International Journal of Engineering Education Vol. 31, No. 2, pp. 521–528, 2015
Printed in Great Britain
0949-149X/91 $3.00+0.00

User Attention Analysis for E-learning Systems—Towards Intelligent Tutoring

YÜCEL UĞURLU
Department of Creative Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
E-mail: yucel.ugurlu@ni.com

We propose a new approach for e-learning systems that incorporate user voice activity information to build collaborative and intelligent learning environments. The proposed algorithm is based on the pitch frequency, power amplitude, and duration of a voice activity signal. Voice activity detection was integrated with the e-learning system for teaching LabVIEW-based graphical programming and it was tested on a group of students. The experimental results showed that the proposed approach is practical and highly suitable for real-time applications. Finally, student survey results are introduced to measure the e-learning satisfaction and user attention analysis approach.

Keywords: human–computer interaction; e-learning; LabVIEW; voice activity detection; intelligent tutoring

1. Introduction

The emergence of new technologies and the Internet has resulted in a major revolution in education in the form of e-learning. Today, e-learning is a popular alternative for students, engineers, and executives for training and development because of its fully customizable and user-oriented framework. Thousands of technical and management courses, including degree and certificate programs, are now being offered by universities, professional development centers, and industry training facilities worldwide [1, 2].

E-learning differs from classroom-based training in many ways and requires accurate planning, monitoring and execution for successful adoption. There are several benefits of e-learning such as: it is usually less expensive to deliver, it is self-paced, it is faster, it provides consistent content, it works from anywhere at any time, it can be updated quickly and easily, it can lead to an increased retention and a stronger grasp on the subject, and it can be easily managed for large groups of students [3, 4].

However, the most frequently cited challenge of e-learning is the amount of time required to develop and maintain an e-learning course. The success of e-learning rests on the fundamental requirement that instructors and students possess adequate technical skills to use e-learning tools effectively. Moreover, e-learning requires more responsibility and self-discipline for the learner to keep up with a more free and unconstrained learning process and schedule [5, 6].

In e-learning, ignoring the human–computer interaction (HCI) factors that come into play may lead to total failure. Therefore, next-generation e-learning systems should be sufficiently smart to be able to adapt to a student's learning style and assure high standards of accessibility and usability in order to make learner's interaction with the system as natural and intuitive as possible [7–11]. The primary objective of our work is to include, in addition to traditional interactions, computer-interaction data such as videos and audio signals in e-learning systems. Furthermore, we expect to understand user behaviors to build intelligent and collaborative learning environments [12, 13].

The rest of this paper is organized as follows: in Section 2, the proposed voice-activity-based user attention analysis approach is explained; in Section 3, the approach is tested on a group of students using an e-learning system developed to teach the graphical programming language of LabVIEW. Further, the results are assessed and student survey results are introduced; in Section 4, concluding remarks are provided, and future work is considered.

2. User attention analysis

Voice activity detection (VAD) is a technique used in speech processing in which the presence or absence of a human voice is detected. It is an important component of speech processing techniques, such as speech enhancement, speech coding, and automatic speech recognition. Various voice activity algorithms have been developed that provide various features and involve trade-offs between latency, sensitivity, accuracy, and computational cost in communication systems [14–16].

However, using VAD to formulize HCI data such as sleep, speech, silence, or emotions is a new concept in e-learning systems. In this research, VAD is embedded in the e-learning system. An appropriate power and pitch frequency of a voice

* Accepted 1 September 2014.

signal implies that the student is conversing with others. In general, e-learning systems are passive, and a voice signal is not required to interact with computers. Therefore, similar to in traditional teaching classes, user voice activity should be incorporated to evaluate user attentiveness.

Sounds produced by humans are created from the vibration of vocal chords that interrupts the flow of air and produces a frequency range of approximately 50–500 Hz. The voiced speech of a typical adult male will have a fundamental frequency of 85–180 Hz, and that of a typical adult female will be 165– 255 Hz [17, 18].

The basic principle of VAD is that it extracts measured features or quantities from the input signal and then compares these values with thresholds usually extracted from noise-only periods. A block diagram of the proposed intelligent e-learning system is shown in Fig. 1.

First, a voice signal $x(k)$ is acquired according to Shannon's sampling theorem, which states that a continuous signal must be discretely sampled at a minimum of double the frequency (Nyquist frequency) of the highest frequency present in the signal [19].

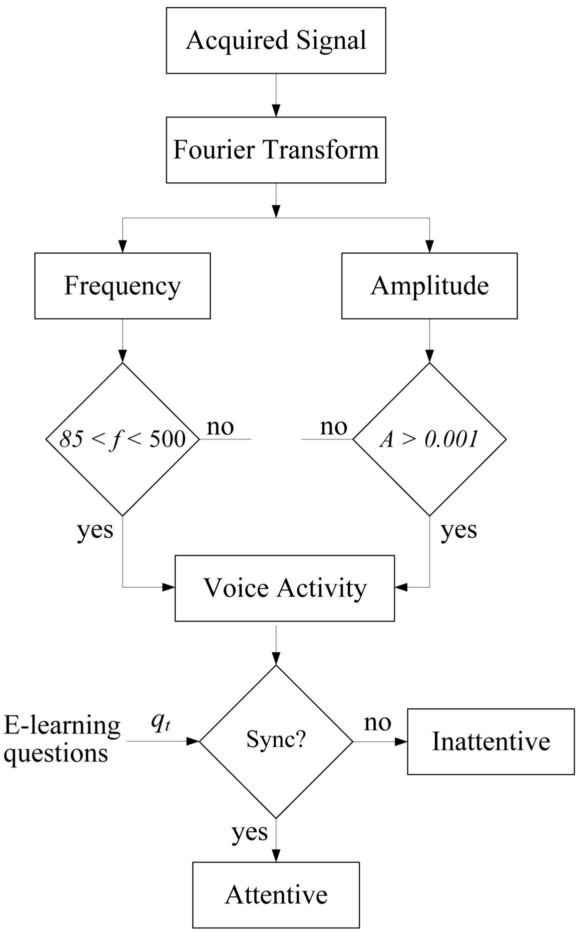

Fig. 1. Algorithm of the proposed intelligent e-learning system.

Next, the speech signal is characterized by a sequence of peaks that occur periodically at the fundamental frequency of the speech signal. In contrast, during unvoiced intervals, the peaks are relatively smaller and do not occur in any discernible pattern. Thus, the maximum peak amplitude during an analysis interval can be used to determine the amplitude of the signal in a simple manner and help distinguish between voiced and unvoiced speech segments.

A pitch detection algorithm is used to estimate the pitch or fundamental frequency of speech or a musical note. This can be done in the time or frequency domain or both. A Fourier transform, which expresses a signal in terms of the frequencies of the waves that constitute that signal, is an efficient approach for understanding the characteristics of speech signals. The Fourier transform of a signal is represented as:

$$X(n) = \frac{1}{N}\sum_{k=0}^{N-1} x(k)e^{-\frac{j2\pi kn}{N}} \quad \text{for}\, n = 0, N-1 \qquad (1)$$

in which $x(k)$ is the amplitude at time sample k, N is the number of samples or frequency points of attentiveness, n is a frequency value from 0 to $N - 1$, and $X(n)$ is the spectral-domain representation of $x(k)$.

A Fourier transform generates a power spectrum of the input signal and yields the main frequency component and amplitude value. User voice activity can be detected using:

$$v_t = \begin{cases} 1 \;\; 500 > f \geq 85 \quad \mathit{and} \quad A \geq 0.001, \\ 0 \qquad\qquad\qquad\qquad\quad \mathit{otherwise}, \end{cases} \qquad (2)$$

where t is the time in seconds, f is the fundamental frequency, and A is the power amplitude. For v_t = 1, the fundamental frequency (f) and amplitude (A) of the signal matches the range of the speech signal.

The basic principle of a VAD is that it extracts measured features or quantities from the input signal and then compares these values with thresholds usually extracted from noise-only periods. Voice activity is declared if the measured values exceed the thresholds. Otherwise, there is no speech activity or noise, i.e., it is assumed that the user is silent [20].

An e-learning system requires voice interaction to determine user attentiveness, which is represented numerically as:

$$q_t = \begin{cases} 1 \quad \mathit{question}\text{ asked}, \\ 0 \qquad\quad \mathit{otherwise}, \end{cases} \qquad (3)$$

where t is the time in seconds.

Equation (2) can be used to detect all kinds of sound activities including speech, emotions and noise. However, to realize a more robust system, user voice activity should be synchronized with the timing of the e-learning questions. Then, the user attention signal can be identified using the formula:

$$s_t = \begin{cases} 1 & q_t = 1 \ \ and \ \ \sum_{t}^{t+5} v_t \geq 1, \\ 0 & otherwise, \end{cases} \tag{4}$$

where t denotes the time in seconds. We assume that a user's voice activity would be high after the system requests user voice interaction in the case of attentive situations. No voice input or significant delay in the timing (more than 5 seconds) is an indicator of inattentiveness.

Intelligent e-learning is a new learning and teaching medium that uses an intelligent tutoring system so that online learning can adapt to a student's level of knowledge. Intelligent systems are designed to simulate human reasoning and learning, thus reducing the need for human intervention in the application process. Intelligent tutoring provides students with customized educational content and the unique feedback that they need and when they need it [21]. In the present study, we introduce an approach based on intelligent techniques, which in turn are based on user activity data, and build an adaptive framework.

However, user access information and duration is insufficient for building intelligent and adaptive systems. The relationship between e-learning content and user reaction should be measured and used for further analysis and evaluation. For this purpose, the integration of e-learning access and information on voice interactivity with the system should be implemented to build better systems.

An intelligent e-learning system records HCI data and uses this information to determine a user's attention, which is then used to recommend topics or to improve content. Therefore, an intelligent e-learning system should incorporate the behavior of users using image and signal processing [22]; typical examples are tracking face and head postures and detecting voice activities and emotions. An integrated voice activity signal detection component in an e-learning system is an essential source of information because it is an easy and common setup for home, school, or office.

E-learning systems includes a series of topics, and the accessed topics can be represented as:

$$T = (T_1, T_2, T_3, ..., T_n) \tag{5}$$

where n is the number of completed topics for a user. Similarly, the corresponding user attention level based on the VAD is represented as:

$$Z = (Z_1, Z_2, Z_3, ..., Z_n) \tag{6}$$

The attention level for each individual topic can be calculated as follows:

$$Z_i = \begin{cases} Attentive & \sum_{i=1}^{m} s_t \geq 1 \\ Inattentive & otherwise, \end{cases} \tag{7}$$

where i is the topic number, and m is the number of samples taken in a single topic.

If there is weak user attentiveness in certain topics, these topics can be displayed as incomplete or keywords of these topics can be used to build adaptive systems. Each topic will be recorded in the system according to its completion rate and student attentiveness. This information can be used to build an intelligent e-learning or tutoring system [23–25].

In conclusion, users learn at varying rates with different levels of knowledge and understanding. In traditional courses, teachers regularly adjust the homework based on their students' performance. Similarly, intelligent e-learning systems can provide feedback based on the user's interaction and behavior in order to encourage better performance.

3. Experimental results

To measure the performance of the proposed approach, we tested it on users of National Instruments' e-learning portal. This portal consists of more than 200 topics and 30 h of learning time, all pertaining to teaching graphical programming to engineering students in LabVIEW. LabVIEW is a graphical development environment for creating flexible and scalable virtual instruments. It offers an intuitive environment that is tightly integrated with measurement hardware, so engineers and scientists can quickly produce solutions for data acquisition, data analysis, and data presentation [26–28]. The e-learning content includes various single-topic videos, each 5–10 min in duration, readings, and short quizzes. In our e-learning system, we asked a simple question of the user at the end of each topic whether she/he wanted to proceed to the next topic. At this point, the user is required to respond by voice. Here we only checked the user voice activity rather than the actual answer to understand the user attentiveness.

A typical learning environment for LabVIEW graphical programming is shown in Fig. 2. In practice, a unique user ID and password are provided to access all e-learning functions [29]. Since users log into the system using their user IDs, it is easy to track their access history.

In our VAD system for LabVIEW, a standard sound card and microphone was used to acquire

Fig. 2. E-learning environment, which explains a typical data acquisition method, for LabVIEW.

voice signals [30]. LabVIEW includes simple virtual instruments (VIs) for analog input and output using a typical sound card built into many PCs. This is convenient for laptops because the sound card and microphone are usually built-in.

Human speech signals contain a significant amount of energy (under 2.5 KHz). Therefore, we used a 44 KHz sampling rate of the speech signal for a 16-bit monaural recording using the Acquire Sound Express VI in LabVIEW. Next, Tone Measurement Express VI was used to calculate the frequency and amplitude of the input signal. This system finds the tone with the highest amplitude in the signal and calculates its amplitude and frequency. Then, the amplitude and frequency information was used to determine if any user voice activities existed during the e-learning time. The front panel and block diagram of the LabVIEW program are shown in Figs. 3 and 4, respectively.

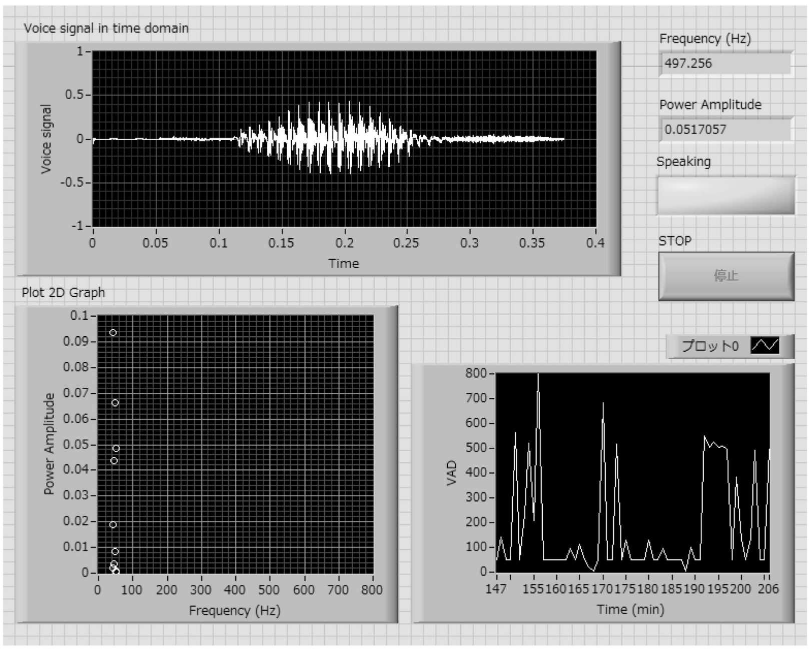

Fig. 3. Front panel of LabVIEW-based VAD program.

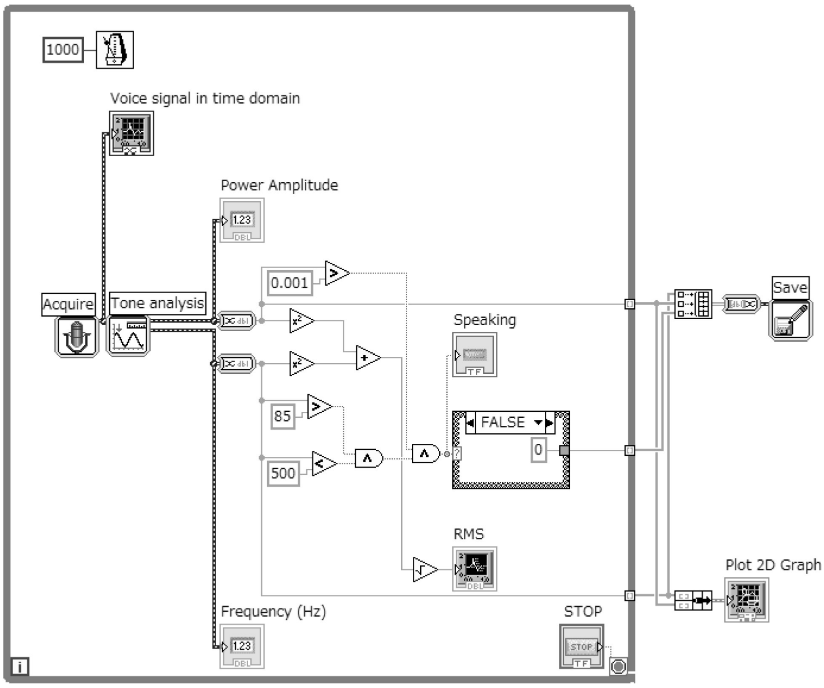

Fig. 4. Block diagram of LabVIEW code.

In our experimental setup, we collected voice samples every second for 60 min of e-learning for a single user. The voice activity signals were analyzed based on the proposed algorithm in Section 2. We assume that user voice activities should be synchronized with the timing of e-learning questions. For simplicity, we calculated the maximum amplitude and frequency values of the voice signal for each minute of the 60-min experiment. Fig. 5 shows the timing diagram of e-learning questions and voice activity of a single user. As shown, the user is required to provide some voice input to indicate the attention level for each topic. To identify the voice activity, the frequency and power amplitude of the signal are analyzed.

During the 60 minutes duration of e-learning, seven different topics were studied by the user. According to the experimental results, we found that user attentiveness was sufficiently good during the 8th, 14th, 22nd, 31st, 45th, and 57th minute, since there was timely voice activity. However, poor user attentiveness was detected at the 38th minute because of no voice activity. In our experiment, time

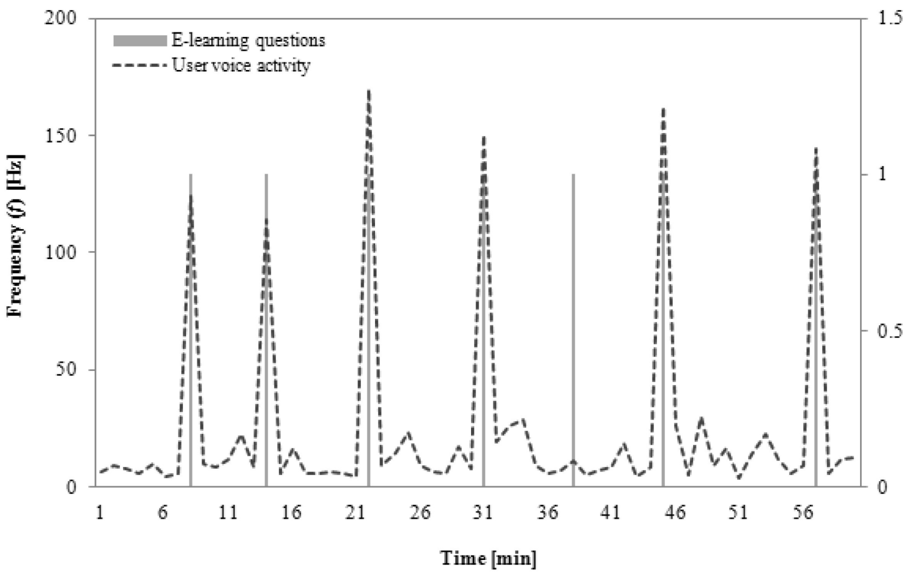

Fig. 5. Timing of e-learning questions and user voice activity.

Table 1. User attention analysis for the accessed topics

Topic no.	Content (T_i)	User attentiveness (Z_i)
Topic 1	LabVIEW introduction	Attentive
Topic 2	LabVIEW basics	Attentive
Topic 3	Express VIs	Attentive
Topic 4	Graphs and charts	Attentive
Topic 5	Troubleshooting, debugging	Inattentive
Topic 6	Subroutines	Attentive
Topic 7	Measurement introduction	Attentive

stamp information was used to check the user attentiveness. Table 1 summarizes the results of the user attentiveness analysis.

Next, the proposed approach was applied to a group of students. For this purpose, sixteen students asynchronously accessed the e-learning system in our laboratory environment. We provided a user ID and password for each student and gave them operational instructions separately. We should note that students had various programming experience with LabVIEW. There were three students who had already had some fundamental programming experience with LabVIEW, but most of them were at beginner level. Then, 60 minutes of self-paced e-learning time was given to the students to practice topics of interest in LabVIEW. At the same time, the VAD system was activated to record and analyze student attentiveness. Fig. 6 shows the attentiveness analysis results of sixteen students using user voice activity information.

For simplicity, topics receiving low attention are shown in the negative (lower) region of the graph. In general, students were highly attentive using the e-learning system, and their voice activity was highly synchronized with the timing of questions. Based on the proposed method, only four students (1, 3, 11, and 15) were inattentive for just five topics (8, 12, 16, 19, and 21) out of 69 total access during 60 min of e-learning time. According to student feedback, inattentiveness was mainly due to e-learning content or student interest. The results show that understanding student attention and attitude for e-learning system is very important to develop an intelligent tutoring framework.

Next, in order to evaluate the e-learning system as well as to understand the students' perception for attentiveness analysis, we conducted a survey for all e-learning participants after the training session. Fig. 7 shows the student survey results.

E-learning system satisfaction and the user attention analysis approach was evaluated by the students who participated in the e-learning session. The weighted average value was 4.4 for e-learning satisfaction and 3.7 for the usefulness of user attention analysis. The e-learning system was evaluated highly by students; however, their rating of the user attentiveness approach was somewhat unclear, being higher than 'neutral' but lower than 'good.' On the other hand, instructors and teachers were quite interested in the proposed approach to understand student learning schemes and to implement continuous improvement of e-learning systems.

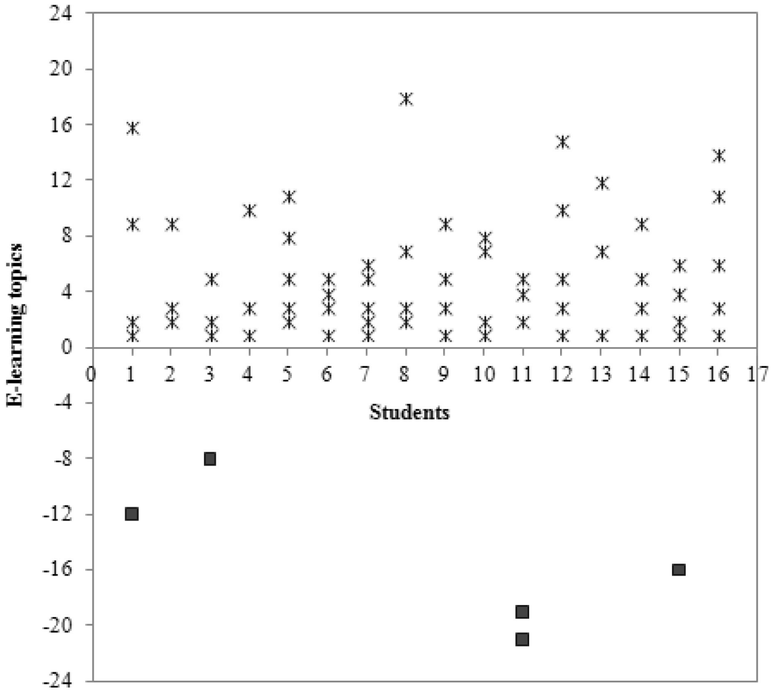

Fig. 6. User attentiveness for a group of students (negative values for topics indicate inattentiveness).

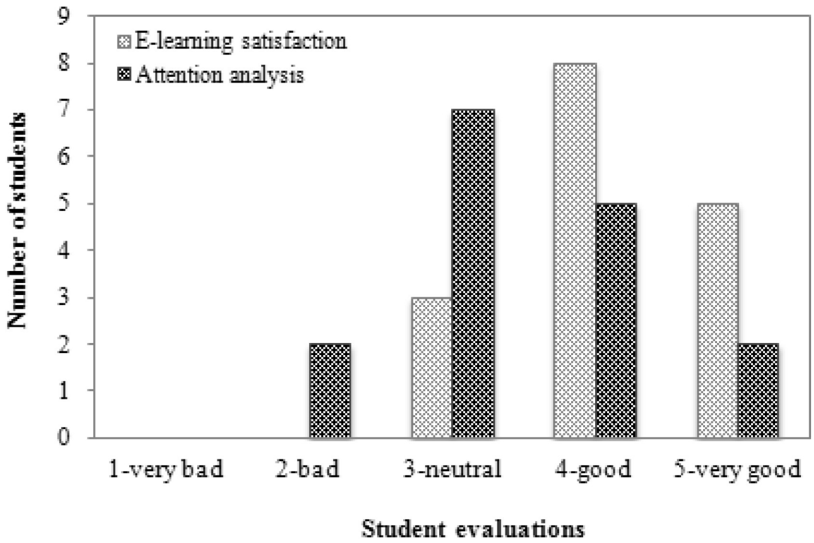

Fig. 7. Student evaluation of e-learning and attention analysis method.

4. Conclusions and future work

This paper proposed a new approach that incorporates user voice activity to build intelligent and collaborative tutoring systems. The proposed approach is inexpensive and practical for real-time applications. The experimental results show that user voice activity information is easy to integrate into existing e-learning systems. In our case study, we used a limited number of students and only one hour of training time. Therefore, the proposed approach should be tested on larger data sets and longer e-learning sessions to explore the common inattentive situations of students. In this way, user attention analysis can provide us with a better understanding of students' learning schemes, motivation, and the quality of e-learning systems.

In our case study, we only integrated user voice activity into the e-learning system for speaking and silent situations. However, user speech information carries important information, and it can also be used to explore student emotions as an extension of this research. In this case, a robust speech recognition algorithm should be considered. Our future study will be dedicated towards integrating a robust speech recognition system to understand user emotions in addition to user voice activity. Finally, these approaches will be combined to build intelligent and adaptive e-learning systems such as automatically recognizing topics that receive low attention for technical reviews and recommending highly interesting topics for individual users.

Acknowledgement—The author would like to thank National Instruments Japan for providing free access to their e-learning system.

References

1. R. C. Clark and R. E. Mayer, *E-Learning and the Science of Instruction*, Pfeiffer, 2007.
2. R. Ubel, Engineers turn to e-learning, *IEEE Spectrum Magazine*, **37**(10), 2000, pp. 59–63.
3. V. Cantoni, M. Cellario and M. Porta, Perspectives and challenges in e-learning: towards natural interaction paradigms, *Journal of Visual Languages and Computing*, 15, 2004, pp. 333–345.
4. L. M. Regueras, E. Verdú, M. F. Munoz, M.A. Perez, J. P. de Castro and M. J. Verdú, Effects of competitive e-learning tools on higher education students: A case study, *IEEE Transactions on Education*, **52**(2), 2009, pp. 279–285.
5. M. Rosenberg, *E-Learning: Strategies for Delivering Knowledge in the Digital Age*, McGraw-Hill, 2001.
6. D. Hasegawa, Y. Ugurlu and H. Sakuta, A case study to investigate different types of intrinsic motivation in using an e-learning system, *IEEE Global Engineering Education Conference*, Berlin, Germany, 2013, pp. 362–366.
7. T. Daradoumis, S. Caballé, J. M. Marquès and F. Xhafa, Intelligent collaborative e-learning systems and applications, *Studies in Computational Intelligence*, Springer, 2010.
8. A. R. Ahmad, O. Basir and K. Hassanein, Adaptive user interfaces for intelligent e-learning: issues and trends, *International Conference on Electronic Business*, Beijing, China, 2004, pp. 923–934.
9. A. Sears and J. A. Jacko, *The Human–Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Applications*, NewYork, 2008.
10. D. Te'eni, J. Carey and P. Zhang, *Human Computer Interaction: Developing Effective Organizational Information Systems*, John Wiley & Sons Ltd, Hoboken, 2007.
11. A. Kirlik, *Adaptive Perspectives on Human–Technology Interaction*, Oxford University Press, Oxford, 2006.
12. S. Stankov, V. Glavinic and Marko Rosic, *Intelligent Tutoring Systems in E-Learning Environments: Design, Implementation and Evaluation*, IGI Global, USA, 2011.
13. V. Aleven, J. Kay and J. Mostow, Intelligent tutoring systems, *International Conference on Intelligent Tutoring Systems*, Pittsburgh, USA, 2010, pp. 218–220.
14. J. Ramírez, J. C. Segura, J. M. Górriz and L. García, Improved voice activity detection using contextual multiple hypothesis testing for robust speech recognition, *IEEE Transactions on Audio, Speech, and Language Processing*, **15**(8), 2007, pp. 2177–2189.
15. R. Lawrence and J. Bing-Hwang, *Fundamentals of Speech Recognitions*, Prentice Hall New Jersey, 1993.

16. D. Ying, Y. Yan, J. Dang and F. K. Soong, Voice activity detection based on an unsupervised learning framework, *IEEE Transactions on Audio, Speech, and Language Processing*, **19**(8), 2011, pp. 2624–2633.
17. I. R. Titze, *Principles of Voice Production*, Prentice Hall, 1994.
18. R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice*, London, Taylor and Francis Ltd., 1987.
19. N. Kehtarnavaz and S. Mahotra, *Digital Signal Processing Laboratory*, BrownWalker Press, Boca Raton, Florida, 2010.
20. E. Verteletskaya and K. Sakhnov, Voice activity detection for speech enhancement applications, *Acta Polytechnica*, **50**(4), 2010, pp. 100–105.
21. V. Shute and B. Towle, Adaptive e-learning, *Educational Psychologist*, **38**(2), 2003, pp. 105–112.
22. Y. Ugurlu, Smart e-learning: Enhancement of human–computer interactions using head posture images, *International Journal of Engineering Education*, **29**(3), 2013, pp. 568–577.
23. K. Wong, K. Leung, R. Kwan and P. Tsang, E-learning: developing a simple web-based intelligent tutoring system using cognitive diagnostic assessment and adaptive testing technology, *Hybrid Learning*, 6248, 2010, pp. 23–34.
24. M. Virvou, G. Piraeus and S. C. Sidiropoulos, An intelligent tutoring system over a social network for mathematics learning, *International Conference on Information, Intelligence, Systems and Applications*, Piraeus, Greece, 2013, pp. 1–4.
25. P. Phobun and J. Vicheanpanya, Adaptive intelligent tutoring systems for e-learning systems, *Procedia—Social and Behavioral Sciences*, **2**(2), 2010, pp. 4064–4069.
26. J. Jovitha, *Virtual Instrumentation Using LabVIEW*, PHI Learning, 2010.
27. Y. Ugurlu, Measuring the impact of virtual instrumentation for teaching and research, *IEEE Global Engineering Education Conference*, Amman, Jordan, 2010, pp. 152–158.
28. G. W. Johnson and R. Jennings, *LabVIEW Graphical Programming*, McGraw-Hill Publishing, 2006.
29. Y. Ugurlu and H. Sakuta, E-learning for graphical system design courses: a case study, *IEEE International Conference on Technology Enhanced Education*, Amritapuri, India, 2012, pp. 1–5.
30. R. King, *Introduction to Data Acquisition with LabVIEW*, McGraw-Hill Science/Engineering/Math, 2012.

Yücel Uğurlu was born in Turkey and received his BS and MS degrees in electrical and electronics engineering from Ankara University in 1993 and 1995, respectively. He won a Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) scholarship for his postgraduate study in Japan, and received his PhD degree in information processing from the Tokyo Institute of Technology in 1999. After receiving his Ph.D., Dr. Uğurlu joined Olympus Corporation's R&D Center in Tokyo to develop advanced ultra-high-definition displays for natural vision systems. He is currently a research fellow at the University of Tokyo. His research interests include engineering education, learning tools and environments, e-learning, human-computer interaction, machine vision, robotics, and image processing. Dr. Uğurlu is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Japanese Society for Engineering Education (JSEE), and the Japanese Society of Applied Science (JSAS). He received a distinguished member award from JSEE in 2013.