Published in: Möller, Sebastian/Ney, Hermann/Möbius, Bernd/Nöth, Elmar/Steidl, Stefan (eds.): Proceedings of Interspeech 2015. Speech beyond speech. Towards a better understanding of the most important biosignal. Dresden, Germany. 6-10 Sep, 2015. Volume 2. - Baixas: ISCA, 2015. Pp. 806-810.

The Effect of High-Variability Training on the Perception and Production of French Stops by German Native Speakers

Jeanin Jügler[1], *Frank Zimmerer*[1], *Bernd Möbius*[1], *Christoph Draxler*[2]

[1]Saarland University, Saarbrücken, Germany
[2]Ludwig Maximilian University, Munich, Germany

juegler|zimmerer|moebius@coli.uni-saarland.de, draxler@phonetik.uni-muenchen.de

Abstract

We investigated the effect of high-variability training (HVT) on the production and perception of French bilabial voiced and voiceless stops by German native speakers. Stop consonants in the two languages differ with respect to several articulatory and acoustic features. German learners of French (Experiment Group) trained the perception of word-initial bilabial stops spoken by six French native speakers using identification tests, whereas subjects of a Control Group did not perform a training. Additional perception and production tests of French words including bilabial, alveolar, and velar stops in all word positions were performed to capture the impact of HVT. Subjects were found to be quite good at distinguishing voiced and voiceless stops. However, voiceless stops received lower correctness scores than voiced ones and subjects of the Experiment group were able to further increase their scores after training. Results for production are mirror-inverted showing that subjects of the Experiment Group successfully produced longer negative VOT values but did not show an improvement for voiceless stops.

Index Terms: high-variability training, stops, French, German, second language learning

1. Introduction

When learning a foreign language (L2) most people usually retain a foreign accent which results from interferences of the native language (L1). Not only do learners of an L2 have problems to produce sounds as well as suprasegmental structures correctly. They also show difficulties perceiving the phonetic and phonological differences produced by a native speaker in comparison to their own non-native productions [1, 2, 3]. Learners have particular problems with L2 sounds that are phonetically similar to sounds of the learner's L1. For example, German and French both differentiate between phonologically voiced and voiceless stops /b d g/, /p t k/. French stops are distinguished in terms of fully voiced plosives vs. voiceless unaspirated ones with a rather short Voice Onset Time (VOT). In contrast, German stop sounds are distinguished in terms of voiceless unaspirated plosives with a short VOT vs. voiceless aspirated ones with a long VOT [4, 5]. French learners of German and German learners of French most likely transfer phonetic knowledge of their respective L1 to their L2 production which might result in difficulties in intelligibility.

Difficulties may also occur when the native language lacks a contrast appearing in the L2. For example, native Japanese speakers show difficulties in perceiving and producing English /ɹ/ and /l/ (e.g., [6, 7]). Intensive high-variability perceptual training, where high-variability refers to the use of multiple model speakers producing the stimuli in identification tasks, is known to contribute to a better performance in both perception and production of English minimal pairs by Japanese native speakers [8, 9, 10]. In these studies, participants completed an identification task training for minimal pairs contrasting between /ɹ/ and /l/ recorded by five speakers of American English. The training phase was carried out over a period of 3-4 weeks and feedback was provided immediately for correct and incorrect answers. The overall identification accuracy showed a significant improvement after training. Perceptual evaluation of the learners' productions by American English listeners showed that the production of /ɹ/ and /l/ words after training received higher rankings than before training. Furthermore, improvements could be maintained even three months after training.

The positive effect of high-variability training (HVT) was also shown in other studies such as [11] which investigated the learning of Mandarin tones by American English native speakers. Training resulted in more native-like productions and a perceptual improvement which was retained even after six months.

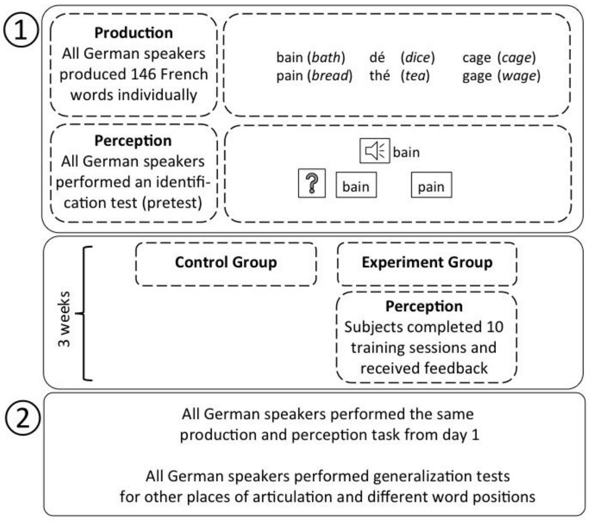
[12] compared low-variability and high-variability training for Dutch learners of Japanese geminate and singleton variants of /s/, showing that HVT leads to a better performance than low-variability training. In addition, a transfer of knowledge for identification of untrained stops and affricates was shown.

Although not HVT, [13] showed that Chinese learners of French were able to improve their perception and production by training on synthetic syllables of a /bu/-/pu/ continuum. The effect was transferred to labial stops followed by /a i/, dental and velar syllable-initial stops as well as voiceless natural stimuli. Intervocalic stops did not improve significantly.

Building new phonetic categories resulting from perceptual training with synthetic stimuli was also shown for English native speakers learning the three-way distinction between voiced, voiceless unaspirated, and voiceless aspirated stops differing in voice onset time [14]. Training on one place of articulation was shown to be transferable to other places [15].

Since HVT seems to be a good method to improve both the production and the perception of difficult L2 sounds, we investigated the effect on the production and perception of French fully voiced and voiceless unaspirated plosives with a short VOT by German native speakers. As mentioned before, German stops are distinguished differently which results in difficulties regarding the correct pronunciation and perception. Because both German and French have a binary distinction of voiced and voiceless sounds it would be interesting to see how well German learners of French are able to hear the differences between voiced and voiceless stops and whether and to what extent they are able to adopt a near-native French pronunciation.

Figure 1: Overview of the study's procedure



1
Production
All German speakers produced 146 French words individually
bain (bath) dé (dice) cage (cage)
pain (bread) thé (tea) gage (wage)
Perception
All German speakers performed an identification test (pretest)
bain
?
bain
pain
3 weeks
Control Group
Experiment Group
Perception
Subjects completed 10 training sessions and received feedback
2
All German speakers performed the same production and perception task from day 1
All German speakers performed generalization tests for other places of articulation and different word positions

2. Experiment

The effect of HVT was investigated regarding the perception and production of French stops by German native speakers. A *Control* and an *Experiment* Group were tested in the experiment. Subjects of the Experiment Group received feedback and completed a set of training sessions whereas subjects of the Control Group did not receive any feedback nor training.

Each group consisted of one male and five female speakers (19-32 years, M: 23.4 years, SD: 3.4 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 or employees at Saarland University.

2.1. Material

For each sound contrast (/b-p/, /d-t/, /g-k/), French minimal pairs differing in initial, medial, and final word position were used (see examples 1 for /b/-/p/ contrasts).

(1) **b**ain [bɛ̃] (*bath*) vs. **p**ain [pɛ̃] (*bread*)
dé**b**it [debi] (*debit*) vs. dé**p**it [depi] (*pique*)
trom**b**e [tʀɔ̃b] (cloudburst) vs. trom**p**e [tʀɔ̃p] (*trumpet*)

We decided to concentrate on word initial /b/-/p/ contrasts for the training sessions. Word medial and final bilabial minimal pairs as well as alveolar and velar pairs in all word positions were included in generalization tests. These tests where used to examine whether any improvements can be transferred to different word positions and different places of articulation (see [8, 9, 10, 13, 15]).

2.2. Procedure

The experiment comprised a production and a perception task. Participants tested on two different days in our lab whereas the training was performed online from home (Figure 1).

During the first appointment, subjects of both groups were asked to produce 146 French words, which included words of all places of articulation and word positions, to record a baseline for later comparisons. These words were part of the perception test of the pre- and post-test as well as in the generalization tests (Table 1). Recordings were made in quiet office rooms using a head mounted microphone (16 kHz, 16 bit) on an M-AUDIO Fast Track USB device. Recordings were saved on a Windows Laptop using a custom-made software developed at LORIA ("Corpus-Recorder", [16]). The words were presented to each speaker in a randomized order.

Table 1: *Number of French words used in the identification tests. Words differed between tests.*

	/bp/	**/dt/**	**/gk/**	∑
Pre-/ Post-Test	64 initial	8 initial	8 initial	80
Generalization	8 initial 6 medial 8 final	8 initial 6 medial 8 final	8 initial 6 medial 8 final	66
				146

Afterwards, the perception test was performed by all participants (pre-test). It was set up as an online experiment using the Percy framework [17, 18]. Participants of both groups were asked to listen to isolated French words spoken by a male (34, Bitche, Lorraine) and a female (28, Strasbourg, Alsace) French native speaker and were presented with two buttons displaying the voiced and voiceless orthographic variant of the auditory stimulus word. They had to decide which variant was presented. Participants did not receive any feedback in this part of the experiment and were allowed to listen to the presented word only once.

The post-test production and perception experiment was performed three weeks later and extended by three generalization tests, which included additional words differing in word position and place of articulation (Table 1). .

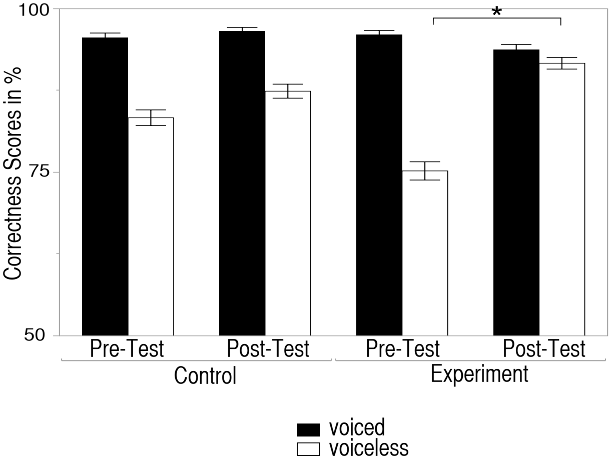
In the three weeks between the first and second appointment, subjects of the Experiment Group had to perform ten training sessions at home. They were instructed to distribute the sessions evenly and only perform one training per day. It was also suggested that they perform the training in a quiet environment and use headphones all the time. The training included 60 bilabial French words taken from the pre- and post-test. At this point, words produced by six different French native speakers (three male, three female) were presented to the participants, including the two French speakers from the pre- and post-test. Participants of the control group did not perform any training sessions.

In the training sessions, feedback was given by changing the color of the pressed button: green for correct response, i.e. the word matched the audio played, red for a mismatch between audio and response.

The following predictions were made:

1. Perception of French voiced and voiceless bilabial stops is affected by L1 interferences, resulting in a moderate error rate in the identification test.
2. HVT improves the perception and production of voiced and voiceless bilabial stops for subjects of the Experiment Group.
3. Improvements can be transferred to other places of articulation and word positions.

Figure 2: Correctness scores in % for participants of the Control and Experiment Groups for trained stop items.



3. Analysis and Results

3.1. Perception

When participants responded correctly to the auditory stimulus word, the response was labeled as '1', whereas incorrect responses were labeled as '0'. Due to differences in the implementation of voicing between the languages, we analyzed voiced and voiceless stops separately. We also distinguished between the identification of items that received trained and untrained items from the generalization test.

The values were entered into a linear mixed model in JMP [19] with Correctness as dependent factor and Speaker and Item as random factor. For trained items, Test (first vs. second identification test) and Group (Control vs. Experiment) were included as independent factors as well as their interaction. For untrained items we used Group, Word Position (initial vs. medial vs. final) and Place of Articulation (POA) as independent factors, as well as all possible interactions.

3.1.1. Trained Items

The results of the statistical analysis for voiceless trained stops shows a main effect for Test (F(1,3787)=93.22, p<0.0001) with higher correctness scores for the post-test (90%) than for the pre-test (79%). The interaction Test×Group also shows an effect (F(1,3787)=34.02, p<0.0001). Planned post-hoc tests indicate a significant difference between the pre- (75%) and post-test (92%) of the Experiment Group, all other comparisons did not reach significance.

As for voiced stops, only the interaction Test×Group reached significance (F(1,3787)=5.91, p<0.05). Post-hoc tests show a significant difference between pre-test (96%) and post-test (94%) of the Experiment Group showing a drop by two percent. Again, no other comparison showed a significance.

3.1.2. Untrained Items

For voiceless untrained items, the model suggests a main effect for word position (F(2,24)=21.53, p<0.0001). Post-hoc tests show that initial (91%) and medial (94%) stops are identified with significantly higher scores than final stops (68%). The interaction Group×Word Position shows an effect (F2,1532)=2.32, p<0.05) and the same picture emerges from post-hoc comparisons. Initial and medial voiceless stops of the Control and Experiment Group get significantly better scores than final stops, but differences are not found between groups.

The same holds true for voiced untrained stops. Word Position shows a main effect (F(2,24)=18.75, p<0.0001). Initial (98%) and medial (94%) voiced stops get significantly better scores than final voiced stops (81%). Post-hoc tests of the significant interaction Group×Word Position (F(2,1532)=4.83, p<0.01) shows the same picture, although final voiced stops of the Control Group (86%) received better scores than final voiced stops of the Experiment Group (75%).

3.2. Production

Duration of VOT was measured using Praat [20]. Again, we analyzed voiced and voiceless stops as well as trained and untrained stops separately. Values were entered into a linear mixed model with VOT as dependent factor and Speaker and Item as random factors. For trained items, Test (first vs. second identification test) and Group (Control vs. Experiment) were included as independent factors as well as their interaction. For untrained items we used Test, Group, Word Position (initial vs. medial vs. final) and Place of Articulation (POA) as independent factors, as well as all possible interactions

3.2.1. Trained Items

The results of the statistical model indicate a main effect for Test voiceless stops (F(1,697)=16.81, p<0.0001, illustrating shorter VOT values for the post-test (60 ms) than for the pre-test (65 ms). No other factors reached significance.

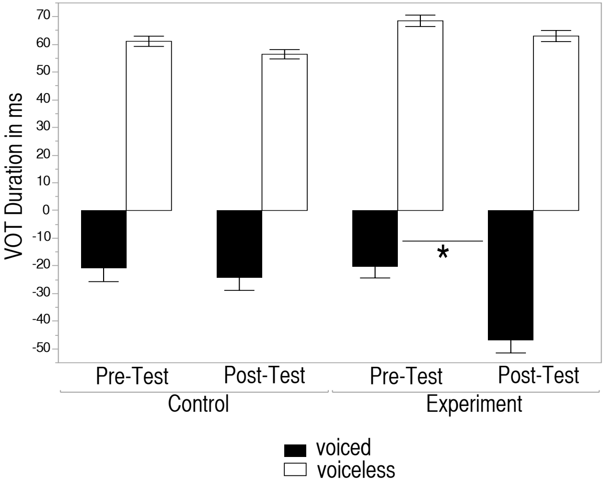
As for voiced stops, Test shows a significant effect (F(1,719.4)=17.98, p<0.0001), showing larger negative VOT values for the post-test (-36 ms) than for the pre-test (-21 ms). Additionally, the interaction Test×Group shows an effect (F(1,719.5)=10.69, p<0.01). A planned post-hoc test shows that VOT of the pre-test of the Experiment Group is produced with a significantly longer negative VOT (-46 ms) in comparison with the pre-test (-21 ms) of the Control Group. No other comparisons reached significance.

3.2.2. Untrained Items

The model indicates a main effect of POA (F(2,41759,15)=10.28, p<0.0001) and Word Position (F(2,80381.75)=19.78, p<0.0001). Post-hoc tests show that alveolar (79 ms) and velar (81 ms) voiceless stops are produced significantly longer than bilabial stops (67 ms). Because training was only carried out for bilabial stops, a transfer of knowledge might have been easier on unkown/untrained stops with the same place of articulation. Futhermore, post-hoc tests for word position show that initial (71 ms) and medial (72 ms) stops behave similarly but are produced differently from final stops (90), which might result from final lengthening processes.

The interaction Word Position×Test is significant (F(2,15940.42)=3.92, p<0.05). Since we recorded both trained and untrained stops twice, it is possible to compare untrained stops of the generalization before and after the training session on inititial bilabial stops. Similar to the factor Word Position, Post-hoc tests show that initial and medial stops from pre- and post-test are produced significantly shorter than final stops of both test conditions. Additionally, final stops of the post-test are produced longer than stops of the pre-test. Lastly, Group (F(1,10190.28)=5.02, p<0.05) indicates longer VOT values for the Experiment Group.

Figure 3: Mean VOT values for participants of the Control and Experiment Groups for trained stop items.



As for voiced stops, the model shows a main effect for POA (F(2,63427.71)=7.71, p<0.001) and Word Position (F(2 349193.61)=42.44, p<0.0001). Post-hoc tests indicate that bilabial voiced stops (-7 ms) behave differently than alveolar (7) and velar (10 ms) stops. Voiced stops in all three word positions are distinguished from each other, whereas medial stops are produced with a longer negative VOT (-18 ms) than bilabial stops (-1). Final stops are produced voiceless (34 ms).

The interaction Word Position×Group indicates a main effect (F(2,27460.76)=3.34, p<0.05). Post-hoc tests show that initial stops of the Control Group as well as medial stops of both groups are produced as fully voiced stops. Initial stops of the Experiment groups are articulated with a short positive VOT (8 ms) which is significantly different from the Control Group and medial stops. Again, final stops were produced with a positive VOT by the Experiment Group, whereas final stops of the Control Group (20 ms) were produced shorter than stops of the Experiment Group (50 ms). Furthermore, an effect was found for Group indicating that voiced stops of the Control Group (-3 m) were produced with shorter VOT values on average than the Experiment Group (12 ms).

4. Discussion

The analysis of the production and perception of voiced and voiceless stops shows two different aspects of behavior of German learners of French. Subjects of the Control and Experiment Group performed relatively well in the identification of stops with a general score of above 85% already before training (Table 3). Since both French and German have a two-way distinction of stops, developing a strategy to distinguish between voiced and voiceless stops, although phonetically marked differently, seems to be rather straightforward (see also [21]).

Figure 2 shows that voiced bilabial stops tend to be identified better than voiceless stops. Since they already achieved a correctness score above 90%, it may be difficult to improve significantly further. Training seems to help participants of the Experiment Group, who were able to improve their score by 17% for voiceless stops. It can also be said that participants seem to do equally well for all places of articulation (above 80%).

Regarding the production, it was found that training helps to produce voiced stops with a longer negative VOT (Figure 3).

Table 2: *Absolute numbers (and percentage) of initial fully voiced stops and voiced stops with voicing that does not last through the entire closure.*

	Control		Experiment	
	Pre-Test	Post-Test	Pre-Test	Post-Test
CD_b	10 (2.5)	6 (1.5)	21 (5.2)	21 (5.2)
CD_d	5 (6.9)	5 (6.9)	18 (25.0)	22 (30.6)
CD_g	3 (4.2)	3 (4.2)	9 (12.5)	12 (16.7)
-VOT_b	59 (14.5)	64 (15.7)	69 (16.9)	103 (25.3)
-VOT_d	8 (11.1)	6 (8.3)	3 (4.2)	13 (18.1)
-VOT_g	7 (9.7)	11 (15.3)	4 (5.6)	8 (11.1)

Table 3: *Mean correctness scores (%) for the three test conditions by participants of the Control and Experiment Group.*

		Mean Correctness %
Pre-Test	Control	0.89
	Experiment	0.86
Post-Test	Control	0.92
	Experiment	0.93
Generalization	Control	0.89
	Experiment	0.87

Table 2 shows that subjects of the Experiment Group produced an increased number of initial bilabial fully voiced stops in the post-test, from 17% to 25%. This may be an indicator for developing an awareness of the correct production of voiced stops.

It is interesting to see that for untrained stops, bilabial voiced and voiceless stops behave differently than alveolar and velar plosives. They tend to have a shorter positive VOT for voiceless stops and a longer negative VOT for voiced ones. This might be an indicator that training on bilabial stops does have a larger impact on plosives with the same place of articulation.

Another interesting fact is that voiced final stops behave differently than initial and medial ones. They are produced with longer and positive VOT values. Firstly, this can be a result of a final lengthening processes. But additionally, it might also be associated with the phonological process of final devoicing in German and thus explaining the second behavior. Since training did not include final stops, German native speakers would not have necessarily learned to produce fully voiced stops at the end of a word (or syllable). Differences in production and perception are therefore not only a product of phonetic differences but are also influenced by phonological rules. It would be interesting to see whether a training on final stops would result in a more French-like production. Without this training, the necessary information is inaccessible to the participants.

We conclude, that high variability training seems to have a beneficial effect on the production and perception of French voiced and voiceless stops by German native speakers. The results reported here may be seen as a challenge for accounts claiming a close link between production and perception, because the behavioral patterns of the participants were mirror-inverted - rather than parallel, as one might expect - for the production and perception tasks.

5. Acknowledgements

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

6. References

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

[2] J. Flege, “Second-language Speech Learning: Theory, Findings, and Problems,” in *Speech Perception and Linguistic Experience: Issues in Cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 229–273.

[3] J. Kingston, “Learning Foreign Vowels,” *Language and Speech*, vol. 46, no. 2-3, pp. 295–349, 2003.

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

[5] E. Pustka, *Einführung in die Phonetik und Phonologie des Französischen.* Berlin: Erich Schmidt Verlag, 2011.

[6] K. S. MacKain, C. T. Best, and W. Strange, “Categorical perception of English /r/ and /l/ by Japanese bilinguals,” *Applied Psycholinguistics*, vol. 2, pp. 269–390, 1981.

[7] R. A. Yamada and Y. Tohkura, “The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners,” *Perception & Psychophysics*, vol. 52, pp. 376–392, 1992.

[8] J. S. Logan, S. E. Lively, and D. B. Pisoni, “Training Japanese listeners to identify English /r/ and /l/: A first report,” *Journal of the Acoustic Society of America*, vol. 89, no. 2, pp. 874–886, 1991.

[9] A. R. Bradlow, R. A. Akahane-Yamada, D. B. Pisoni, and Y. Tohkura, “Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production,” *Journal of the Acoustic Society of America*, vol. 101, no. 4, pp. 2299–2310, 1997.

[10] ——, “Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production,” *Percept Psychophys*, vol. 61, no. 5, pp. 977–985, 1999.

[11] Y. Wang, M. M. Spence, A. Jongman, and J. A. Sereno, “Training American listeners to perceive Mandarin tones,” *Journal of the Acoustic Society of America*, vol. 106, no. 6, pp. 3649–3658, 1999.

[12] M. Sadakataa and J. M. McQueen, “High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates,” *Journal of the Acoustic Society of America*, vol. 134, no. 2, pp. 1324–1335, 2013.

[13] B. L. Rochet, “Perception and Production of Second-Language Speech Sounds by Adults,” in *Speech Perception and Linguistic Experience: Issues in Cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 379–410.

[14] C. L. McClaskey, D. B. Pisoni, and T. D. Carrell, “Transfer of training of a new linguistic contrast in voicing,” *Percept Psychophys*, vol. 34, no. 4, pp. 323–330, 1983.

[15] D. B. Pisoni, R. N. Aslin, A. J. Perey, and B. L. Hennessy, “Some Effects of Laboratory Training on Identification and Discrimination of Voicing Contrasts in Stop Consonants,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 8, no. 2, pp. 297–314, 1982.

[16] V. Colotte, “Corpus Recorder. [Version 1.2],” 2013.

[17] C. Draxler, “Percy – an HTML5 framework for media rich web experiments on mobile devices,” in *Proc. Interspeech*, Florence, 2011, pp. 3339–3340.

[18] ——, “Online experiments with the Percy software framework – experiences and some early results,” in *Proc. LREC*, Reykjavik, 2014, pp. 235–240.

[19] SAS Institute Inc., “JMP [Version 11.0.0],” 1989–2007.

[20] P. Boersma and D. Weenink, “Praat: doing phonetics by computer. [Version 5.4.06],” http://www.praat.org, 2013.

[21] C. T. Best and P. A. Hallé, “Perception of initial obstruent voicing is influenced by gestural organization,” *Journal of Phonetics*, vol. 38, pp. 109–126, 2010.