1 \le k \le M$

- The initial state distribution.

The state transition diagram and probability of *the system* is depicted in Fig. 2 and Table 1 respectively.

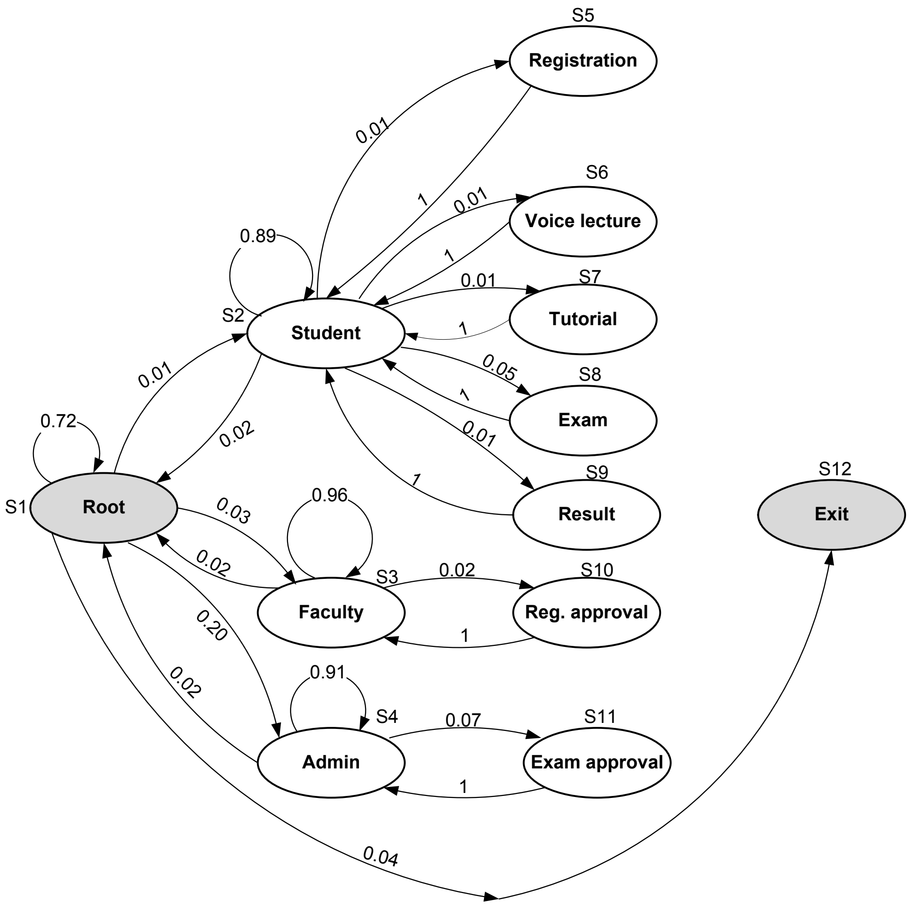

Fig. 2: State transition for HMM-based speech interactive education

The learner is at conversational state S1. Each dialogs determine the next dialogs to transit. Fig. 2 contains the format of state transition probabilities. The probability of transiting to the next state is computed in Equation 1.0. 1.1, 1.2 and 1.3 with the values shown in Table 1 For example, the probability of moving from state S1 to S2, S3, S3 and S4 is 0.20. Transitions are specified using VoiceXML menu item, which define the next document and dialogs to use. If a VoiceXML menu item does not refer to a document, the current document is assumed. The program will end when a dialogs does not specify a successor or when it encounters an exit command.

The variables used in our model are defined as follows (Azeta, 2012): En*i* = total number of edges leaving state *i*, *i* = 1,…,n. Ok = observation,   *k*=1…r. CS = current state. W*j* = weights assigned to alphabets. P*si* = Probability of state i, and i = 1,…,n. P(En*i*) = Probability of En*i.*

$$P(\mathrm{En}i) = \frac{1}{\mathrm{En}i} \qquad (1.0)$$

$$\mathrm{P}si(\mathrm{Ok}) = \frac{\mathrm{P(En}i)}{\mathrm{W}j}, \; i = 1,..,\mathrm{n}. \quad j = 1, \ldots, \mathrm{m}. \qquad (1.1)$$

$$\mathrm{total}_{\mathrm{Ps}} = \sum_{i=1}^{n} \mathrm{P}si(\mathrm{Ok}), \qquad k = 1, \ldots, \mathrm{r} \quad \text{and} \quad i = 1, \ldots, \mathrm{n} \qquad (1.2)$$

$$\mathrm{P}si(\mathrm{CS}) = 1 - \mathrm{totalps} \qquad (1.3)$$

Table 1: State transition probabilities

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**
States	**Root**	**student**	**faculty**	**admin**	**register**	**lecture**	**tutorial**	**exam**	**result**	**reg. approval**	**Exam approval**	**Exit**
S1	0.72	0.01	0.03	0.20	0	0	0	0	0	0	0	0.04
S2	0.02	0.89	0	0	0.01	0.01	0.01	0.05	0.01	0	0	0
S3	0.02	0	0.96	0	0	0	0	0	0	0.02	0	0
S4	0.02	0	0	0.91	0	0	0	0	0	0	0.07	0
S5	0	1	0	0	0	0	0	0	0	0	0	0
S6	0	1	0	0	0	0	0	0	0	0	0	0
S7	0	1	0	0	0	0	0	0	0	0	0	0
S8	0	1	0	0	0	0	0	0	0	0	0	0
S9	0	1	0	0	0	0	0	0	0	0	0	0
S10	0	0	1	0	0	0	0	0	0	0	0	0
S11	0	0	0	1	0	0	0	0	0	0	0	0
S12	0	0	0	0	0	0	0	0	0	0	0	0

3.2 Architectural design of HMM-based speech education

The HMM-based speech education architecture (see Fig. 3) consists of some VoiceXML software suites including Registration, Voice lecture, Tutorial, Examination and Results. The application modules are contained in Intelligent Voice-enabled e-Education System (iVeES). For this prototype implementation, we used free Voxeo hosting service to provide a quick way to get started in order to eliminate huge capital expenses in subscription.

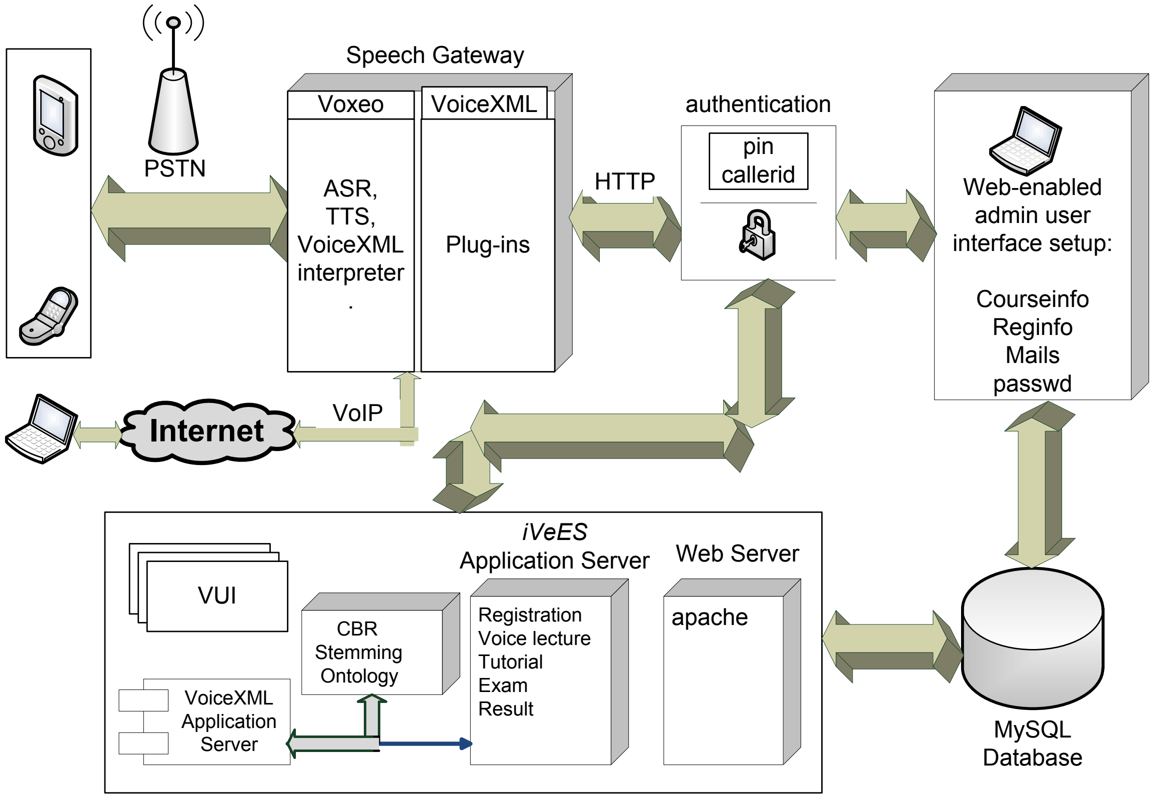

Fig. 3: Architectural design of HMM-based speech education

The architecture shown in Fig. 3 was drawn using modular architectural design. The description of the architecture is presented as follows: The learner connects through a telephone and PSTN to the VoiceXML interpreter through Voxeo Speech gateway, The VoiceXML interpreter executes the call interaction with a caller using the instructions of a VoiceXML script supplied by the voice commands in application server, The interpreter calls TTS and ASR as plugins to complete its tasks, The VoiceXML interpreter communicates via web protocols (HTTP) to VoiceXML application server, The VoiceXML application server delivers the application including VoiceXML text pages, and the web application server queries the database via apache to dynamically retrieve information and the VoiceXML interpreter TTS speaks with the caller.

3.3 Implementation Results Summary

A prototype part of the VoiceXML application (voice user interface) was deployed on a Voxeo voice server on the web and accessed from a mobile phone and land phone using the format:<source country int. dial out #> <destination country code><destination area code><generated voice network 7 digit #>.

Fig. 4 contains screen shot of a sample tutorial question. The faculty or course lecturer select course code, course title and type tutorial question for which answers are sought. The date is automatically selected by the system. It is the responsibility of the faculty to select the search method as either partial of perfect. Partial means displayed answers will include both the exact match and matches that are similar, whereas perfect search showcases answers that are exact

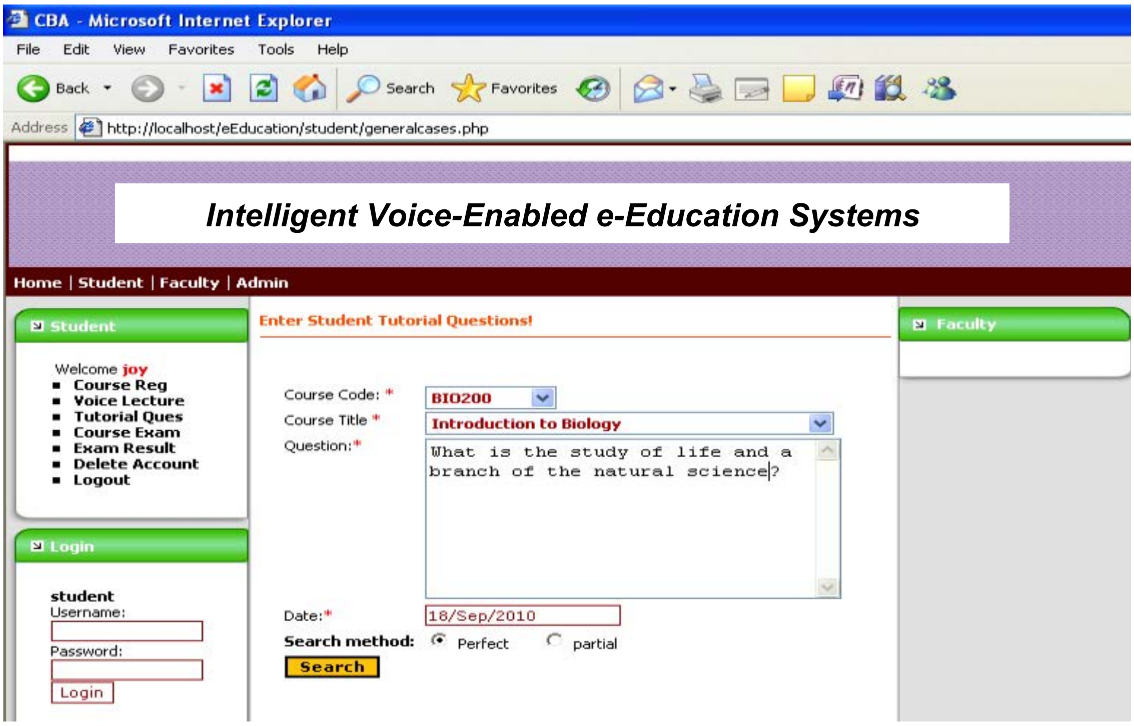

Fig. 4: A screen shot of tutorial questions web interface

4 CONCLUSION

In this paper, a prototype HMM-based e-Education system has been provided and implemented using VoiceXML to proffer solution for the sight impaired students desirous of acquiring further education. The future research direction, for this study is two-fold. First, evaluation of the system using PARAdigm for DIalogue System Evaluation (PARADISE), and second, voice biometric techniques based on speech data will be added as additional means of security mechanism to enhance the authentication of candidates for examination.

REFERENCES

[1] Azeta A. A., Ayo C. K., Atayero A. A. and Ikhu-Omoregbe N. A. (2009) “Application of VoiceXML in e-Learning Systems”, Cases on Successful E-Learning Practices in the Developed and Developing World: Methods for the Global Information Economy. Chapter 7, a Book Published in the United States of America (USA) by Information Science Reference (an imprint of IGI Global). Pp. 92-108. Available online at: http://www.igi-global.com/reference/details.asp?ID=35211&v=tableOfContents

[2] Raghuraman M. B. (2004),”Design and Implementation of V-HELP system – a voice-enabled web application for the visually impaired. A Master’s project presented to the faculty of the graduate college in the University of Nebraska in partial fulfillment of requirements for the degree of Master of Science. Major: Computer Science. August 2004. Retrieved 21 March 2010 from http://hcc.unl.edu/netgroup/alumnidocs/mukesh.MS.pdf

[3] Voxeoprophecy(2003), “The Voxeo Prophecy Platform - Free Download”, Available online at: www.voxeo.com/prophecy

[4] Azeta A. A.(2012), “Developing A Computerized VoiceXML-Based Application For E-Education: Design, Implementation and Evaluation”, Academic Ph.D thesis printed as a textbook by Lambert Academic Publishing.

[5] Datamonitor (2006), “An Introductory Guide To Speech Recognition Solutions”, Understanding the technology, the vendors and the market Industry white paper published Datamonitor. (pp 1-20)

[6] Palowireless (2009), “VoiceXML Development Tools”, Published By: Global industry Analysts Date Published: Jan 2009. Retrieved 5th November 2010 from http://www.palowireless.com/voicexml/devtools.asp

[7] Gallivan P., Hong Q., Jordan L., Li E., Mathew G., Mulyani Y., Visokey P. and Tappert C., (2002), “VoiceXML Absentee System, Proceedings of MASPLAS'02. The Mid-Atlantic Student Workshop on Programming Languages and Systems Pace University. Retrived online 10th January 2010 from http://csis.pace.edu/csis/masplas/p10.pdf

[8] Motiwalla, L.F. and Qin J. (2007),” Enhancing Mobile Learning Using Speech Recognition Technologies: A Case Study”. Eighth World Congress on the Management of eBusiness, 2007. (WCMeB 2007). Volume , Issue , 11-13 July 2007 Page(s):18 – 18

[9] Motiwalla L. F. (2009), “A Voice-enabled Interactive Services (VòIS) Architecture for e-Learning”, International Journal on Advances in Life Sciences, vol 1 no 4, year 2009 Page 122 – 133. Retreived 30th October 2010 from http://www.iariajournals.org/life_sciences/lifsci_v1_n4_2009_paged.pdf

[10] Garcıa V. M. A., Ruiz M. P. P., and Perez J. R. P(2010), “Voice interactive classroom, a service-oriented software architecture for speech-enabled learning”, Journal of Network and Computer Applications, 33 (2010) 603–610, Elsevier. pp. 603-610.ilGil A. and García-Penalvo F. J. (2008), “Learner Course Recommendation in e-Learning Based on Swarm Intelligence”, Journal of Universal Computer Science, vol. 14, no. 16 (2008), pp. 2737-2755.

[11] Narada W. (1996), “Definition of Hidden Markov Model”. Available online at http://jedlik.phy.bme.hu/~gerjanos/HMM accessed 30th June 2008.