IJFEAT
INTERNATIONAL JOURNAL FOR ENGINEERING APPLICATIONS AND TECHNOLOGY

Smartphone-Based Pronunciation Improving Technique for Ambient Environment-A Review

Sakshi S. Samdurkar [1], Shreyash A. Vagare [2], Prachi D. Tote[3], Prof. Aditya P. Bakshi[4]
1Student,Dept of CSE, J.D.I.E.T, Yavatmal, Maharashtra, India, ***samdurkarsakshi101@gmail.com***
2Student,Dept of CSE, J.D.I.E.T, Yavatmal, Maharashtra, India, ***shreyashsavn8@gmail.com***
3Student,Dept of CSE, J.D.I.E.T, Yavatmal, Maharashtra, India, ***prachi.tote77@gmail.com***
4Assistant Professor,Dept of CSE, J.D.I.E.T, Yavatmal, Maharashtra, India, ***bakshi.aditya.ab@gmail.com***

ABSTRACT

In an ambient intelligence (AmI) environment, electronic devices that comprise the Internet of things (IoT) network work together seamlessly to provide a wide variety of applications and intelligent services to users. It brings intelligence to everyday's environment and makes those environment sensitive to us. Computer-assisted pronunciation training (CAPT), a widely used application in the traditional Internet environment that corrects user's pronunciation, is a promising service for transition to the AmI environment. However, the migration of the CAPT to the AmI environment is challenging due to its high computational requirements that is at odds with the low computational capacity of IoT members. In this paper, we propose a smartphone-assisted pronunciation learning technique based on a lightweight word recommendation method that exploits built-in functions supported by IoT members and a computationally moderate word selection method. The experimental evaluation of the proposed method demonstrates that the user pronunciation is significantly improved without incurring unacceptable computational costs for a smartphone platform.

Keywords- *Ambient intelligence, Internet of things, Intelligent activity, Bag of phonemes, Computer assisted Pronunciation training system.*

1. INTRODUCTION

Ambient intelligence (AmI) refers to electronic environments that are sensitive and responsive in the presence of users [1]. It brings intelligence to everyday's environment and makes those environment sensitive to us. It is an exciting paradigm in IT in which people are empowered through digital environment i.e. aware of their presence and context and is sensitive, adaptive and responsive to their needs , habbits , gestures and emotions.

AmI potentially gives more control to humans by making their environment more responsive to intended actions and reducing the physical effort that is required to perform the task. At the same time AmI can take away the control when the environment performs the wrong actions.

Ambient Intelligence is a vision on the future of consumer electronics, Telecommunications and electronics that was originally developed in the late 1990s by Eli Zelkha and his team at Palo Alto Ventures for the time frame 2010–2020. In an ambient intelligence world, devices work in concert to support people in carrying out their everyday life activities, tasks and rituals in an easy, natural way using information and intelligence that is hidden in the network connecting these devices. As these devices grow smaller, more connected and more integrated into our environment, the technology disappears into our surroundings until only the user interface remains perceivable by users.

Ambient intelligence (AmI) refers to electronic environments that are sensitive and responsive in the presence of users. In an AmI environment, users can be supported by information as well as intelligence provided by electronic devices that comprise the Internet of things (IoT) network [3]. In this scenario, it is expected that members of the IoT network, such as RFID, sensors, appliances, and smartphones, should be distinguishable without manual configuration. The devices should also be able to communicate with each other through the data and information that they gather or produce , and must have the ability to make decisions to facilitate intelligent activities .To make the AmI environment user-friendly, a wide variety of applications, including traditional Internet applications,must be provided explicitly when requested by the users or by implicit prediction according to the situation .Computer-assisted pronunciation training (CAPT) is one such popular Internet application that can be adapted to the AmI environment.

A common feature of ambient intelligence is that many objects are inter-connected and act in unison, which is also a challenge in the Internet of Things. There has been a shift in research towards integrating both concepts, considering the Internet of Things as representing the future of computing and communications.

The Internet of Things (IoT) is a technology and a market development based on the inter-connection of everyday life objects with each other, applications and database data. These devices, objects or things (such as laptops, smartphones, onboard computers, video systems, household appliances, intelligent buildings, wireless sensor networks, ambient devices, RFID tagged objects and commodities) are identifiable, readable, recognizable, addressable and even controllable via the Internet.

2. LITRATURE REVIEW

J. Zhao, H. Yuan,W.-K. Leung, H.-Y. Meng, J. Liu, and S. Xia en.tl [2], proposed to enhance the effectiveness of training, researchers investigated synthetic pronunciation training that is based on multimodal speech examples. For example, a system may provide audio-visual speech examples that are composed of the pronunciation sound and visual animation of articulators. Moreover, an approach that generates the feedback by comparing articulation gestures of users and natives was also considered to provide visual insight. However, these systems usually require significant computational resources to conduct a detailed analysis of the speech signal and generate synthetic animation.

X. Qian, H. Meng, and F. Soong en.tl [4], proposed that in conventional CAPT systems, the feedback is generated by mimicking the tutoring strategy of language learning in the real-world; it can be roughly classified as a phonics training strategy and whole-word training strategy. Phonics training employs a phoneme emphasis strategy, which aims to inform the user how to pronounce each phoneme, and this strategy was frequently adapted to the development of conventional CAPT systems. It was argued that phonics training is an effective strategy for _ne-tuning a user's pronunciation because it is able to pinpoint the cause of error from the perspective of phonemes.

L. Wang, X. Feng, and H. M. Meng en.tl [6], proposed that in the early stage of studies in this field, CAPT systems provided a summarized feedback to users, e.g., the pronunciation score. Currently, a wide-variety of studies have been considered to produce effective feedback based on an expanded analysis on users' speech. Because frequently mispronounced phonemes may differ depending on the mother language of participating users, prior knowledge exploited from or encoded to speech corpus was used to enhance the training process. For example, based on the tendency of users to frequently confuse phonemes that are similar to phonemes of their mother tongue CAPT systems are able to provide more informative feedback to users by pointing out the source of the users' confusion.

W. Otto and R. Chester en.tl [7], proposed that as compared to phonics training, whole-word training is considered as an effective strategy for beginning-level users because it follows a natural approach to language learning. Users are able to view alphabets of words rather than phonetic symbols, as well as listen to good examples of pronunciation and then re-pronounce. This is known as a meaning or shape emphasis strategy, and enables users to naturally learn the relation between alphabets and pronunciation, whereby this strategy leads users to self-learning of productive pronunciation rules. Although phonemeon word-based feedback can be generated according to different tutoring strategies, users' pronunciation training is done by repeatedly pronouncing the same phonemes, words, or sentences that are listed by the system.

3. PROBLEM ANALYSIS

CAPT is a computer-based language learning technology that enables users to self-correct their pronunciation using an automatically generated training process. This system is particularly helpful to users who feel uncomfortable participating in oral presentations. It also eliminates the difficulty associated with finding bilingual tutors who are native speakers. Thus, the CAPT system is an effective alternative to traditional pronunciation training for users who want to participate in lessons without tutors or listeners. The limitation of CAPT is that users practice the same words or sentences repeatedly until their pronunciation of predetermined examples becomes acceptable. This process is somewhat tedious from the user's perspective. The situation becomes more challenging when the user has little training in the pronunciation of the target language and the system generates a large amount of feedback . In such cases, users can easily be exhausted owing to the extensive training, resulting in the discontinuation of the learning. To avoid this situation and overcome this problem, we develop our system to provide diverse and effective feedback.

4. PROPOSED WORK

The following are the procedural steps of proposed (SAPT) in IoT network:

Step 1: Firstly the user in AmI environment who wants to use the pronunciation learning application system speaks the test words displayed on the smartphone which is the typical IoT member used by the user.

Step 2: Now the system transfers the speech signals to an IoT member than can support speech recognition processing.

Step 3: Now the IoT member tests how the spoken words are actually recognized. If the words are pronounced in an exact manner then no recommendation of words is provided.

Step 4: If not then, the system generates the data set based on bag of phoneme model. Thus determines the correlation between the phonemes and mispronounced words.

Step 5: Depending upon the correlation between the phonemes and the mispronounced words, few words will be suggested.

Step 6: Out of the suggested words, the words which are having the highest selection probability will be sent to the system.

Step 7: Finally the user Pronuonces and practices the words that are being displayed.

5. SYSTEM DESIGN

The Figure shows the mechanism and the details about the SAPT method:

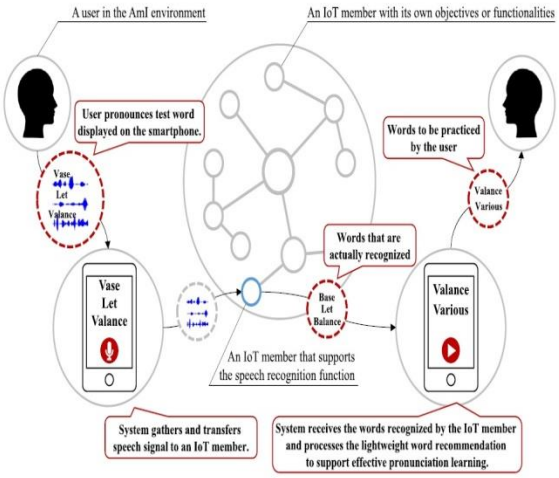

Fig 5.1. SAPT System Design

6. CONCLUSION

A smartphone-based pronunciation improving technique that provides ambient enviornment to users is proposed. Adaptive and personalized pronunciation training is provided by the information given by IOT member by using the bag of phoneme model and correlation analysis. The training process is facilitated based on the most initiative approach, which is to listen to native pronunciation for various words and they pronounce them. With this the users will be able to easily and rapidly correct their pronunciation.

REFERENCES

[1] H. Hagras, D. Alghazzawi, and G. Aldabbagh, ``Employing type-2 fuzzy logic systems in the efforts to realize ambient intelligent environments [application notes]," IEEE Comput. Intell. Mag., vol. 10, no. 1,Feb. 2015

[2] J. Zhao, H. Yuan,W.-K. Leung, H.-Y. Meng, J. Liu, and S. Xia, ``Audiovisual synthesis of exaggerated speech for corrective feedback in computer assisted pronunciation training," in Proc. IEEE Int. Conf. Acoust., Speech Signal

Process., Vancouver, BC, Canada, May 2013, pp. 8218_8222.
[3] L. Atzori, A. Iera, and G. Morabito, ``The Internet of Things: A survey," Comput. Netw., vol. 54, no. 15, pp. 2787_2805, Oct. 2010.
[4] X. Qian, H. Meng, and F. Soong, ``Capturing L2 segmental mispronunciations with joint-sequence models in computer-aided pronunciation training (CAPT)," in Proc. Int. Symp. Chin. Spoken Lang. Process., Nov./Dec. 2010, pp. 84_88.
[5] H. Franco, H. Bratt, E. Shriberg, V. Abrash, and K. Precoda, ``EduSpeak: A speech recognition and pronunciation scoring toolkit for computeraided language learning applications," Lang. Test., vol. 27, no. 3, pp. 401_418, 2010.
[6] L. Wang, X. Feng, and H. M. Meng, ``Mispronunciation detection based on cross-language phonological comparisons," in Proc. Int. Conf. Audio, Lang. Image Process., Shanghai, China, Jul. 2008, pp. 307_311.
[7] W. Otto and R. Chester, ``Sight words for beginning readers," J. Educ. Res., vol. 65, no. 10, pp. 435_443.