Iberspeech 2016
23-25 November 2016, Lisbon, Portugal

TipTopTalk! Mobile application for speech training using minimal pairs and gamification

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

[1]Department of Computer Science
[2]Department of English Philology
University of Valladolid
cristian@infor.uva.es

Abstract. This demonstration describes the TipTopTalk! mobile application, a serious game for foreign language (L2) pronunciation training, based on the minimal-pairs technique. Multiple Spoken Language Technologies (SLT) such as speech recognition and text-to-speech conversion are integrated in our system. User's interaction consists in a sequence of challenges along time, for instance exposure, discrimination and production exercises. The application implements gamification resources with the aim of promoting continued practice. A specific feedback is also given to the user in order to avoid the performance drop detected after the protracted use of the tool. The application can be used in different languages, such as Spanish, Portuguese (European and Brazilian), English, Chinese, and German.

Keywords: serious game, speech technology, computer assisted pronunciation training, gamification, learning analytics, L2 pronunciation, minimal pairs

1 Introduction

There are many software tools that rely on speech technologies for providing to users L2 pronunciation training in the field of Computer Assisted Pronunciation Training (CAPT)[4]. 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 [8][7]. The volume of technological services for smartphones and other smart devices is growing everyday [1]. Currently the most popular mobile and desktop operating systems grant users a free access to several Text-To-Speech (TTS) and Automatic Speech Recognition (ASR) systems. Besides, the combination of adequate teaching methods and gamification strategies will increase user engagement, provide an adequate feedback and, at the same time, keep users active and comfortable [10][9].

This paper describes the software tool TipTopTalk![1] [5][11][12] a second generation serious game application designed for L2 pronunciation training and

[1] https://play.google.com/store/apps/details?id=uva.eca.simm.tiptoptalk

testing. It is a two-years project focused on advanced research in speech training technology, such as speech recognition and text-to-speech conversion and the successful joint integration of them in a multilingual and multimodal information retrieval system. The languages considered in the project are Spanish, Portuguese (European and Brazilian), English, simplified Chinese, and German.

The rest of the paper is structured as follows. Section 2 offers an overview of our system, the application dynamics and the user interface. Section 3 describes the demonstration's script. Finally, section 4 provides the conclusions and future work.

2 Description of the system

2.1 General overview of TipTopTalk!

Three main elements are involved in our system, an Android client application, an own web server and external services provided by Google. See references [13][12][11] for more specific details. Figure 1 represents the conceptual architecture of the Android client application. The *Control* module includes the application's business logic. The *minimal pairs database* is accessed by the *Control* component in order to extract the minimal pairs lists of each language. The *Game Interface* component presents each pair to the users in accordance with the game dynamics. The *Control* component makes use of an *ASR component* that translates spoken words into text. When the patterns produced by the ASR component match those of the target words, the pronunciation is correct. The *TTS component* is used to generate a spoken version of any required word. It allows users to listen to a model pronunciation of the words before they try to pronounce them themselves. We use both Google's free ASR and TTS system. However, TipTopTalk! adapts to any ASR or TTS that works with Android.

A *Configuration* component selects the language in which the ASR and TTS components operate. Furthermore, it allows selecting among different sets of minimal pairs according to the language to be tested. Results will show the capital importance of a proper selection of minimal pairs. The *minimal pairs database* –which constitutes the knowledge database of the system– can be updated in order to improve the system or to include new challenges.

Finally, a *Game Report* is generated at the end of each game. This report registers user dynamics, including the timing of the oral turns (both for recognition and for synthesis) and the results obtained. We gather relevant quantitative data from all emerging events in the visual interface of the application with which we feed a daily log for each user in order to determine whether her or his pronunciation skills are improving. In addition, we send depersonalized user's interaction events to our Google Analytics account in order to compute how often a given event has occurred.

2.2 Pedagogical activities cycle

TipTopTalk! follows a learning methodology based on the sequencing of three different learning stages: exposure, discrimination and pronunciation [3]. It relies

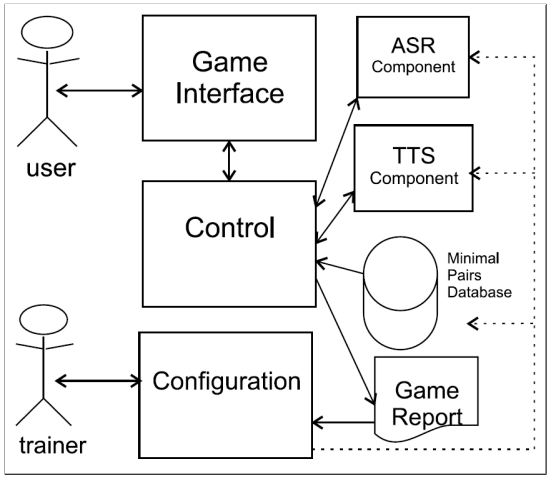


Fig. 1. Conceptual components of the client's system.

on the use of minimal pairs. They raise users' awareness of the potential risks of generating wrong meanings when phonemes are not properly produced[2]. The lists of minimal pairs used by the tool are selected by expert linguists in order to obtain the best possible results. TipTopTalk! tries to adapt this methodology with gamification elements since it is a serious game.

As a consequence, there are three main game modes. The first one is the exposure mode, players become familiar with the distinctive phonemes within sequences of minimal pairs selected by a native linguist and presented at random. The aural correlate of each word is played a maximum of five times. Then, users decide whether to move on to next round of words, or to record their own realization of the words to compare it with the TTS version.

Secondly, in the discrimination mode, users test their ability to discriminate between the elements of minimal pairs. They listen to the aural correlate of any of the words in each pair and must match it with the correct written form on the screen. As part of the gamification strategy, the game randomly asks users to pick the word that has not been uttered, rather than the uttered one. At higher levels of difficulty, the phonetic transcription of each word, otherwise visible, is removed. These strategies aim at the promotion of user adaptation and engagement.

Finally, in the pronunciation mode, participants are asked to separately read aloud (and record) both words of each minimal pair. A real-time feedback is provided instantly. Native model pronunciations of each word can be played as many times as the user needs. Speech is recorded and played using third party ASR and TTS applications.

2.3 Gamification

TipTopTalk! adapts to the player in function of the interaction results giving a specific feedback. New training modes are suggested based on the results of the current one. For instance, in discrimination mode, if an user achieves the maximum score, advancement to a pronunciation mode will be suggested. Otherwise, going back to exposure mode will be automatically recommended after a low score has been attained in discrimination. Each TipTopTalk! teaching strategy has its visual user interface containing different game elements. Figure 3 shows three visual user interface screenshots of the main game modes, that is, exposure, discrimination and pronunciation.

Gamification is an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement[6]. In TipTopTalk! users add points to their *phonetic level* and reach several achievements dependent on the mode and difficulty level (see Figure 2 (b)). There are also different language-dependent leaderboards, based on scores attained and the number of completed rounds, where all players are ranked to increase engagement through competition (see Figure 2 (a)).

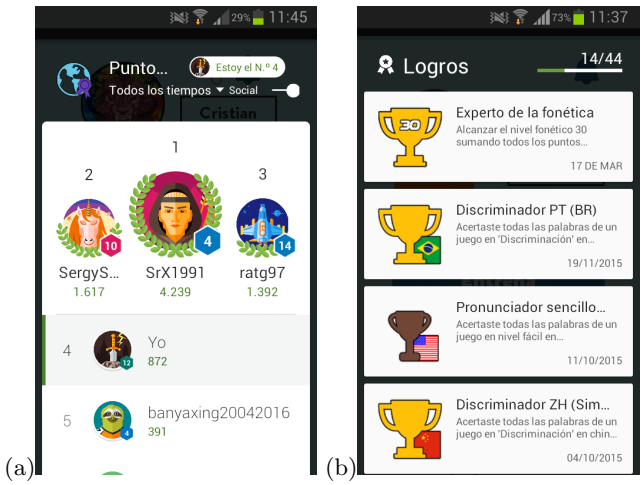


Fig. 2. Examples of gamification elements: a leaderboard (a) and a list of user's trophies (b)

Sharing results via social networks plays an important role in the gamification strategy by virtue of the competitiveness that it promotes. There are other gamification elements such as a limited time to complete the current round or a game; the granting of more or less points depending on the difficulty level and the number of attempts required for completion; the allotting of a number of reserve lives to allow further playing; the dispensation of an amount of *clear tickets* which allow users to skip the current round and move on to next one; and the

graphical display of the visual percentage of a game list result. Finally, we incorporate a system of push notifications that sends motivational and challenging messages to users in order to trigger their engagement.

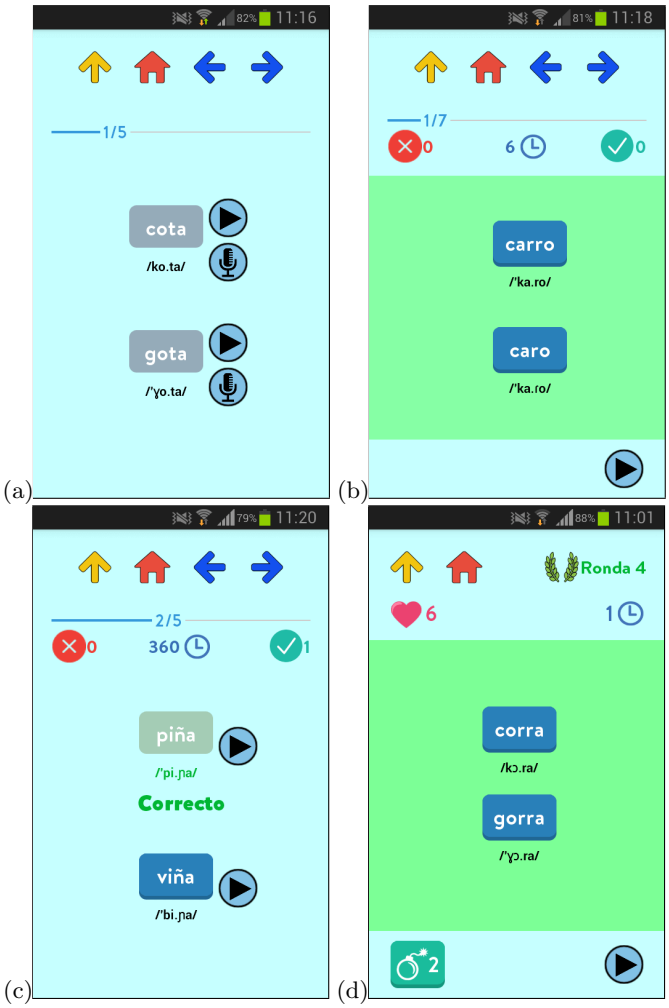
3 Activities in the demonstration

The demonstration will consist on an interactive session showing all different modes in the client application (see 2.2). People will be able to ask for help during the presentation. At the beginning, all attending people can download the application with a given URL or taking a photo of a QR picture. Once downloaded, the demonstration begins choosing the Spanish language. The first step is to complete an exposure activity, listening to and repeating all words. The first image (a) of Figure 3 shows a basic round of the exposure training mode. There is a menu-options bar at the top in which users can exit the current game, go forward to the next round or go back. There is also a status bar below the menu-options bar that indicates to users the current round. The system allows us to register whether users play the model for both words at the beginning of each round. Orthographic forms and phonetic transcriptions are displayed at the center of the screen. We keep track of the number of times users synthesize a word or record themselves. We save the recorded voice in a file for subsequent analyses and corpus compilation.

The second screenshot (b) of Figure 3 (discrimination mode) includes new elements such as a timer at the top and both discrimination wrong and correct counters. There is a background colour as a gamification element. If the colour is green, users must choose the word they think is being played. However, if the background colour is red, they must choose the wrong one. In the right bottom corner there is a button that plays another time the sound of the word.

The third screen capture (c) in Figure 3 represents a snapshot of a pronunciation mode round. This part of the game introduces more feedback elements than the previous. When the user utters the test word correctly, the related elements change their base color to green, and the word gets disabled as a positive feedback message appears. Otherwise, a message appears containing the words recognized by the ASR (different from the test word) together with a non-positive feedback. The mispronounced word changes its base color to red and remains active before it gets disabled only after five unrecognized realizations by the user. There is a limit of five wrong attempts per word.

The last screenshot (d) represents a round of *Infinite Mode* with the variant of the discrimination mode. The aim of this mode is to complete the highest number of rounds possible. There are new elements such as number of remaining lives at the left-top corner, the current round at the top-right corner and a skip-rounf button at the left-bottom corner. Discrimination and pronunciation challenges are presented randomly in each round. Users start with a finite number of lives that will decrease in one each time they fail. Also, the game's difficulty level increases with each round. For instance, from the tenth round on, the chance that the orthographic representation a word is substituted by asterisks is raised



(a) (b) (c) (d)

Fig. 3. Visual user interface of exposure (a), discrimination (b), pronunciation (c) and *Infinite* (discrimination variant) (d) modes.

to 50%. From the twentieth round on, a 50% chance that the TTS button is absent is introduced. The amount of time allotted for round completion is also progressively reduced.

4 Conclusions and future work

In this demonstration we presented a serious game implemented by a mobile application leaning on third party services. The main goal of our system is to provide a tool for improving L2 pronunciation with gamification elements. The client application was developed for Android version 2.3.3 and using the Eclipse development environment. On the one hand, it connects to an own web server. It works under a GNU/Linux operating system gathering data such as log files, messages and picture files. On the other hand, it relies on several Google services, for instance Google Voice Search, Google Analytics and Google Play Games.

TipTopTalk!'s dependence on both external ASR and TTS systems for assessing speech production may be a long-term problem since they are black-box systems. We are considering the possibility of using other open source platforms or creating a new one adapted specifically.

There are some points that can be improved in future versions. We are now working on some international collaborations to expand the range of available languages. We are also working in the portability to other mobile operating systems. Finally, despite the introduction of gamification elements, an habituation factor leads to a fall in interest and performance after protracted use. This suggests us to be able to incorporate mechanisms to provide real particularized feedback based on automatically identified errors.

Acknowledgements. This work was partially funded by the 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. We would like to thank Andreia Rauber, Anabela Rato and Junming Yao for their contribution of the minimal pairs lists.

References

1. Campbell, S.W., Park, Y.J.: Social implications of mobile telephony: The rise of personal communication society. Sociology Compass 2(2), 371–387 (2008)
2. Celce-Murcia, M., Brinton, D.M., Goodwin, J.M.: Teaching pronunciation: A reference for teachers of English to speakers of other languages. Cambridge University Press (1996)
3. Cámara-Arenas, E.: Native Cardinality: on teaching American English vowels to Spanish students. S. de Publicaciones de la Universidad de Valladolid (2012)

4. Escudero-Mancebo, D., Carranza, M.: Nuevas propuestas tecnológicas para la práctica y evaluación de la pronunciación del español como lengua extranjera. Actas del L Congreso de la Asociación Europea de Profesores de Espanol, Burgos (2015)
5. Escudero-Mancebo, D., Cámara-Arenas, E., Tejedor-García, C., González-Ferreras, C., Cardenoso-Payo, V.: Implementation and test of a serious game based on minimal pairs for pronunciation training. SLaTE-2015 pp. 125–130 (2015)
6. Kapp, K.M.: What is Gamification? The Gamification of Learning and Instruction: Gamebased Methods and Strategies for Training and Education, San Francisco, CA: Pfeiffer 13, 1–24 (2014)
7. Kartushina, N., Hervais-Adelman, A., Frauenfelder, U.H., Golestani, N.: 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(2), 817–832 (2015)
8. Linebaugh, G., Roche, T.: Evidence that L2 production training can enhance perception. Journal of Academic Language & Learning. 9(1), A1–A17 (2015)
9. McFarlane, A., Sparrowhawk, A., Heald, Y.: Report on the educational use of games. TEEM (Teachers evaluating educational multimedia), Cambridge (2002)
10. Muntean, C.I.: Raising engagement in e-learning through gamification. In: Proc. 6th International Conference on Virtual Learning ICVL. pp. 323–329 (2011)
11. Tejedor-García, C., Cardenoso-Payo, V., Cámara-Arenas, E., González-Ferreras, C., Escudero-Mancebo, D.: Playing around minimal pairs to improve pronunciation training. IFCASL (2015)
12. Tejedor-García, C., Cardenoso-Payo, V., Cámara-Arenas, E., González-Ferreras, C., Escudero-Mancebo, D.: Measuring pronunciation improvement in users of CAPT tool TipTopTalk! Interspeech pp. 1178–1179 (2016)
13. Tejedor-García, C., Escudero-Mancebo, D., Cámara-Arenas, E., González-Ferreras, C., Cardenoso-Payo, V.: Improving L2 production with a gamified computer-assisted pronunciation training tool, TipTopTalk! IberSpeech 2016: IX Jornadas en Tecnologías del Habla and the V Iberian SLTech Workshop events