AUDITORY FEEDBACK METHODS TO IMPROVE THE PRONUNCIATION OF STOPS BY GERMAN LEARNERS OF FRENCH

Jeanin Jügler & Bernd Möbius

Computational Linguistics & Phonetics, Saarland University, Saarbrücken, Germany
juegler|moebius@coli.uni-saarland.de

ABSTRACT

The effect of manipulation of a speaker's voice as well as exposure to a native speaker's utterance was investigated regarding the pronunciation of stops by German learners of French. Three subject groups, a *Control (CG)*, a *Manipulation (MG)*, and a *Native Speaker (NG)* Group, were recorded on two subsequent days. The MG was presented with a manipulation of their voice on the second day and the NG listened to a native French speaker, while the CG did not receive any feedback. Results show that speakers of the MG and NG were able to extract useful information from the respective feedback and successfully adapted to it. Participants were able to reduce their voice onset time values, although speakers of the NG reduced it to a greater extent.

Keywords: non-native speech, French, German, feedback, stops.

1. INTRODUCTION

Speaking a foreign language calls for more than in-depth knowledge of grammar or vocabulary. However, this concept of language learning seems to represent the general state of thinking considering most second language (L2) teaching. Many L2 class rooms show only minimal effort to teach correct pronunciation or include specific pronunciation tasks to make learners aware of their problems and deviations [10, 18]. Since a poor pronunciation will make it harder to be understood by interlocutors [13] a special focus on pronunciation in L2 teaching is crucial. However, visualizing and explaining pronunciation problems sufficiently is a challenge for teachers. It is also problematic for the learners to perceive their own mistakes or deviations from the native pronunciation [1, 8]. This is especially true if phonetic and phonological knowledge of the native language (L1) interferes with the phonetic and phonological system of the foreign language [1, 8, 9]. As a rule, a learner is challenged by several phonological and phonetic differences between L1 and L2.

However, the attitude towards pronunciation teaching seems to change. The development of computer-assisted language learning (CALL) and especially computer-assisted pronunciation training (CAPT) systems has received increased interest over the years (e.g. Euronounce [6], Dutch-CAPT [14]). It was shown that feedback helps to improve pronunciation in a foreign language (e.g. [16]) although shortcomings of current CAPT systems are still characterized by the lack of good feedback [7].

An overview of several CAPT systems shows a variety of visual (e.g., vocal tract, waveform, articulation scores, pitch curve) and/or auditive (e.g., reference speaker, resynthesis of voice) feedback methods in order to display pronunciation errors (e.g. [6, 11, 14, 16]). Yet, some systems use feedback methods that are difficult to interpret (e.g. [6, 11]), such as waveforms, which might give the wrong impression that the learner's oscillogram has to look exactly like the waveform of the native speaker [15].

Manipulation of the learner's voice seems to be an efficient feedback method from which learners are able to extract useful information. Bissiri and Pfitzinger [2] investigated the effect of resynthesis of the learner's own voice on learning lexical stress in German by Italian native speakers and showed that resynthesis has a motivating effect.

The aim of this paper is to test the influence of altered auditory feedback of a learner's own voice on the production of French stops by German native speakers. Since learners have difficulties perceiving their own mistakes and their deviations from a native pronunciation [1], exposure to a native speaker's utterances was also tested.

German and French mark the distinction between voiced and voiceless stops /b d g p t k/ differently. French speakers differentiate between fully voiced plosives and voiceless unaspirated ones with a rather short Voice Onset Time (VOT). In contrast, German shows a distinction by voiceless unaspirated plosives with a short VOT and voiceless aspirated ones with a long VOT [12]. As a consequence of these differences, learners of German and French, respectively, are expected to transfer the phonetic knowledge of their native language to their production in the L2.

2. EXPERIMENT

The effect of manipulated feedback of a learner's voice and exposure to a native speaker's utterances was investigated regarding the pronunciation of stops by German learners of French. The experiment involved three subject groups. A *Control (CG)*, a *Manipulation (MG)*, and a *Native Speaker (NG)* Group were tested for a set of French and German sentences containing minimal pairs contrasting in word-initial stops. The CG subjects did not receive any feedback whereas the MG subjects were presented with their own manipulated voice throughout the experiment. The NG subjects listened to utterances of a French native speaker (female, 28, Strasbourg) who also served as the golden speaker for later comparisons.

Each experimental group consisted of five female and five male native German speakers (19-38 years, M: 23.7 years, SD: 3.9 years) with basic knowledge of French (A1-A2 level according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)). All participants were students at Saarland University.

2.1. Materials

For each sound contrast (/b-p/, /d-t/, /g-k/), seven French minimal pairs differing in syllable-initial position were embedded in a short sentence. These experimental targets were all nouns preceded by [e] (e.g. *les*) to ensure a consistent preceding segmental context. Furthermore, the stops of interest were directly followed by a vowel (see examples (1) and (2) for the contrast /g-k/ in French).

(1) *Les cages à oiseaux sont très petites.*
(The cages for birds are really small.)
(2) *Les gages sont payés à la fin du mois.*
(The wages will be payed at the end of the month.)

To facilitate a cross-language comparison, the same number of minimal pairs was recorded by the same learners for German. Each target word was preceded by [ə] (e.g. *weiße*) (examples (3) and (4)).

(3) *Der weiße Guss auf dem Kuchen ist lecker.*
(The white icing on the cake is delicious.)
(4) *Der erste Kuss ist etwas Besonderes.*
(The first kiss is something special.)

In addition to the experimental targets, four training targets for each stop in both languages were included. Training targets began with a stop and were followed by a vowel but were not minimal pairs. The complete set of French and German sentences (Table 1) was recorded by all three groups.

Table 1: *Number of German and French sentences included in the study.*

	experimental	training	$\sum$
French	42	24	66
German	42	24	66

2.2. Manipulation

Manipulation was carried out manually and pertained only to the duration of VOT. Since German speakers differentiate between voiceless unaspirated stops with a short VOT and voiceless aspirated stops with a long VOT, subjects most likely show a longer VOT for French plosives than French native speakers. Therefore, VOT had to be shortened based on values of a French reference speaker (golden speaker). Manipulation was carried out for the training set of sentences and was only applied if the aspiration was longer than the golden speaker's.

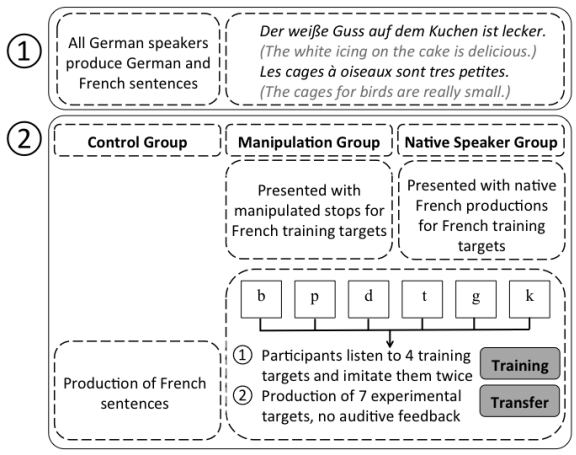
It can be argued that not only length of VOT is crucial for the production and perception of voiced and voiceless stops since in contrast to German, French voiced stops are characterized by a fully voiced closure. We decided against a manipulation of phonation of the closure. It is not straightforward to modify the signal to become voiced. A modification of the phonation is technically challenging and often yields unsatisfactory perceptual results. Preserving the natural stimulus quality was deemed to be of higher importance.

2.3. Procedure

Recordings took place on two subsequent days. They were made in quiet office rooms using a head-mounted microphone (16 kHz, 16 bit) on a M-AUDIO Fast Track USB device. Recordings were saved on a Windows Laptop using a custom-made software that was developed at LORIA ("Corpus-recorder", [5]). The sentences were presented to each speaker in a randomized order.

On the first day, German and French utterances were recorded as a baseline for later comparisons. The second recording session on the subsequent day differed for the three groups (Figure 1). The CG subjects were asked to read the set of French sentences from the previous day once again without receiving any additional information or feedback on their pronunciation. For the MG and NG, the structure of the recordings was modified. They received auditory

Figure 1: Overview of the study's procedure.



feedback, either a manipulated version of their own recording (MG) or a native French utterance (NG), and had to record the sentences again. To ensure that the subjects did not simply imitate what they heard but were able to transfer knowledge obtained from the feedback, the recording session was divided into a *training* and a *transfer* block.

In the training block, four training targets embedded in a sentence had to be imitated twice. Subsequently, participants moved on to the transfer block where they were asked to produce sentences including the experimental targets without any additional auditive presentation. Participants only worked on one stop at a time to allow them to develop a specific strategy and concentrate on it.

Since feedback is only useful if sufficient information is provided [4], the focus on stops was pointed out to the subjects before the second recording and target words were highlighted. Subjects were also generally informed that there is a difference between the pronunciation of French and German stops.

2.4. Hypotheses

The following predictions were made for VOT:

1. There is no significant difference between the first and second recordings for the CG subjects. Improvements from repetition only are expected to be small.
2. VOT values of the second recording will be significantly shorter than in the first recording for the MG and NG subjects due to an improvement induced by auditory feedback.
3. VOT in the second recording does not differ significantly from the French reference speaker for the MG and NG subjects.

Table 2: *Mean VOT values (ms) of the experimental groups for the first and second French recording and reference values of the golden speaker*

		Rec 1	Rec 2
Control	*voiced*	15	15
	voiceless	49	46
Manipulation	*voiced*	13	13
	voiceless	54	47
Native Speaker	*voiced*	11	12
	voiceless	51	38
Golden Speaker	*voiced*	3	
	voiceless	30	

3. ANALYSIS AND RESULTS

Duration of VOT was labeled using Praat [3]. The VOT of fully voiced stops, which is defined by negative VOT values, was represented by 1 ms for convenience and later treated as 0 ms in the analysis.

VOT values were analyzed using JMP [17]. Overall 119 values were excluded because the speaker showed a hesitation while producing a stop or failed to produce the target sound. Subsequently, values were entered into a linear mixed model with VOT as the dependent factor, SPEAKER and ITEM as random factors, GENDER, MATERIAL (training/transfer targets), FOLLOWING SOUND, and a combination of GROUP (Control/Manipulation/Native Speaker), TEST (first/second recording), LANGUAGE (French vs. German), and PHONATION (voiced/voiceless) as independent factors.

The results of the statistical analysis indicated no main effect of GENDER ($F(1,26.13)=0.65$, $p=0.4271$), which confirms the overall expectation that the articulation of plosives is not gender specific. MATERIAL showed no effect ($F(1,334.8)=1.05$, $p=.3068$), which indicates that targets of the training (imitation) phase did not differ from the targets of the transfer phase. Therefore, if participants were able to imitate the manipulated or native stops, respectively, they were also able to transfer this production strategy to the experimental targets. FOLLOWING SOUND showed a significant influence on VOT ($F(13,94.99)=3.31$, $p<0.001$). The factor combination showed a main effect ($F(19,543.2)=49.55$, $p<0.0001$). Student's t-test were carried out to take a closer look at specific contrasts.

Firstly, no significant differences were found for the first recordings by the three experimental groups for either voiced or voiceless stops, which facilitates a comparison across subject groups. Comparing the

Figure 2: VOT values of voiced and voiceless stops for the first and second French recordings.

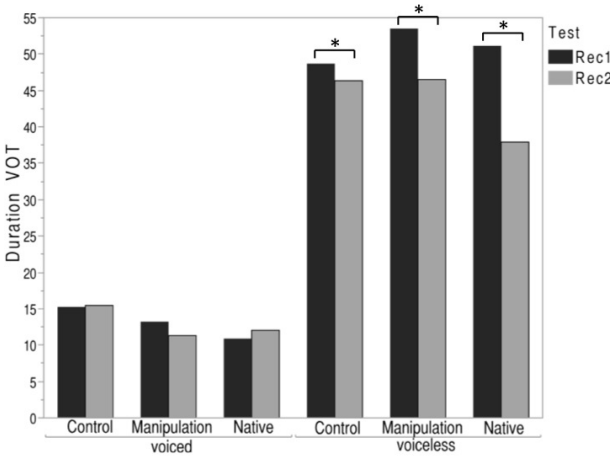
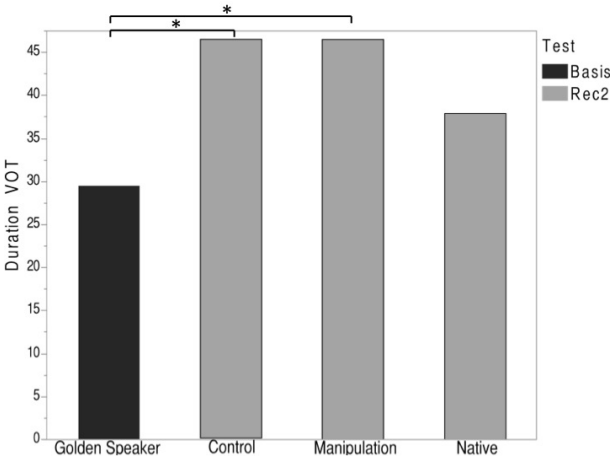


Figure 3: VOT values for voiceless stops of the second French recording in comparison to the golden speaker.



performance of the first and second recording, a significant reduction for VOT was found only for voiceless stops for all three groups (Figure 2). The mean VOT values in Table 2 indicate that the reduction for CG amounts to only 3 ms whereas the difference for MG and NG is 7 and 13 ms, respectively. It is doubtful that a reduction of 3 ms is sufficient for a noticeable perceptual effect.

A comparison of voiceless stops of the second recording showed that only speakers from the NG group managed to reduce their VOT to a level that is not significantly different from that of the golden speaker (Figure 3).

Due to space limitation a comparison of German and French productions is not included here.

4. DISCUSSION

This study examined the influence of two auditory feedback methods on improving the pronunciation of voiced and voiceless stops of German learners of French. It was shown that the manipulation of the speaker's voice, here the reduction of VOT, had a motivating effect on the production of these sounds (MG subjects). The improvement was only effective for voiceless but not for voiced stops. The same finding holds for the exposure to utterances of a native French speaker (NG subjects). Again, participants were able to reduce the duration of VOT and even more so than the Manipulation Group. A comparison of the Native Speaker Group with the golden speaker showed no significant difference. This demonstrates the benefit of being exposed to native productions, since learners were able to reduce VOT to the level of a native speaker while speakers of the Manipulation Group failed to do so.

It is striking that for both feedback methods, the training and transfer phase for each stop was quite short. Only four imitation and seven training targets were included. Even with such few training targets an improvement for voiceless stops was observed. It would be interesting to examine the strength of this effect while training for a longer time and with more training targets.

However, auditory feedback only improved the pronunciation of voiceless but not voiced stops. It might be argued that manipulation of VOT is not a sufficient method. But then, exposure to native speech also failed to show an effect for improving the pronunciation of voiced stops, although it comprises more pertinent information than only VOT duration. It is unclear whether learners are not able to perceive the differences or whether they are not able to apply the knowledge extracted from the native utterances. To examine this question, a high-variability study is currently under way in which learners train the perception of stops. It might also be advantageous for the learner to be exposed to more than one native speaker.

We also intend to perform a perception experiment to verify whether French native speakers actually perceive the reduction in VOT as a more French-like production.

5. ACKNOWLEDGEMENTS

This research was funded by ANR and DFG via the IFCASL project (www.ifcasl.org). We wish to thank our colleagues for discussions which helped improve this article.

6. REFERENCES

[1] Best, C. 1994. The emergence of native-language phonological influences in infants: a perceptual assimilation model. In: Goodman, J., Nusbaum, H., (eds), *The development of speech perception: The transition from speech to spoken words*. Cambridge, MA: MIT Press 167–224.

[2] Bissiri, M., Pfitzinger, H. 2009. Italian speakers learn lexcial stress of German morphologically complex words. *Speech Communication* 51, 933–947.

[3] Boersma, P., Weenink, D. 2013. Praat: doing phonetics by computer. [Version 5.3.51]. http://www.praat.org.

[4] Butler, D., Winne, P. 1995. Feedback and self-regulated learning: a theoretical synthesis. *American Educational Research Association* 64, 245–281.

[5] Colotte, V. 2013. Corpus Recorder. [Version 1.2].

[6] Demenko, G., Wagner, A., Cylwik, N., Jokisch, O. 2009. An audiovisual feedback system for acquiring L2 pronunciation and L2 prosody. *Proc. ISCA Workshop on Speech and Language Technology in Education* Warwickshire. 25–28.

[7] Engwall, O., Wik, P., Beskow, J., Granström, 2004. Design strategies for a virtual language tutor. *Proc. ICSLP* Jeju Island. 1693–1696.

[8] Flege, J. 1995. Second-language Speech Learning: Theory, Findings, and Problems. In: Strange, W., (ed), *Speech Perception and Linguistic Experience: Issues in Cross-language research*. Timonium, MD: York Press 229–273.

[9] Hincks, R., Edlund, J. 2009. Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology* 13, 32–50.

[10] Hirschfeld, U., Trouvain, J. 2007. Teaching prosody in German as a foreign language. In: Trouvain, J., Gut, U., (eds), *Non-Native Prosody. Phonetic Description and Teaching Practice*. Berlin, New York: Mouton de Gruyter 171–187.

[11] Jokisch, O., Koloska, U., Hirschfeld, D., Hofmann, R. 2005. Pronunciation Learning and Foreign Accent Reduction by an Audiovisual Feedback System. *Affective Computing and Intelligent Interaction: Lecture Notes in Computer Science* 3784, 419–429.

[12] Künzel, H. 1977. *Signalphonetische Untersuchung deutsch-französischer Interferenzen im Bereich der Okklusive. Forum Linguistikum, 10*. Bern, Frankfurt, Las Vegas: Peter Lang.

[13] Mennen, I., Schaeffler, F., Docherty, G. 2007. Pitching it differently: A comparison of the pitch ranges of German and English speakers. *Proc. 16th ICPhS* Saarbrücken. 1769–1972.

[14] Neri, A., Cucchiarini, C., Strik, H. 2008. The effectiveness of computer-based corrective feedback for improving segmental quality in L2-Dutch. *ReCall* 20, 225–243.

[15] Neri, A., Cucchiarini, C., Strik, H., Boves, L. 2005. The pedagogy-technology interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning* 15, 441–467.

[16] Precoda, K., Halverson, C., Franco, H. 2000. Effects of speech recognition-based pronunciation feedback on second-language pronunciation ability. *Proc. Integrating Speech Technology in (Language) Learning* Dundee. 102–105.

[17] SAS Institute Inc., 1989–2007. JMP [Version 11.0.0].

[18] Setter, J., Jenkins, J. 2005. State-of-the Art Review Article. *Language Teaching* 38, 1–17.