INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Vietnamese learners tackling the German /ʃt/ in perception

Anke Sennema[1], *Silke Hamann*[2]

[1]University of Vienna, Austria
[2]University of Amsterdam, The Netherlands
anke.sennema@univie.ac.at, silke.hamann@uva.nl

Abstract

Previous observations from didactic studies have indicated that Vietnamese learners of German as a foreign language often fail to realize consonantal clusters in German [1, 2, 3]. The present study investigated whether this problem occurs already at the level of perception, i.e., whether Vietnamese learners find it difficult to perceive the difference between a cluster and a single consonant. We focused on the discrimination between the German cluster /ʃt/ and the single consonants /t/ and /ʃ/, both in onset and coda position. Due to different phonotactic restrictions on coda consonants in Vietnamese, we expected the coda position to pose a bigger challenge for correct discrimination than the onset position. With an AX discrimination task, we tested how 83 university students from Hanoi perceived these contrasts. Our findings show that only the distinction between /ʃt/-/ʃ/ in coda position posed a real challenge to our listeners. We attribute this difficulty to the weak and non-native auditory cues for the plosive in this position. For all other contrasts our participants performed surprisingly well. We propose that this is due to the influence of English as first L2 that facilitates the acquisition of phonological contrasts in German as an L3.

Index Terms: L2 sound perception, consonants, phonemic category, position, Vietnamese

1. Introduction

The native language (L1) has a strong influence on how we perceive and produce sounds in a second language (L2). For Vietnamese learners of German it was shown in previous studies that they have problems producing the German contrast in vowel length [1]. Another noticeable phenomenon concerns the articulatory distinction between the consonants [ç], [ʃ] und [s], and the substitution of consonants in syllable final position, such as [ç] and [x] by [k], or [l] by [n] [3, p. 49]. A third point mentioned frequently for Vietnamese learners is their problem with consonantal clusters, both in German [1-3] and English [4, 5].

While previous studies have focused on the production side [1-5], the present study investigates whether the perception is equally affected. In our study, we concentrate on the initial stages at which perceptual sensitivity to German phonotactics sets in, by examining adult German as foreign language (Gfl) learners with up to 18 months of L2 exposure.

The topic of our study is the German consonant cluster /ʃt/, which can occur both in onset and coda position. Right from the beginning stages of German language instruction on level A1 (according to the *Common European Framework of reference for languages*, CEFR), the cluster is part of the basic vocabulary.

2. Vietnamese and German phonology

2.1. Consonants in Vietnamese

Vietnamese is a language with obligatory consonant in the onset [6]. The following description is based on the Hanoi variety of Vietnamese, where any of the consonants given in Table 1 (based on [7, p. 382]) can occur in the onset. The segments given in brackets only occur in a small number of loanwords. In words that orthographically start with a vowel grapheme, the corresponding vowel is preceded by a glottal stop [11, p. 25].

Table 1: *Vietnamese onset consonant inventory.*

	Labial	**Dental/ Alveolar**	**Palatal**	**Velar**	**Glottal**
Plosive	(p) ɓ	tʰ t ɗ	tɕ	k	
Fricative	f v	s z		x ɣ	h
Nasal	m	n	ɲ	ŋ	
Approximant	w	l (r)	(j)		

In coda position, Vietnamese only allows the voiceless stops [p, t, k] and the nasals [m, n, ɲ, ŋ] [2].

Neither in onset nor coda position does Vietnamese allow consonant clusters [7, 8]. Consonant sequences can occur across syllable boundaries, though Vietnamese is described as a predominantly monosyllabic language.

2.2. Consonants in German

The consonantal phonemes of German that can occur in onset position are given in Table 2 (from [9]). Segments in brackets are again restricted to loanwords. Like in Vietnamese, syllables that orthographically start with a vowel grapheme have a corresponding vowel that is preceded by a glottal stop.

Table 2: G*erman onset consonant inventory.*

	Labial	**Dental**	**Post-alveolar**	**Palatal**	**Velar**	**Glottal**
Plosive	p b	t d			k g	
Fricative	f v	z	ʃ (ʒ)	(ç)		h
Affricates	pf	ts	tʃ (dʒ)			
Nasal	m	n				
Liquid		l		j		
Rhotic					ʀ	

All of the consonants given in Table 2 can occur in coda position in German, apart from the voiced obstruents, which are realized voiceless due to final devoicing [9]. Additionally, the velar nasal [ŋ] and fricative [x] are allowed in coda position [9].

http://dx.doi.org/10.21437/Interspeech.2019-2832

With respect to consonant clusters, German allows clusters of two or three consonants word initially, however clusters with three consonants are restricted to some combinations such as [pfl] [ʃtR]. In coda position up to five subsequent consonants are possible (though this depends on phonological interpretation and assumptions such as extrametricality, see e.g. [10]).

3. Perception study

Our investigation consisted of two parts. First, we tested whether early Vietnamese-speaking Gfl learners have difficulties in distinguishing the consonant cluster /ʃt/ from the single phonemes /ʃ/ and /t/.

We also were interested in whether Vietnamese listeners differed in the perception of /ʃt/ depending on the position of this cluster (in onset or coda position). Our hypothesis was that Vietnamese learners have difficulties in perceiving the difference between the German cluster /ʃt/ and single consonants /t/ or /ʃ/, especially in coda position, thus confirming findings reported for production by [2] and [3].

To test whether phonotactic constraints of Vietnamese had an effect on how clusters are perceived in contrast to single consonants, we compared the rate of correct discrimination for all items that consisted of non-identical word pairs. Given that Vietnamese does not allow clusters [7, 8], we expected participants to discriminate single consonant contrasts better than clusters.

To test both hypotheses, we designed a perception experiment with an AX discrimination task using Praat [11].

3.1. Materials

Cluster items were 36 minimal pairs of real words (nouns, verbs, prepositions) contrasting /ʃt/ with /ʃ/ or /t/ either in initial or final position, see the examples in (1). The words did not contain any further clusters.

(1)	a) initial:	[ʃta̯ɪn] - [ʃa̯ɪn]	“stone - shine”
		[ʃta̯ʊp] - [ta̯ʊp]	“dust - deaf”
	b) final:	[mɪʃt] - [mɪt]	“mix-3SG - with”
		[ta̯ʊʃt] - [ta̯ʊʃ]	“swap-3SG - swap-IMP”

Each cluster item pair was presented in random order in four combinations, namely as A-B, B-A, A-A and B-B (with different tokens in all combinations). As some of the cluster item pairs shared words (e.g. *Stau* - *Schau* and *Stau* - *Tau*), we ended up with a total of 110 cluster items. In addition, 38 minimal pairs containing monosyllabic words without any clusters but contrasting either in initial or final consonant were included, see the examples in (2).

(2)	a) initial:	[ba̯ɪn] - [va̯ɪn]	“leg - wine”
		[ho:n] - [lo:n]	“mockery - wages”
	b) final:	[za:l] - [za:t]	“hall - seed”
		[ha̯ʊt] - [ha̯ʊs]	“skin - house”

The pairs contrasting in single consonants were presented in the orders A-B, A-A and B-B. Some single consonant pairs shared again words, resulting in 115 single consonant items. In total, we had 215 stimuli pairs.

We used only monosyllabic words to reduce the memory load and to ensure that we tested the learners’ perceptual abilities rather than how well they memorized words.

All items were produced by a phonetically trained female speaker of Standard High German, recorded with an Olympus VN-7800PC digital voice recorder.

3.2. Participants

24 Vietnamese Gfl students of the Hanoi University (HANU) and 59 Vietnamese Gfl students of the Vietnam National University (VNU), all aged between 18 and 22, took part in the experiment. They were tested in Hanoi on a voluntary basis. At the time of testing, none of them reported any hearing impairment. The participants were enrolled in German language classes at a level of A1 or A2 (according to the CEFR). All spoke the Hanoi variety of Vietnamese.

3.3. Procedure

The listeners’ perception accuracy was tested by an AX discrimination task. Participants were instructed in German that they had to listen to two words at a time and then had to indicate whether the two words were the same or not by clicking the respective button (same or different) on a computer screen. During the experiment, both the instructions and the button labels were presented on the computer screen in German and in Vietnamese. Participants listened to the word pairs via headphones in a fairly quiet classroom surrounding at their universities. Before the experiment started, they were presented with two filler pairs (one same, one different) to familiarize them with the procedure. It took participants about 15 minutes to complete the task. In addition to the perception test, they filled in a consent form and a questionnaire on their linguistic background.

3.4. Analyses and results

We analyzed the accuracy of the responses to all stimuli pairs (83*215=17,845). Participants’ mean performance correct for these pairs was 87.34% (s.d. 10%), ranging from 48.8% to 98.6% correct. For further analyses, we looked separately at identical cluster pairs, were we expected the answer “same”, and non-identical cluster pairs, were we expected the answer “different”. For the identical cluster pairs, we were interested whether participants performed equally well for the pairs with cluster compared to the pairs with fricative only or plosive only, and whether position (onset or coda) influenced this performance.

For the identical cluster pairs, we performed a generalized linear mixed effects model in R (glmer from the package lme4; [12]) with response (‘same’ or ‘different’) as dependent variable, pair (cluster, fricative or plosive) * position (onset or coda) * log-performed months of tuition as within-subjects factors, word pair as random intercept and participant as random intercept and slope.

There was no significant difference between responses to stimuli pairs with clusters and responses to stimuli pairs with fricative /ʃ/ or plosive /t/ (88.4% correct), and none between pairs with fricative /ʃ/ and plosive /t/ (86.7% correct). However, position of contrast had a main effect: if the contrast was in onset position, participants responded with ‘same’ (correct) significantly more often ($p = 0.0201$). We conclude from these results that participants find word pairs with the cluster /ʃt/ not inherently different from word pairs with single obstruents.

We performed another generalized linear mixed effects model for the non-identical cluster pairs, again with response as dependent variable, and pair (cluster compared to fricative vs. cluster compared to plosive) * position (onset or coda) * log-performed months of tuition as within-subjects factors, word pair as random intercept and participant as random intercept and slope. Both pair and position showed main effects (more correct responses for the pairs cluster-plosive and for onset position) and a significant interaction ($p < 0.0001$), as shown in Figure 1.

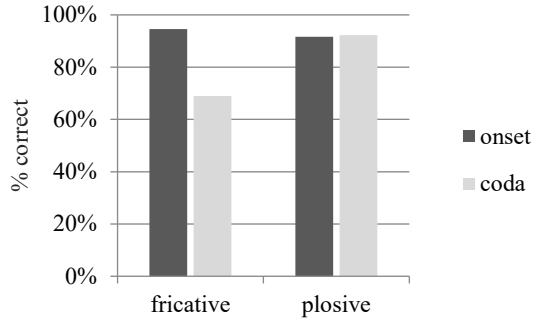

Figure 1: *Discrimination (% correct) for non-identical pairs (where /ʃt/ is compared either to fricative /ʃ/ or to plosive /t/), in onset or coda position.*

As can be seen from Figure 1, participants performed worse for the comparison cluster-fricative in coda position than for any of the three other comparisons. We conclude from this that in coda position, Vietnamese listeners have problems distinguishing /ʃt/ from /ʃ/.

In an additional explorative analysis, we checked whether order of the presentation of the words within the word pairs (AB or BA) played a role by including this as factor in both models, but it showed no effect.

In two last generalized linear mixed effects models, we compared the performance for all single consonants pairs with the pair /ʃt/-/t/ and /ʃt/-/ʃ/. We used again response as dependent variable, pair (/ʃt/-/t/ or /ʃt/-/ʃ/. vs. single consonant pairs) * position (onset or coda) * log-performed months of tuition as within-subjects factors, word pair as random intercept and participant as random intercept and slope. For the comparison with /ʃt/-/t/, only a main effect of position could be found, while for the comparison with /ʃt/-/ʃ/, we also found an interaction with pair ($p = 0.00521$), see Figure 2.

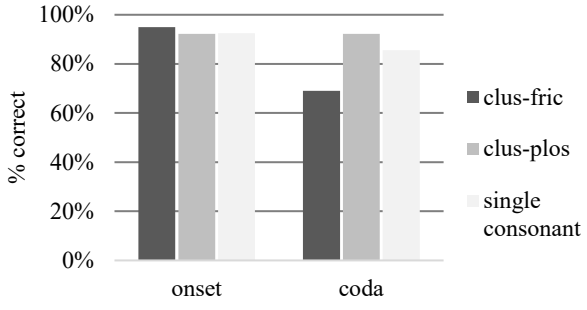

Figure 2: *Discrimination (% correct) for* /ʃt/-/ʃ/ *and* /ʃt/-/t/ vs. *single consonants, in onset or coda position.*

We infer from these findings that any of the tested contrasts are easily distinguished in onset position. In coda position, the contrast /ʃt/-/t/ does not cause significantly different responses than other pairs containing only single consonants, but /ʃt/-/ʃ/ is difficult to distinguish for Vietnamese learners.

In none of the models, months of German tuition had an effect: Their experience with German had no influence on the performance of our participants.

4. Discussion

We found that Vietnamese learners of German at beginners' level can distinguish clusters from single consonants fairly well, and that their performance did not improve with increased input of German. A possible explanation for this unexpected finding is provided in section 4.1.

The Vietnamese Gfl learners performed worse at distinguishing pairs in coda position than in onset position. This is in accordance with our expectations and is due to the much smaller coda inventory and less perceptual cues in coda than in onset position in Vietnamese. Especially the comparison of /ʃt/ with /ʃ/ in coda position proved difficult for Vietnamese Gfl learners. We propose two explanations for this finding in section 4.2.

4.1. Cross-linguistic influence in language acquisition

Next to the widely attested transfer from the L1 of the multilingual learners, transfer from previously acquired non-native languages is regarded as a significant source of cross-linguistic influence. Therefore, foreign language acquisition has started to differentiate between the acquisition of an L2 and other subsequent languages ([13] [14] [15]). For the acquisition of a third language (L3) phonology, Hammarberg & Hammarberg [16] reported in a longitudinal case study the tendency to resort to L2 articulatory settings that override L1 transfer at the initial stages of L3 acquisition. Wrembel [17] corroborates these findings for the language pair German L2 - English L3. Suggestions on pedagogical implications for the transfer of complex consonant clusters for learners with L2 English and L3 German in a learning and teaching context are made by Marx & Mehlhorn [18].

In Vietnam, English language tuition is compulsory at school level. In terms of phonotactics, English allows consonant clusters both in onset, as in *string* [stɹɪŋ], and coda position, e.g., *fast* [fɑːst]. With regard to the high scores of correct discrimination of non-native consonant distinctions in our experiment, the influence of English as first L2 on the acquisition of phonological contrasts in German as L3 seems probable. Thus, our learners were presumably drawing on their linguistic knowledge and cognitive experience of English as L2.

4.2. Auditory and lexical non-saliency of coda /ʃt/-/ʃ/

Two reasons for the difficulty of /ʃt/-/ʃ/ compared to /ʃt/-/t/ in coda position can be given. The first explanation concerns the salience of auditory cues in coda position. In the /ʃt/-/t/ pairs there is a salient auditory cue in the form of the friction noise indicating that the two are different. In addition, this cue is available right from the beginning of the coda contrast. In the case of /ʃt/-/ʃ/, on the other hand, the contrast is only perceivable when the auditory cues of stop closure and burst occur, i.e., at the very end of the word. The closure cue in coda position is rather weak, and on its own can be interpreted as signaling the word end. Furthermore, the final burst cue is something that

does not natively occur in Vietnamese, therefore Vietnamese listeners are probably not sensitive to it.

A second reason for the difficulty of the coda /ʃt/-/ʃ/ contrast might be the fact that all word pairs containing this contrast were pairs of identical verbs that contrasted only in their inflection (indicative vs. imperative). Though this grammatical contrast is highly relevant for native German listeners, beginning Vietnamese learners of German might not have had enough experience with the imperative yet and perceived the two forms of each pair as being the same lexical item.

5. Conclusions

Our study showed that for Vietnamese language learners of German, only the contrast /ʃt/-/ʃ/ in coda position posed a real challenge. We attribute this problem to the fact that the contrast in this pair only hinges on the auditory cues of silent closure phase and stop burst. Especially the latter is problematic for Vietnamese listeners, who in addition to not having any clusters in their native language, do not phonetically release bursts in coda position.

For all other contrasts tested in this study, the Vietnamese listeners performed surprisingly well. We attribute this finding to the influence of English as first L2 that facilitates the acquisition of phonological contrasts in German as an L3. Results thus support the notion of transfer of linguistic knowledge and cognitive experience from L2 to L3.

In future work, we will analyze production tokens of the stimuli that were recorded from the participants after the perception experiment. Phonetic analyses of the realization of the cluster and its phonemes can be complemented by listener ratings, and by way of triangulation this will supply insight into what makes particularly the cluster /ʃt/-/ʃ/ in coda position prone to errors.

6. Acknowledgements

This research was funded by a grant of ASEA-Uninet. We thank Professor Dang Thi Thu Hien at Hanoi National University (HANU) and Prof. Ta Thi Hong Hanh at Vietnam National University Hanoi (VNU) for their substantial help in organizing the testing of the Vietnamese students.

Portions of this work were presented at “Phonetik und Phonologie im deutschsprachigen Raum (P&P) 2018”. Vienna, September 2018.

7. References

[1] H. Dieling, *Phonetik im Fremdsprachenunterricht Deutsch.* Berlin, München, Leipzig: Langenscheidt, 1992.

[2] H. P. Kelz, “Vietnamesisch,” in U. Hirschfeld, H. P. Kelz, U. Müller (ed.) *Phonetik international. Grundwissen von Afrikaans bis Zulu. Kontrastive Studien für Deutsch als Fremdsprache*. Waldsteinberg: Popp, 2013.

[3] D. Diehl, “Integration von Phonetik in den DaF-Unterricht am Goethe-Institut Hanoi,” in U. Hirschfeld, F. Lange, E. Stock (ed.) *Phonetische und rhetorische Aspekte der interkulturellen Kommunikation*. Frank & Timme: Berlin, 2016.

[4] N. Nguyen, *Interlanguage Phonology and the Pronunciation of English Final Consonant Clusters by Native Speakers of Vietnamese*. Ohio, Ohio University, 2008.

[5] A. Osburne, “Final cluster reduction in English L2 speech: A case study of a Vietnamese speaker.” *Applied Linguistics* vol. 17, no. 2, pp. 164-181, 1996.

[6] X. G. Le, *Phonetische Interferenzen der vietnamesischen Deutschlernenden*. Jena: Thüringer Universitäts- und Landesbibliothek, 2014.

[7] J. P. Kirby, “Vietnamese (Hanoi Vietnamese),” *Journal of the International Phonetic Association*, vol. 41, no. 3, pp. 381-392, 2011.

[8] L. C. Thompson, *A Vietnamese grammar*. Seattle: University of Washington, 1965.

[9] T. A. Hall, *Phonologie*. Berlin: de Gruyter, 2011.

[10] R. Wiese, *The Phonology of German*. Oxford: Oxford University Press, 2000.

[11] P. Boersma, D. Weenink, *Praat. Computer program*, version 6.0.22. Retrieved from www.praat.org in November 2016

[12] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effectsmodelsusinglme4. *Journal of Statistical Software,* 2015.

[13] G. De Angelis, *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters, 2007.

[14] J. Cenoz, B. Hufeisen, U. Jessner (ed.) *Crosslinguistic Influence in Third Language Acquisition. Psycholinguistic Perspectives*. Clevedon: Multilingual Matters, 2001.

[15] R. Llama, W. Cardoso, L. Collins, “The influence of language distance and language status on the acquisition of L3 phonology” *International Journal of Multilingualism* vol. 7, no. 1, pp. 39-57, 2009.

[16] B. Hammarberg, B. Hammarberg, “Re-setting the basis of articulation in the acquisition of new languages: A third-language case study,” in B. Hufeisen, R. Fouser (ed.) *Introductory Readings in L3*. Tübingen: Stauffenburg Verlag, 2005.

[17] M. Wrembel, “L2-accented speech in L3 production,” *International Journal of Multilingualism*, vol. 7, no. 1, pp. 75-90, 2010.

[18] N. Marx, G. Mehlhorn, “Pushing the positive: encouraging phonological transfer from L2 to L3.” I*nternational Journal of Multilingualism*, vol. 7. no. 1, pp. 4-18, 2010.