Measuring pronunciation improvement in users of CAPT tool TipTopTalk!

Cristian Tejedor-García[1], *David Escudero-Mancebo*[1], *Enrique Cámara-Arenas*[2], *César González-Ferreras*[1], *Valentín Cardeñoso-Payo*[1]

[1]Department of Computer Science
[2]Department of English Philology
University of Valladolid

cristian.tejedor.91@gmail.com

Abstract

We present a L2 pronunciation training serious game based on the minimal-pairs technique, incorporating sequences of exposure, discrimination and production, and using text-to-speech and speech recognition systems. We have measured the quality of users' production during a period of time in order to assess improvement after using the application. Substantial improvement is found among users with poorer initial performance levels. The program's gamification resources manage to engage a high percentage of users. A need is felt to include feedback for users in future versions with the purpose of increasing their performance and avoiding the performance drop detected after protracted use of the tool.

Index Terms: computer assisted pronunciation training, leaning analytics, L2 pronunciation, minimal pairs [1].

1. Introduction

There are many software tools that rely on speech technologies for providing users with L2 pronunciation training [1]. While such tools undoubtedly engage users in learning-oriented practice, there have been very few attempts to objectively assess the actual improvement attained by them [2][3]. In the present paper, we show the performance results of users of TipTopTalk! [4][5], a serious game designed for L2 pronunciation training and testing.

We will describe the software tool, the test campaign, the learning analytics technique we have implemented and, finally, the results. Discussion, conclusions and prospects for future development are included.

2. Program description

TipTopTalk! provides learning practice with minimal pairs (pairs of words identical except for one single phoneme). Training proceeds along cycles of exposure, discrimination and production stages [6]. With each exposure turn, the tool reproduces a pair iteratively to allow users to perceive a particular phonemic contrast. During the discrimination phase, users are aurally presented with one of the components of particular pairs, and they must identify which of two written words is being pronounced by the program. For exposure and discrimination exercises the same TTS system is used. In production exercises, the system asks the user to orally produce one of the words of particular pairs, and then the recorded word is analyzed by an automatic speech recognition system that assesses its accuracy.

The three activity types are articulated into a serious game that includes stimulating images and sounds, a scoreboard, a rank of best players, a timer, etc. At the beginning of each session users are required to select between two playing modalities: Training or Challenge. They can also choose to focus on exposure, discrimination or production routines, and select the phoneme or pair of phonemes they want to practice with. A third playing mode alternates the three activity types in order to test and rank the abilities of the player. Figure 1 shows different states of the program.

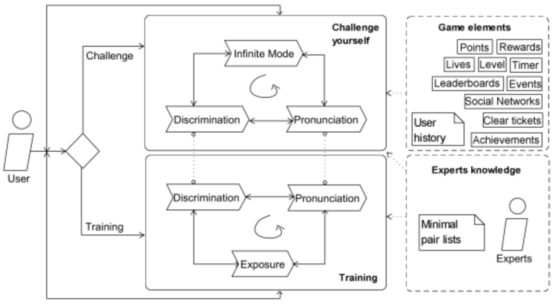

Figure 1: *Flow chart of the activities proposed to users.*

TipTopTalk! uses Google's TTS system in exposure and discrimination exercises, and Google's ASR system for assessing user's production. Results are saved as a JSON format log file that anonymously compiles all possible data diachronically. The application runs with a list of 793 minimal pairs for American English, 168 for Chinese and 155 for European Spanish. Currently it is being added new words for German and European and Brazilian Portuguese. Pairs are grouped within categories of phonemic contrasts.

3. Assessment

We carried out a three-week test campaign after distributing the tool among students of computing engineering and English philology, and students of Chinese. Up to 100 users registered for practicing with TipTopTalk!, and 58% of them made extensive use of it with up to 6000 interactions during the first week. Some of the participants remained engaged for several weeks, registering more than 11000 interactions.

This campaign has generated a database with 88000 entries containing information about the use of the tool made by each user in relation to the different exercises. This set of records *R*, can be interpreted as

[1] Thanks to Ministerio de Economía y Competitividad y Fondos FEDER project key: TIN2014-59852-R Videojuegos Sociales para la Asistencia y Mejora de la Pronunciación de la Lengua Española, and Junta de Castilla y León project key: VA145U14 Evaluación Automática de la Pronunciación del Español Como Lengua Extranjera para Hablantes Japoneses.

$$R = E \bigcup D \bigcup P \bigcup O \qquad (1)$$

where E stands for the entries corresponding to exposure turns, D for those corresponding to discrimination exercises, P for user productions and O for control manipulations (for example, changes of activity, logging in or out of the system, etc.). Discrimination exercises are characterized as

$$D = \bigcup_u \bigcup_k D_{u,k} \qquad (2)$$

where $D_{u,k}$ represents the discrimination attempts of user $u=1..U$ of the words of a kind of pair $k=1..K$. $D_{u,k}$ constitutes a sequence of chronologically ordered attempts,

$$D_{u,k} = (d_1..d_{N_{u,k}}) \qquad (3)$$

where $N_{u,k}$ stands for the number times that user u tries to discriminate words of a kind of pair k. A function of quality $f_D(D_{u,k}, w, s)$ computes the average number of correct answers made in a window of w attempts, beginning at the position $s = 1..N_{u,k} - w$, in $D_{u,k}$.

For user u's production of words of a kind of pair k, we have the sequence

$$P_{u,k} = (p_1..p_{M_{u,k}}) \qquad (4)$$

where p_i represents the attempts to pronounce words of a kind of pair k with the fact that the game allows up to 5 attempts per word and $M_{u,k}$ stands for the number times that user u tries to pronounce words of a kind of pair k. The function of quality $f_P(P_{u,k}, w, s)$ where $s=1..M_{u,k}$ measures the quality of a user u's pronunciation attempts words of a kind of pair k within a window of w words (with up to 5 attempts) beginning at position s. Function f_P accounts for the position of the target word within a list of predictions by the ASR, the reliability indicators generated by the ASR system, the number of attempts made by the user, and the possibility of homophones.

The difference between the value of f at a given s, relative to the value of f for $s=0$ will tell us about the performing evolution of the user both in the discrimination and production of the different pairs and/or phonemes.

4. Results and discussion

The analysis of results concerning the improvement functions f_D and f_P point to significant correlations between the user, the kind of pair that is being discriminated or the word produced, and the number of trials (ANOVA test) both in discrimination and production phases.

Figure 2 shows the evolution of functions f_D and f_P at s. We show their average values varying u and k with a size window of w=6. In order to interpret the dependence of u, we distinguish three categories of users depending on the values of f in the initial window. We assume this value to be representative of the initial competence of each user before using the tool for the first time.

User's performance shows improvement along time both in discrimination and pronunciation tasks. In the production mode, it is users with a poorer initial level who undergo the most significant progress. Users with a higher initial level (some of them are, in fact, native speakers) register an initial drop in performance which we think attributable to the playability variables introduced in order to make the game more challenging (for example, in discrimination exercises users must click on one or the other word within a pair depending not only on the word they hear, but also on the background color displayed).

A general drop in performance quality is detected with regard to pronunciation after some time. The average user progresses initially towards an optimal point after which the values of f begin to fall. We believe this decrease in performance has to do with habituation and gradual loss of interest in the game.

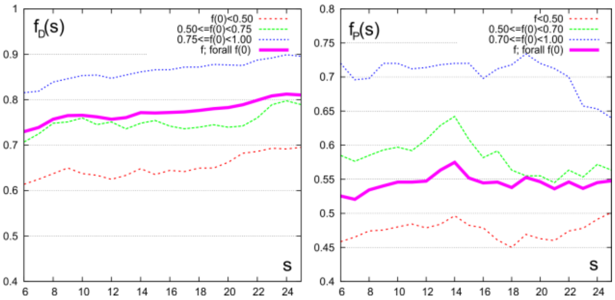

Figure 2: *Evolution of the function of quality along time of use in discrimination (left-hand diagram) and production (right-hand diagram).*

5. Conclusions and future work

Experimental results show that the use of TipTopTalk! is conducive to an improvement in L2 pronunciation and phoneme discrimination among users with a low initial level. Despite the introduction of gamification elements, a habituation factor leads to a fall in interest and performance after protracted use. This suggests the convenience of introducing feedback mechanisms to assist and guide users especially when a performance drop is detected.

There is a high correlation between particular phoneme contrasts and performance results. We are currently working on the definition of a new version of the program that includes exercises adapted to difficulties concerning specific contrasts. This new version will allow us to analyze aspects of use in relation to exposure and discrimination when the same kind of pronunciation difficulties is encountered repeatedly.

6. References

[1] D. Escudero and M. Carranza, "Nuevas propuestas tecnológicas para la práctica y evaluación de la pronunciación del español como lengua extranjera", *Congreso AEPE 2015*.

[2] G. Linebaugh and R. Thomas. "Evidence that L2 production training can enhance perception". *Journal of Academic Language & Learning 2015*, 9 (1): 1–17.

[3] N. Kartushina and A. Hervais-Adelman and U. H. Frauenfelder and N. Golestani. "The Effect of Phonetic Production Training with Visual Feedback on the Perception and Production of Foreign Speech Sounds." *The Journal of the Acoustical Society of America 138*, no. 2 (August 2015): 817–32.

[4] D. Escudero-Mancebo and E. Cámara-Arenas and C. Tejedor-García and C. González-Ferreras and Valentín Cardeñoso-Payo, "Implementation and Test of a Serious Game Based on Minimal Pairs for Pronunciation Training", *SLaTE*. pp. 125-130, 2015.

[5] C. Tejedor-García, V. Cardeñoso-Payo, E. Cámara-Arenas, C. González-Ferreras, and D. Escudero-Mancebo, "Playing around minimal pairs to improve pronunciation training," *IFCASL 2015*.

[6] E. Cámara-Arenas, Native Cardinality: on teaching American English vowels to Spanish students, *S. de Publicaciones de la Universidad de Valladolid*, Ed., 2012.