

Evaluating Improvements in Speaking Ability in Passive Speakers Aided by a Virtual Avatar

Maria Samson*

Author

ABSTRACT

This paper discusses computer assisted language learning through related works and by also discussing a prototype which would further explore the effects of learning Spanish through a virtual agent.

1 INTRODUCTION

Digital language learning services have been steadily rising in popularity [5]. After a certain point, however, it is not necessary to continue memorizing new vocabulary, but to enhance one's conversational speaking abilities. The best method of accomplishing this is through simple practice – conversing with someone fluent in the language [3]. Most people, for whatever reasons, do not have access to a tutor that can evaluate their skills and help them improve. It is for these reasons that this proposal puts forth an application that assists in language learning by engaging with the user through various scenarios in the selected language, where the user will be able to respond using a microphone. We hypothesize that participants will have increased accuracy and confidence in their Spanish-speaking abilities as a result of conversing with a virtual avatar over 2 sessions. For the purposes of this prototype, only the Spanish language will be used.

2 MOTIVATION

One of the most effective ways to quickly learn a new language is through immersion and practicing speaking [3]. While the most effective way to practice this is by finding a fluent partner [6], unfortunately many foreign language learners do not have access to a native-speaker with whom to practice conversing with. Others are ashamed of their current speaking abilities, and would prefer to practice alone to build some preliminary proficiency. This is especially prevalent in passively bilingual speakers – those who have native or near-native listening comprehension, but limited speaking abilities, who happen to make up the participant pool of the experiment being conducted [4]. A solution to this is through the use of virtual conversation partners, where the user speaks into a microphone and the bot responds accordingly. But what is the purpose of this project beyond creating a tool? The aim of this project is to determine if participants are able to retain information, and perform at an equal or greater level than their first trial.

3 STATE-OF-THE-ART

Among the first iterations of a product that assists in more interactive language learning was the development of audio tapes. Rosetta Stone is one such brand that plays a sample sentence before prompting the user to respond. One glaring issue with this is that the speaker receives no feedback on their pronunciation whatsoever. Language tapes are quite limited in their interactivity, which is something that has been remedied thanks to smartphones' built-in microphones. Currently, there are several similar language-learning applications currently available, such as Duolingo and Babbel, both of which

offer speech practice, but only in the form of pronunciation practice, rather than through quick-thinking conversation [1]. Duolingo does have a well developed method for perfecting pronunciation, something that most technologies in this domain lack thus far, but is similarly lacking in that users cannot maintain a prolonged conversation with it [1]. Spanish in particular has been quite heavily studied in terms of learning a second language [8], which is helpful for the purposes of this experiment, as there is existing research to build upon. It has been established that computer assisted language learning provides users with entertainment and novelty, allows for more easily accessible interaction, and provides learners with the chance to practice their skills in a slightly more natural setting [9]. While it is clear that robots have a better aptitude for assisting in language learning [7] [2] – at least in children – we hope to explore if there is a similar, though perhaps less pronounced, effect with a virtual agent instead.

4 METHODOLOGY

4.1 Design

The core component of this project is a virtual agent that speaks to the user in a casual way by mimicking a real conversation one would have with a friendly acquaintance. This agent will be henceforth referred to as Ana. The user will select a language to begin practicing at the beginning of the conversation, to begin, this language will be Spanish. Once a language has been selected, Ana appears and introduces itself to the user before then asking the user questions about themselves in a polite conversational manner. These questions are determined to be generic enough and at an appropriate level that the user will likely have some vocabulary sufficient to answer the question. For each question or prompt, the user will be instructed to respond using a provided microphone as the input mechanism. This microphone will not be registered by Ana, but will allow the experiment proctor to respond using a Wizard of Oz approach. The proctor will command Ana to respond using a variety of prerecorded, computer generated sentences, provided that what the user said made sense, otherwise it will politely inform the user that it did not understand, and if they could please repeat that. After several unsuccessful attempts from the user in a row, Ana will simply change the subject. The user will use whatever computer they have at their disposal to interact with Ana, which will be screen-shared via the experiment proctor.

The prototype of Ana will appear quite simply: Ana, a cartoon-looking woman, is standing in what appears to be a cafe (see Figure 1). We decided to use a two-dimensional virtual avatar in order to avoid accidentally entering the "uncanny valley" with her as it is easier to avoid in 2D [10]. Accidentally entering the uncanny valley would inevitably have confounding effects on the experiment due to participants feeling uncomfortable from her uncanniness. The presence of the agent alone is what we are studying, so there is no need to include fancy detail or graphics animations. She will have a few preset facial expressions (happy, confused, neutral), but nothing beyond that. She is operated by the experiment proctor using the keyboard as a soundboard to play prerecorded, computer-generated questions, answers, and responses, unbeknownst to the participant.

*e-mail: samsonmn@rams.colostate.edu

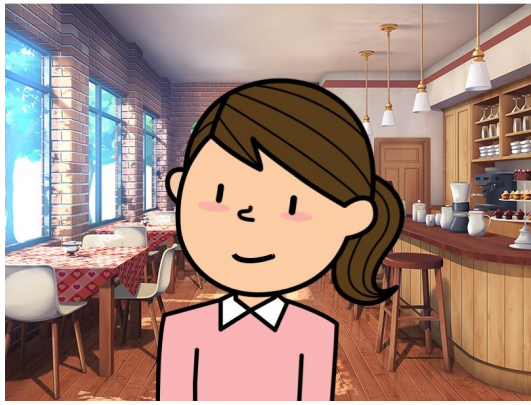


Figure 1: An image of how Ana appears to the user.

4.2 Evaluation

4.2.1 Participants

A total of 10 participants were used in this study, of ages varying between 13-24 (average age $M=17.8$ years, $SD=3.5$ years. 6F, 4M). They are all relatives of the experimenter (siblings, cousins, cousins' children), and are all passively bilingual due to having at least one Spanish-speaking parent – meaning they have native or near-native listening/comprehension fluency, but limited speaking ability. Given the participants' similar upbringings, they also have largely similar vocabularies deriving from northern Mexico, which adds an added layer of control to the experiment. It also allows us to evaluate based solely on speaking ability, as there are extremely low chances of speech errors being attributed to inaccurate comprehension.

4.2.2 Measures

Speaking improvements were measured through a survey and observations recorded over two conversation sessions one week apart. In addition to measuring changes in participants' perception of their speaking abilities, we also measure their gains through quantitative aspects listed below. These measurements were gathered by recording each session (with consent) then going back and counting/calculating each point individually. These sessions sought to evaluate various aspects of their speaking abilities, including:

- response time; time between Ana finishing her statement/question and user response. Taken for responses greater than 2 words or longer than 1 second.
- rate of speech; user's spoken words per minute per response. Taken for responses greater than 2 words or longer than 1 second.
- speech error rate; severe pronunciation or grammatical errors, measured by how often "Ana", in this case the experiment proctor, does not understand the user.
- comprehension error rate; when the user does not understand something Ana says, and/or responds with something irrelevant to the subject (as determined by the experiment proctor). This measure will likely be extremely low, given the participants' passive bilingualism, however it would be relevant given a group of non-native participants.

4.2.3 Conditions

The independent variable here is the vehicle in which users are practicing their Spanish speaking abilities, in this case via Ana.

The dependent variables were discussed above in Measures. Control variables include; regional dialects (northern Mexico), approximate comprehension ability (native/near-native), upper threshold for speaking ability (participants never or rarely speak Spanish), and method through which the session is administered (over Discord screen-sharing). Random variables include; age, gender, geographical location, monitor-size, Ana's volume, and anything else not specified as a control variable.

4.2.4 Procedure

Over a Discord call, after making general small talk, the experimenter explained the objective of this experiment and outlined the procedure. The participant then filled out a survey logging their current Spanish abilities. This survey was used to exclude participants with speaking fluency over the desired threshold for this experiment. The survey logged the user's self reported (1-5) Spanish listening comprehension skills, Spanish speaking skills, Spanish reading skills, frequency in which they listen to Spanish, frequency in which they speak in Spanish, and frequency in which they read in Spanish.

Using Discord's screen-sharing feature, the experimenter pulled up Ana and demonstrated what she sounds like. This also allowed the participant to make any necessary changes to their computer system, such as increasing the volume. Once the participant indicated they were ready, the proctor began using Ana to ask questions to the user, and generally engage in conversation. When Ana's conversation options were exhausted, or the participant indicated they wished to end the conversation, the participant was then asked to fill out the same survey as before. One week later, the procedure was repeated, using different prerecorded questions and answers from Ana, but which were approximately the same level of difficulty to respond to.

5 RESULTS AND DISCUSSION

Upfront, there is the obvious problem of the experiment's small number of participants, therefore these results should be taken with a grain of salt, as they are representative of only a small niche portion of Spanish speakers. One of the questions that confused several of the younger participants was "¿Qué tal?" Apparently this is a lesser used phrase in the contexts they are used to, and they have never heard it before.

5.1 Self-Reported Speaking Ability

Self-reported speaking ability gains are measured on a scale of 1-5 using the participants' responses from the survey administered before and after each session, where 1 is low speaking ability and 5 is high speaking ability. The first post-session indicated a trend toward the median of 2, with one respondent decreasing from a 3 to a 2, many of those who responded 1 increasing to a 2, and all those who started with 2 maintaining that score. The pre-session results were nearly the same. From the end of session 1 to the beginning of session 2, respondents tended to revert to their initial responses. The post-session results from session 2 were markedly higher than all previous responses. This supports the hypothesis that conversing with Ana will result in increased confidence in a participant's speaking abilities.

5.2 Measured Speaking Ability

The changes in measured speaking ability neither support nor disprove the hypothesis, as there is an expected change in error rate, but an unexpected change in rate of speech, with both results being relatively statistically insignificant.

5.2.1 Error Rate

Error rate is measured by comparing the total number of words spoken against the number of erroneous words. As seen in Figure 3, general there is a slight decrease in the error rate overall, with the

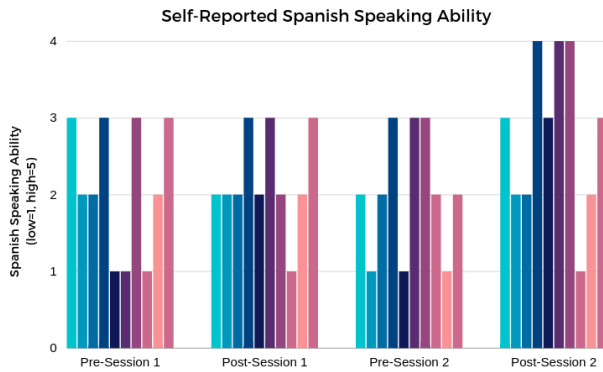


Figure 2: Change in participants' self-reported speaking ability before and after 2 sessions.

median difference between session 1 and session 2 being -1.56. Only one participant has a notably higher error rate in session 2, which even then was not a large margin, with most participants maintaining or improving their accuracy. This suggests that between sessions 1 and 2, participants may have made some sort of effort to become more aware of the vocabulary being used around them, or that they made an effort to improve their accuracy in terms of vocabulary or grammar.

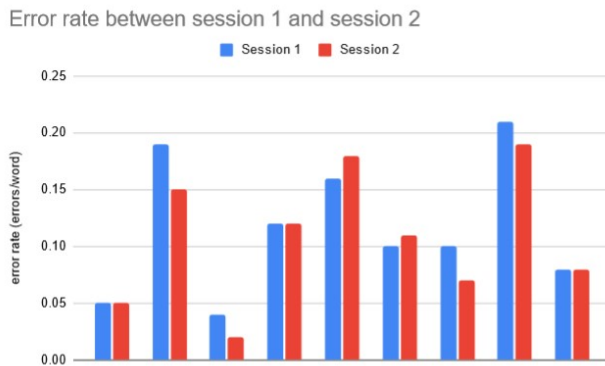


Figure 3: Change in participants' self-reported speaking ability before and after 2 sessions.

5.2.2 Rate of Speech

The average rate of speech is measured by comparing each response's words per second, and then taking the average of all of a participant's responses. As shown in Figure 4, there is virtually no change in average rate of speech, with the median difference between sessions 1 and 2 being -0.02. We expected the rate of speech to increase, however the median between sessions 1 and 2 staying nearly unchanged. There appears to be inconsistent results between participants, with some increasing their rate of speech, and other decreasing, but none really maintaining.

6 CONCLUSION

This paper discussed the results of an experiment that sought to determine the effects of conversing with a virtual agent named Ana on a group of passively bilingual Spanish-speaking participants. To do this, we had the participants fill out a survey to report their current speaking abilities before and after each session of conversation, of

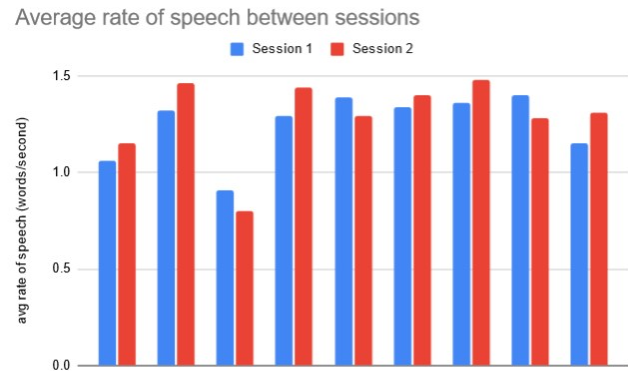


Figure 4: Change in participants' self-reported speaking ability before and after 2 sessions.

which there were 2 total. We found that there was a small increase in the confidence of the participants after the 2 sessions, but there were inconsistent results that indicated minimal to no increase in actual measured speaking ability. Future work should probe whether the results of this experiment can be translated to a group of participants who do not come from Spanish-speaking households.

7 ACKNOWLEDGEMENT

Special thanks to my mother, Susana Ramirez-Aragon, for obtaining and relaying the contact information of each of the participants.

REFERENCES

- [1] J. Bajorek. L2 pronunciation tools: The unrealized potential of prominent computer-assisted language learning software. *Issues and Trends in Educational Technology*, 5:60–87, 2017. doi: 10.2458/azu.itet.v5i1_bajorek
- [2] R. Berghe, J. Verhagen, O. Oudgenoeg-Paz, S. van der Ven, and P. Leseman. Social robots for language learning: A review. *Review of Educational Research*, 89:259–295, 2018. doi: 10.3102/2F0034654318821286
- [3] S. Burger, A. Weinberg, and M. Wesche. Lessons learned from implementing french immersion studies program at the university of ottawa. *Integrating Content and Language in Higher Education*, 6:21–43, 2015. doi: 10.18192/olbiwp.v6i0.1130
- [4] R. Chumak-Horbatsch. Early bilingualism: Children of immigrants in an english-language childcare center. *Psychology of Language and Communication*, 12:3–27, 2008. doi: 10.2478/v10057-008-0001-2
- [5] H. effective are mobile. Y.-t. sung and k.-e. chang and j.-m. yang. *devices for language learning? A meta-analysis*, 16:68–84, 2015. doi: 10.1016/j.edurev.2015.09.001
- [6] S. Emde, J. Schneider, and M. Köter. Technically speaking: Transforming language learning through virtual learning environments (moos). *The Modern Language Journal*, 85:210–225, 2002. doi: 10.1111/0026-7902.00105
- [7] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme. Social robot tutoring for child second language learning. *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016. doi: 10.1109/HRI.2016.7451757
- [8] H. A. Murillo-Reyes. Factores cognitivos que influyen en la adquisición del español como segunda lengua. *Revista Electrónica de Conocimientos, Saberes y Prácticas*, 8:159–169, 2020. doi: 10.5377/recsp.v3i1.9798
- [9] D. Ratnaningsih, F. Nofandii, D. Purba, and D. Wiratno. The influence of computer assisted language learning (call) to improve english speaking skills. *Research, Society and Development*, 8, 2019. doi: 10.33448/rsd-v8i10.1413
- [10] M. Seymour, L. Yuan, A. Dennis, and K. Riemer. Crossing the uncanny valley? understanding affinity, trustworthiness, and preference for more realistic virtual humans in immersive environments. *Proceedings of*

the Annual Hawaii International Conference on System Sciences, 2019.
doi: 10.24251/HICSS.2019.213