Japañol, a Computer Assisted Pronunciation Tool for Japanese Students of Spanish Based on Minimal Pairs

T. Kimura[1], C. Tejedor[2], M. Machuca[3], A. Ríos[3] and D. Escudero[2]

[1] Seisen University, [2] Universidad de Valladolid, [3]Universitat Autònoma de Barcelona

Keywords— Computer Assisted Pronunciation Training, mobile application, gamification, L2 Spanish, L1 Japanese

I.   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) [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. The combination of adequate teaching methods and gamification strategies are expected to increase user engagement, provide an adequate feedback and, at the same time, keep users active and comfortable [2, 3].

This study describes the "Japañol" mobile application, a specific and controlled version of TipTopTalk! [4], a serious game for anywhere anytime self-learning, especially designed for Japanese learners of Spanish as a foreign language, that allows users to train and to test their pronunciation skills using their own Android mobile phones or Windows PCs.

II.   BRIEF DESCRIPTION OF THE SYSTEM

A. General overview

Three main elements are involved in our system: an Android client application, an own web server and external services provided by Google. Fig. 1 represents the conceptual architecture of the Android client application. We use both Google's free ASR and TTS system. 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 with which we feed a daily log for each user in order to determine whether her/his pronunciation skills are improving.

B. Minimal pairs

The lessons are made up with minimal pairs considering the specific problems in the phonemic contrasts produced by Japanese speakers in Spanish as a foreign language. The current version of Japañol consists of 7 lessons, each one of which treats one contrast and provides 12 minimal pairs. These are: (1) /s/:/θ/ as in *casa* "house": *caza* "hunting"; (2) /θ/:/f/ as in *zumo* "juice": *fumo* "I smoke"; (3) /fu/:/xu/ as in *fuego* "fire": *juego* "play"; (4) /l/:/ɾ/ as in *pelo* "hair": *pero* "but"; (5) /l/:/rr/ as in *lisa* "grey mullet": *risa* "laugh"; (6) /ɾ/:/rr/ as in *pero* "but": *perro* "dog"; and (7) /fl/:/fɾ/ as in *flotar* "to float": *frotar* "to rub." We focus in these minimal pairs because they are recognized as especially challenging for Japanese L1 speakers and in order to keep a reasonable total time for the completion of the full training activity.

C. Structure of each lesson

The user can choose any of the 7 lessons to start with, but once s/he has entered a lesson, the order of five modes (Explanation, Exposure, Discrimination, Pronunciation and Mixed mode) must be followed. In the Explanation mode, the articulation of each of the consonants that constitute the minimal pair is explained in Japanese by read explanation from the speaker, written explanation on the screen, animation of sagittal section of the articulators and by videos taken in front of the face of a native Spanish speaker pronouncing the consonants in question. The learner listens to the sound s/he is learning, reads and listens to the explanation, watching the animation and the video.

Having acquired the knowledge of how to pronounce each of the target sounds, the user moves on to the Exposure mode. Here the learner listens to sample pronunciation of some minimal pair words with the same two target sounds, like *cede* "s/he transfer": *sede* "venue"; *vez* "time": *ves* "you see"; and *Díaz* "a surname": *días* "days," generated by the TTS system. The learner is asked to record her/his own pronunciation and to compare it with the sample pronunciation by listening to both sounds.

In the Discrimination mode, the learner listens to the native pronunciation and matches it with the correct written form on the screen. In the Pronunciation mode, the learner reads aloud and records the word that appears on the screen. The recording is judged

by Google's free ASR system and a real-time feedback (the recognized word and a pronunciation hint) is provided instantly. There is a limit of five wrong attempts per word to avoid discouragement. After three consecutive failures, users can listen to the word as feedback. Fig. 2 shows the real-time feedback offered to the user in case of bad pronunciation. Native model pronunciations of each word can be played as many times as the learner needs. Speech is recorded and played using the ASR and TTS applications. Finally in the mixed mode, the learner faces discrimination and pronunciation challenges randomly, in order to consolidate acquired knowledge and skills. Users are guided by the system in order to complete all training modes. When reaching a score below 60% in Discrimination, Pronunciation or Mixed mode, users are recommended to go back to Exposure mode as a feedback resource and then return to the failed mode.
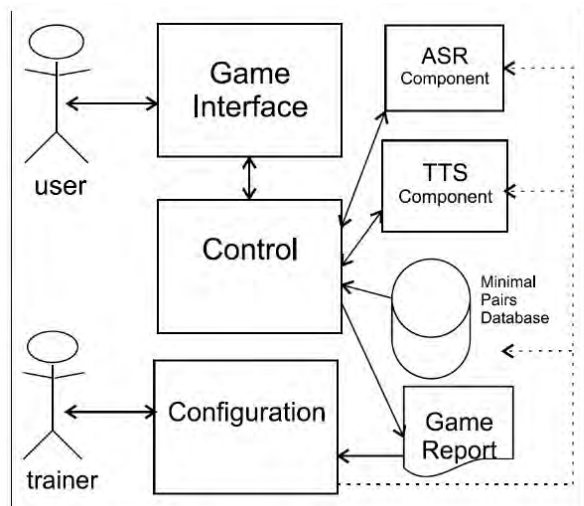

Fig. 1. Conceptual components of the client's system.

Fig. 2. Captured screen of the Pronunciation mode providing real-time feedback.

III.   CONCLUDING REMARK

We have pilot tested this program with a group of 8 native Japanese and obtained satisfactory results. Currently we are collecting data from 25 Japanese university students learning Spanish as their L3 (after Japanese and English), and some preliminary and promising results will be presented at the symposium. We are also planning to improve our system so that it may better meet the needs of the users, by including lessons for some prosodic features, improving individualized corrective feedback, developing the application for other platforms such as iPhone, etc.

REFERENCES

[1] D. Escudero-Mancebo and M. Carranza, "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 Español, Burgos, 2015.

[2] A., McFarlane, A. Sparrowhawk and Y. Heald, "Report on the educational use of games", TEEM (Teachers evaluating educational multimedia), Cambridge, 2002.

[3] C. I. Muntean, "Raising engagement in e-learning through gamification", Proceedings of 6th International Conference on Virtual Learning, 2011, pp.323-329.

[4] C. Tejedor-García, D. Escudero-Mancebo, E. Cámara-Arenas, C. González-Ferreras and V. Cardeñoso-Poyo, "Measuring pronunciation improvement in users of CAPT tool TipTopTalk!", InterSpeech 2016, pp.1178-1179.