The Importance of Recommender and Feedback Features in a Pronunciation Learning Aid

Dzikri Rahadian Fudholi

The Australian National University / Canberra, ACT, Australia
Universitas Gadjah Mada / D.I.Yogyakarta, Indonesia
u5857432@alumni.anu.edu.au

Hanna Suominen

The Australian National University / Canberra, ACT, Australia
Data61, The Commonwealth Scientific and
Industrial Research Organisation / Canberra, ACT, Australia
University of Canberra / Canberra, ACT, Australia
University of Turku / Turku, Finland
hanna.suominen@anu.edu.au

Abstract

Verbal communication — and pronunciation as its part — is a core skill that can be developed through guided learning. An artificial intelligence system can take a role in these guided learning approaches as an enabler of an application for pronunciation learning with a recommender system to guide language learners through exercises and feedback system to correct their pronunciation. In this paper, we report on a user study on language learners' perceived usefulness of the application. 16 international students who spoke non-native English and lived in Australia participated. 13 of them said they need to improve their pronunciation skills in English because of their foreign accent. The feedback system with features for pronunciation scoring, speech replay, and giving a pronunciation example was deemed essential by most of the respondents. In contrast, a clear dichotomy between the recommender system perceived as useful or useless existed; the system had features to prompt new common words or old poorly-scored words. These results can be used to target research and development from information retrieval and reinforcement learning for better and better recommendations to speech recognition and speech analytics for accent acquisition.

1 Introduction

Pronunciation Learning Aid (PLA) is a system for learning to pronounce better. Pronunciation learning is needed because speaking is a hard task for the human brain (Levelt, 1993). In the process of learning, a person uses another person, a book, or another resource to get the knowledge they need. PLA is one of those facilities that enables a learning experience by giving a practice module.

A number of use cases for PLA exist in real life. They encompass the entire spectrum from supporting teachers' work flow in classrooms to computer-assisted virtual learning environments (Figure 1). That is, more and more learning can happen from home and teachers' time can be used more sparingly.

In this short paper, we are introducing an English PLA prototype with a *Recommender System* (RS) and *Feedback System* (FS). RSs are commonly used to recommend movies, books, music, or similar items (Lü et al., 2012), but their applications to language learning are only emerging. On the contrary, FSs for language learning are more established (e.g., using visual feedback (Wen et al., 2006) or *Speech Recognition* applied to pronunciation evaluation (Abdou et al., 2006)). These two systems could work together to facilitate a language learner to do self-practicing as follows: The RS can give specialized guidance to the language learner (Adomavicius and Tuzhilin, 2005) which in our case translates to the PLA users being guided through a series of exercises that are fit for them. After this, the user can practice by reading and pronouncing these recommended
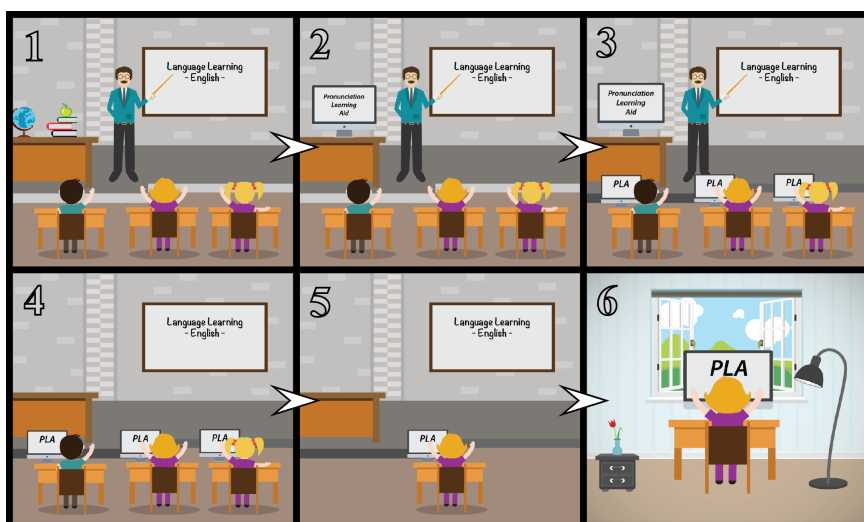
83

Figure 1: Illustration of replacing teaching role to PLA

word/phrases. One of the best options to learn pronunciation is by listening to an example (Leather, 1983). Consequently, providing an audio example is one of the feedback features in our FS.

The focus of the paper is on the result of our user study regarding the importance of each system feature in our PLA. Instead of assuming that both the RS and the FS are needed, we need actual evaluation data to inform our judgment and decision-making regarding the features to include.

The rest of the paper is organized as follows: First, in Section 2, we describe our materials and methods. Then, in Sections 3 and 4, we present and discuss our results, respectively. Finally, in Section 5, we conclude the paper.

2 Materials and Methods

The PLA prototype had the following two main data elements: *item* and *user*. *Item* [word] was the recommended option, which was populated by using the *Ogden's basic word list* (Ogden and Halász, 1935). Each word was also associated with a commonness status by counting its occurrences in the *Europarl* corpus (Koehn, 2005). *User* referred to a specific language learner and *user data* to this person's recorded pronunciation history (i.e., the past learning experience while using the system) and demographic data (e.g., first language, nationality, and age). These recordings were enriched by comparisons to other users' data.

The prototype included the following features. The RS produced the recommendation choices of 1) a *New Common Word*, 2) an *Old Poorly-Scored Word*, and 3) a *New Word from Others' Poorly-*

scored Words. The FS had the feedback features of 1) a *Speech Replay*, 2) a *Pronunciation Example*, and 3) a *Pronunciation Score*. Each of the six features was implemented using a different processing method but based on the same data available in the system.

2.1 Processing Methods

Some of the methods were as simple as reading or counting the data such as counting word frequency whilst others used more advanced machine learning algorithms. In the RS, the New Common Word feature combined the word commonness and the user history to find the most common word that the user has not seen yet. For the Old Poorly-Scored Word feature, it only used the user history to recommend a word with the poorest pronunciation score by the FS. The last feature (i.e., New Word from Others' Poorly-scored Words) analyzed all users' history and demographic data: First, using the *K-Nearest Neighbors* algorithm (Bobadilla et al., 2013), it found similar users to the current user, followed by applying two equally weighted spaces as follows: users' history space with each word score as a dimension and demographic space with demographics as dimensions. Second, the RS built a list of words that were poorly scored in other similar users' histories but not yet seen by the current user. Each of the three RS methods ran once to initialize interaction with the user and again every time the user finished an exercise.

In FS, once one of the three recommendation options was chosen, the system generated three feedback features for that specific word option.

Table 1: Kano conclusion table

		Response to a negative question				
		I like it that way	It must be that way	I am neutral	I can live with it that way	I dislike it that way
Response to a positive question	I like it that way	Questionable	Delighting	Delighting	Delighting	Satisfier
	It must be that way	Reverse	Indifferent	Indifferent	Indifferent	Basic
	I am neutral	Reverse	Indifferent	Indifferent	Indifferent	Basic
	I can live with it that way	Reverse	Indifferent	Indifferent	Indifferent	Basic
	I dislike it that way	Reverse	Reverse	Reverse	Reverse	Questionable

The *Pronunciation Example* feature worked by playing a stored example (i.e., sound file) for the word. Both the *Speech Replay* and *Pronunciation Score* feedback were available after the user had recorded their own speech: The *Speech Replay* feature simply replayed the recording. The *Pronunciation Score* feature was similar to *Automatic Pronunciation Scoring* (Kim et al., 1997). However, due to the time constraint of our research, we had to use random scoring instead in the user study.

2.2 Evaluation Methods

After obtaining the proper ethics approvals and research permissions, we evaluated the importance of each feature in our prototype by conducting a user study. We asked 16 international students to complete our questionnaire after they had tried using the prototype. We used a scenario for each feature so that every respondent had the same experience but with freedom to continue practicing as they wished.

For our questions, we used the *Kano Model* (Yadav, 2016) that built a positive and negative question for each feature to allow concluding whether the user likes a given feature or not. For example, a positive question was "How do you feel if the system is able to replay your recorded speech?" and a negative question was "How do you feel if the system cannot replay your recorded speech?".

To make conclusions from the question pairs for each feature, we used Table 1 (Yadav, 2016). One example of the conclusion was *Delighting*, which meant that the existence of the feature is good. The conclusion of *Reverse* meant that the system is better without the feature.

At the end of the questionnaire, we also asked open questions as follows: "Do you need help to learn pronunciation?", "What difficulties are you having?", "What do you think about the PLA?", "What improvements would you like to see?".
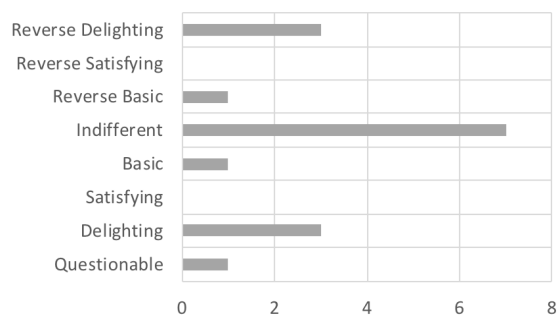

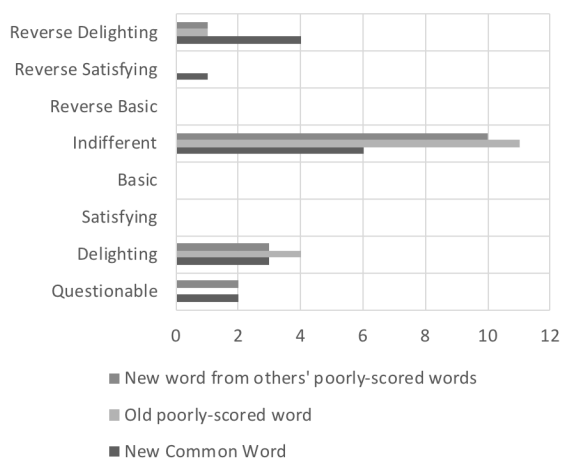Figure 2: The importance of the RS as a whole

Figure 3: The importance of each RS feature

3 Results

In order to have a realistic case where the language learners are using their own personal computers or laptops, we used an online questionnaire in the user study. Alongside the questionnaire link, we provided the respondents a link to download the prototype. The prototype was built in *Java* and each respondent had to install it on their device.

Before assessing the importance of each feature, we addressed the importance of the RS and FS. Most of the respondents were feeling indifferent about the existence of RS (Figure 2). The same number of respondents were feeling delighted and
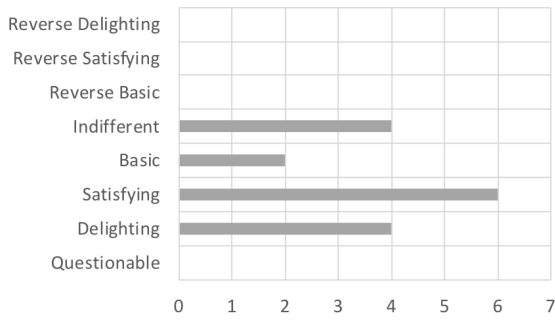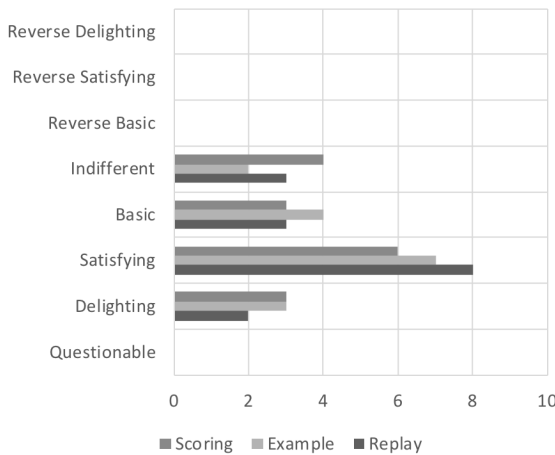
Figure 4: The importance of the FS as a whole

Figure 5: The importance of each FS feature

reverse delighted, and the same conclusion held for the basic and reverse basic. The trends for the features in RS and FS were similar (Figure 3). Most respondents were feeling indifferent and the numbers of respondents feeling delighted and reversely delighted were approximately the same.

We asked the importance of FS in PLA and got the result that a clear majority of respondents were feeling satisfied without anyone feeling the reverse (Figure 4). A similar trend also occurred in the result of each FS feature (Figure 5); most respondents felt satisfied without any reverse feeling.

Based on the answers to the open questions, thirteen respondents needed help to learn pronunciation with the main reason of their accent. Most respondents felt the usefulness of PLA and especially the FS features were desirable. The respondents were keen to use a PLA not only for English but also for Mandarin and French.

4 Discussion

From the result we can see that most respondents were having difficulties with their pronunciation learning, mainly because their foreign accent. They welcomed help from any source, including PLA, to correct their pronunciation.

The role of RS in PLA was somewhat unclear. The results diverged between the RS being needed or not needed with the same number of respondents in both sides while most of them felt indifferent. Some respondents did not know how to begin the exercises and needed the guide to do so. Otherwise, some respondents felt the system recommendation was not the best for them to learn and they know better what they should learn.

For each recommendation options, a new common word was not preferred. The respondents preferred to choose on their own because they did not want to just learn common words. Possibilities to practicing poorly-scored words were requested for.

Including the FS in the PLA was crucial but the RS features could be optional. None of the respondents said that the PLA would be better without the entire FS or any of its features. Their key expectation was to receive feedback. Having examples and replay options was also expected but having correctness scoring as a pronunciation feedback functionality was not an expectation but rather a bonus.

5 Conclusion

As expected, our technology-assisted approach for pronunciation learning was perceived as useful but surprisingly, recommendations were not a key feature for a good system. Instead, receiving feedback was essential in a PLA. However, sixteen respondents is a small sample, and this limits the generalizability of these conclusions.

Acknowledgments

References

Sherif Mahdy Abdou, Salah Eldeen Hamid, Mohsen Rashwan, Abdurrahman Samir, Ossama Abdel-Hamid, Mostafa Shahin, and Waleed Nazih. 2006. Computer aided pronunciation learning system using speech recognition techniques. In *Ninth International Conference on Spoken Language Processing*.

Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749.

Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132.

Yoon Kim, Horacio Franco, and Leonardo Neumeyer. 1997. Automatic pronunciation scoring of specific phone segments for language instruction. In *Eurospeech*.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86.

Jonathan Leather. 1983. Second-language pronunciation learning and teaching. *Language Teaching*, 16(3):198–219.

Willem JM Levelt. 1993. *Speaking: From Intention to Articulation*, volume 1. MIT press, Cambridge, MA, USA.

Linyuan Lü, Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. 2012. Recommender systems. *Physics Reports*, 519(1):1–49.

Charles Kay Ogden and Gyula Halász. 1935. *Basic English*. Kegan Paul, Trench Trubner & Co., Ltd., London, UK.

Sayling Wen, Zechary Chang, and Pinky Ma. 2006. Language learning system and method with a visualized pronunciation suggestion. US Patent 7,153,139.

Sachendra Yadav. 2016. The Kano model — a tool to prioritize the user's wants and desires. Last Accessed: 16 October 2017.