Simicry - A mimicry-feedback loop for second language learning

Preben Wik, Björn Granström

KTH, Centre for Speech Technology CTT, Stockholm, Sweden

`preben@kth.se, bjorn@speech.kth.se`

Abstract

This paper introduces the concept of Simicry, defined as similarity of mimicry, for the purpose of second language acquisition. We apply this method using a computer assisted language learning system called Ville on foreign students learning Swedish. The system deploys acoustic similarity measures between native and non-native pronunciation, derived from duration syllabicity and pitch. The system uses these measures to give pronunciation feedback in a mimicry-feedback loop exercise which has two variants: a 'say after me' mimicry exercise, and a 'shadow with me' exercise.

The answers of questionnaires filled out by students after several training sessions spread over a month, show that the learning and practicing procedure has a promising potential being very useful and fun.

Index Terms: CALL, CAPT, Automaticity, prosody, implicit learning

1. Introduction

There is today compelling evidence that language depends on brain systems that are also used for other functions. The declarative/procedural (DP) model predicts common computational, processing, and anatomic neural substrates for language and non-language functions.[1] This means that general theories of skill acquisition would also be applicable to skill building in language learning. Perhaps the best known general theory of skill acquisition is Anderson's adaptive control of thought (ACT) [2]. The ACT theory describes how declarative knowledge after repeated practice, develops into proceduralized knowledge that can be used to solve a task without accessing long-term declarative memory. Performance speed and accuracy improve as the learner implements these production rules and automaticity develops. Studies on vocabulary and grammar acquisition [3] [4] have shown that general models of skill acquisition also apply to automaticity in second language (L2) acquisition. We assume that the same general principles would also apply to the development of automaticity in pronunciation proficiency in second language acquisition (SLA). Automaticity is the performance of a skill without conscious control. When someone speaks their first language it is to a large extent an automatic process. We often don't think of which words to use or which grammatical constructs to apply, much less how to shape the mouth or move the tongue in order to create the right sounds, or where to place the stress or how to adjust the pitch for a sentence to sound right. An L2 learner on the other hand might struggle with all these aspects of a new language, and from this point of view, what the L2 learner is aiming at is to a large extent to develop automaticity. However, since promoting automaticity is generally believed to require massive repetition experiences and constant practice, it is something that is difficult for a language teacher to provide in a classroom setting, but well suited to do in a computer assisted language learning (CALL) application.

2. Simicry

As discussed above, it is assumed that extended practice, under particular conditions and circumstances, will increase fluency by developing automaticity.

Simicry, a term coined from "Similarity of mimicry", is a framework for practicing L2 pronunciation based on the idea that a mimicry-repetition-feedback-loop will facilitate exposure to large amounts of comprehensible, meaningful input, which according to the literature will promote an implicit process of acquisition, and the development of automaticity.

It is in all languages acceptable to vary the pronunciation of an utterance in many different ways, based on differences in dialect, personality, semantic intent, emotional content etc. The pronunciation may vary for example by placing focus on different words, changing the intonation (pitch), slowing down or speeding up parts of the utterance etc. Even though all languages show variation of this kind, every language will have its own regions of variation that are acceptable, and regions of variation that are unacceptable. Such language specific restrictions is something an L2 learner will initially be unaware of, and due to the fact that they cannot easily be described or understood in explicit terms, may be best taught as implicit knowledge.

We see a pedagogical point in limiting the students degrees of freedom in variation in this learning setting, and suggest that learning to reproduce, i.e. to mimic a native speaker, is an appropriate first step to acquire native-like pronunciation. Although being able to mimic well is not something a native speaker necessarily is good at, it can nevertheless be something that a language student can benefit from in order to get a good pronunciation in a new language. We hypothesize that trying to produce an utterance that as closely as possible corresponds to that of a model utterance, and getting feedback on similarity of mimicry, rather than giving students all the degrees of freedom that a native speakers have in their production, will implicitly force the student to adjust their pronunciation, and learn to pay attention to aspects of the language that would otherwise go by unnoticed.

In Simicry one might for example have one analysis method for measuring similarity in rhythm, and another for similarity in melody (pitch). In an implementation of this where a student is consistently weak in one aspect, say rhythm, one could turn off all other aspects of similarity measures, and only focus on the rhythm. In such a scenario the student could say whatever he or she wanted, using any phonemes and any pitch contour without affecting the score, as long as the rhythm was similar. If the nature of the analysis was such that one region of an utterance could be determined as problematic, one could zoom in on that region and loop over that part until the student had mastered it,

Figure 1: *The Simicry module in Ville*

then zoom back out again and loop over a larger region or the whole utterance.

Exactly which similarity measures that are relevant or salient for Simicry is yet to be determined. Even though the pitch, and duration of an utterance in absolute terms, does not carry linguistic meaning, (but changes in both pitch and duration might), there could be pedagogical benefits in telling the students to focus on getting the length of the whole utterance similar to the target utterance in order to make them focus and concentrate on other aspects of speech than they usually do. The framework is intended to be open for different kinds of similarity measures and allow for different researchers to experiment on the effect that feedback of a new similarity measure could have on student performance. Even though we see the feedback as an important part, it is the massive exposure to meaningful input and the implicit learning generated by the looping construction that is the main focus of the paradigm.

3. Simicry Experiment

Ville is a virtual language teacher for Swedish, developed at The Centre for Speech Technology (CTT), at KTH. It uses an Embodied Conversational Agent (ECA) as a key feature, exploring human computer interaction using the person metaphor rather than the desktop metaphor used in most other computer programs (thus when mentioning Ville in the text, it is sometimes referring to the software in general, and sometimes to the *ECA or person* Ville. A first version of Ville was offered in the fall of 2008 to all foreign students at KTH [5]. Ville has since then been downloaded by approximately 1200 students, who have done more than 5500 perception exercises, 1500 writing exercises, and 15000 recordings. The first version focused on helping students with vocabulary training, providing a model pronunciation of new words and drilling students in memorization exercises. Since then research has focused on developing new interfaces and feedback mechanisms on pronunciation and perception exercises. Some of them are based on explicit phonetic knowledge, offering the students explanations on specific phonetic aspects of the language in understandable terms. For learning the prosody of a foreign language we have created Simicry exercises, since we believe they could be well suited

to train implicit knowledge of not easily formalized aspects of a language.

A test of 8 new capabilities of Ville was done in the spring of 2009 [6], where an early implementation of Simicry was one of the capabilities tested. This paper will report on the Simicry part of a follow up experiment with similar exercises, but where the students had the possibility to work with Ville at home, over a longer period of time, and where a more comprehensive study of the Simicry paradigm, and its implementation in Ville was investigated.

3.1. Simicry in Ville

The Simicry module in Ville, shown in figure 1, has the following features: The interaction is such that once the play button is pushed a mimicry-repetition-feedback loop is entered, and the students only interact through speech. Students may select two different ways to do the actual mimicry interaction. *Say-after*, where Ville says the sentence first, and the students repeat it, and *Shadowing*, where Ville and the student speak at the same time. In *Shadowing*, a count down tick is preceding every utterance to help students start at the same time as Ville. Students may select level of difficulty: If the student's performance is above a certain threshold, Ville moves on to the next sentence; otherwise, the same sentence is presented again, until the student has repeated it successfully. An acceptance-rate slider enables the students to choose this threshold themselves. This adds an aspect of gameplay to the exercise, where students should find a level on the acceptance-rate slider that makes it hard, but not too hard for them to pass, and then try to increase the level of difficulty over time. Students may select what packages to work with. A package is a collection of sentences with a semantic topic, such as: At the bar, at the airport, asking for directions, party conversations etc. There were 11 different packages available for the students to choose from, each package containing 30-50 sentences.

3.2. Measuring similarity

The version of Simicry in Ville implemented in this experiment puts focus on prosody, to investigate if the students will notice or change the prosodic realizations of their utterances. The

score presented to the student as that which is compared with the acceptance-rate slider (which determines whether to loop or move to the next sentence), is calculated as an average of four measures:

- Psyllabicity (a measure of pseudo syllabic units, resembling that of the convex hull algorithm [7]). this is a measure of how many syllables there are in each utterance and the score is calculated by comparing the number of syllables in the native (target) utterance and the student utterance

- Length (the total length of the utterance)

- Timing (The mismatch in duration of each syllable)

- Melody (a normalized F0 correlation score)

All scores result in a measure between 0-100 and the final score is the average of these four scores. This final score is then compared with the acceptance-rate threshold, which determines if the student should move on to the next sentence, or loop (mimic the same utterance again).

3.3. User Test

A group of 42 foreign students (16 female and 26 male) at KTH, studying Swedish as a second language on elementary and advanced elementary levels, volunteered to participate in the study. The language background (L1) of the students is shown in table1.

Table 1: *Language background (L1) of the 42 participants.*

Country	Participants	Country	Participants
Persian	6	Korean	2
English	6	French	1
Chinese	5	German	1
Spanish	5	Italian	1
Russian	4	Lithuanian	1
Turkish	3	Urdu	1
Greek	2	Vietnamese	1
Polish	2	Arabic	1

All students were brought to the CTT lab to do a pre-test that lasted approximately 30 minutes, and which consisted of several language learning exercises in addition to the Simicry exercise being presented in this paper.

The same test was repeated as a post-test after one month of training where the students used Ville at home. All pronunciations produced by the students during the tests and the training were logged on a server at KTH.

In the Simicry task the students were instructed to mimic the sentences as close as possible, and pay special attention to rhythm, length, and intonation. The Simicry task in the pre and post test consisted of 56 phonetically balanced sentences, created as a message from Ville requesting assistance in buying necessary groceries for a weekend trip to a cabin.

After the initial pretest the students were randomly divided into three groups. Participants in group 1 and 2 got a version of Ville to work with at home during the following month, whereas group 3 was a control group, and had no access to the program. There were some differences in the Simicry exercise between the two versions of Ville that were given to group 1 and group 2. Group 1 only worked in *shadowing* mode, and group 2 only worked in *say-after* mode.

4. Analysis and results

Questionnaires were given to the participants in group 1 and group 2 after they had come again to the CTT lab and completed the post test, after one month of training at home. Questions that were relevant for the Simicry part included:

1. Rate the Simicry part based on usefulness:
   1 for not useful at all - 5 for very useful

2. Rate the Simicry part based on enjoyment:
   1 for boring - 5 for fun

3. Rate the Simicry part based on difficulty:
   1 for very difficult - 5 for very easy

4. Do you think that you learned something from the exercises:
   1 for nothing - 5 for a lot

Figure 2 present the mean and standard deviation of the answers to the questionnaire by group 1, which had the *shadowing* version of the program. It is interesting to note that although all students rated the Simicry exercise as fun and useful, they also thought it was difficult, which might reflect the nature of mimicking natively pronounced sentences on the fly. Figure 3 presents the same results for group 2, which used the *say-after* mode of the program. The results of group 1 and group 2 seem to reflect very similar opinions. In average students in group 2 thought that Simicry (in say-after mode) is slightly more useful and slightly more difficult than students in group 1, but the difference is not significant.
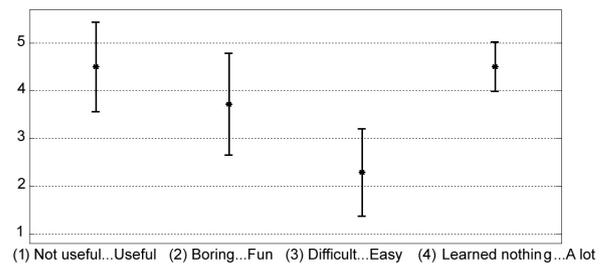
Figure 2: *Replies in mean and Std. to questions for group 1*
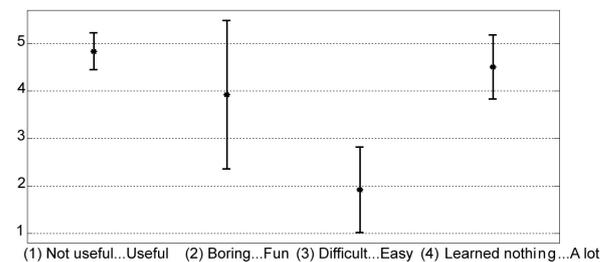
Figure 3: *Replies in mean and Std. to questions for group 2*

The students in group 1 also had a question regarding the preference for *say-after* or *shadowing*, since they had tried both types of interfaces. (*say-after* in the pre and post tests, and *shadowing* while working at home). The question was stated as:

5. You have tried two types of Simicry: Which do you prefer?
Shadowing: 1 for 'I love it' - 5 for 'I don't like it'
Say-after: 1 for 'I love it' - 5 for 'I don't like it'

The histogram of the answers for these two questions are presented in figure 4. The histogram clearly show a preference for the say-after version of Simicry compared to the shadowing.
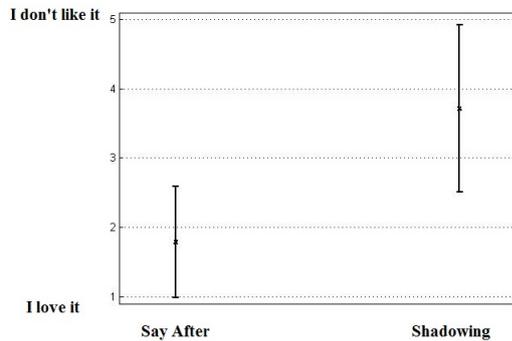
Figure 4: *Replies on preference for say-after or shadowing for group 1*

5. Discussion and Future work

A first peek by a human judge has been performed on the 4704 sentences collected in the pre-post tests (56 sentences x 42 students x 2 sessions). All sentences from all participants were randomized in a pairwise fashion, so the same sentence from the same speaker from the pre test and post test were kept together, but in a randomized order. These were then presented to a judge who could choose between three options: 1) stimuli 1 best, 2) stimuli 2 best, 3) no difference. There were large individual differences among the students, and all groups improved their performance between first and second session, but there were no significant differences between the groups. A more comprehensive listening test of the material, with a multiple judges, and a scaled judgment of the pronunciation is underway.

Although the Simicry implementation that was presented in this paper only gave feedback on aspects of prosody, there is nothing prosody specific about the Simicry paradigm, and we would like to explore it further with other aspects of similarity measures in mind. As discussed above, which measurements of similarity that are the most relevant for a student to get feedback on needs to be investigated further. The measures used in this experiment need to be optimized and several proposals from the literature, for example [8] [9] have been suggested. It should be noted however that good methods for classification of a linguistic feature does not necessarily entail good pedagogical feedback. An interesting question to investigate is: Should similarity measures be language specific, or even phrase, sentence, or word specific, or are there measures that are able to capture language specific variations using some universal features? However, we believe that the simicry-feedback loop paradigm might be pedagogically valuable even with pseudo arbitrary acoustic similarity measures, leaving the learning and abstraction implicit for the language learner herself.

Another, related issue in order to achieve maximum efficiency and retention from working with Simicry is how often, and when to introduce new material to the student. Whether it is vocabulary, or variants of pronunciation that is being trained, the frequency of repetition is not arbitrary. Practice should be scheduled according to some optimal spaced repetition algorithm in order to maximize learning and retention [10]

6. Conclusions

We presented in this paper some different aspects of a paradigm we call Simicry, defined as similarity of mimicry. We investigated some subjective effects of an implementation of Simicry, as part of the language teaching platform Ville. The implementation offered two versions of a mimicry-feedback loop, one with students mimicking sentences in Swedish while they were presented (Shadowing), and one with the students repeating after the sentence was presented (Say-after). The training software was introduced to 42 subjects, split into three groups, a control group, and two groups with the two different versions of Simicry. Students answered questionnaires after one month of training, and the results show clear opinions supporting the Simicry training method as useful and fun, although they clearly preferred the say-after version to the shadowing version. The implementation of Simicry presented in this paper deploys first step ideas on how acoustic similarity should be measured to provide informative feedback to the learner. These aspects are the focus of future work and experiments.

7. Acknowledgements

8. References

[1] M. Ullman, "The neural basis of lexicon and grammar in first and second language: The declarative/procedural model," *Bilingualism: Language and Cognition*, vol. 4, no. 02, pp. 105–122, 2001.

[2] J. Anderson, "ACT: A simple theory of complex cognition." *American Psychologist*, vol. 51, no. 4, pp. 355–365, 1996.

[3] R. DeKeyser, "Beyond explicit rule learning," *Studies in Second Language Acquisition*, vol. 19, no. 02, pp. 195–221, 1997.

[4] P. Robinson, "Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions," *Studies in Second Language Acquisition*, vol. 19, no. 02, pp. 223–247, 1997.

[5] P. Wik and A. Hjalmarsson, "Embodied conversational agents in computer assisted language learning," *Speech communication*, vol. 51, no. 10, pp. 1024–1037, 2009.

[6] P. Wik, R. Hincks, and J. Hirschberg, "Responses to Ville: A virtual language teacher for Swedish," in *Proceedings of the 2009 Workshop on Speech and Language Technologies in Education (SLaTE), Wroxall Abbey Estates, United Kingdom*, 2009.

[7] J. Edlund and M. Heldner, "/nailon/-Software for Online Analysis of Prosody," in *Ninth International Conference on Spoken Language Processing*. ISCA, 2006.

[8] S. Al Moubayed, G. Ananthakrishnan, and L. Enflo, "Automatic prominence classification in swedish," in *Proceedings of Speech Prosody*, Chicago, USA, 2010.

[9] J. Tepperman, T. Stanley, K. Hacioglu, and B. Pellom, "Testing Suprasegmental English Through Parroting," in *Proceedings of Speech Prosody*, Chicago, USA, 2010.

[10] P. Pavlik and J. Anderson, "Using a model to compute the optimal schedule of practice." *Journal of Experimental Psychology: Applied*, vol. 14, no. 2, pp. 101–117, 2008.