



Impact of fluency and segmental categorization in L2: the case of French final fricatives uttered by German speakers.

Anne Bonneau¹

¹Multispeech Group, Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

anne.bonneau@loria.fr

Abstract

This study examines the effects of L1/L2 interferences at the segmental level, and of lack of fluency at the sentence level, on the realizations of French final fricatives by German learners. Due to L1/L2 interference, German speakers tend to devoice French final fricatives. A well-known effect of the lack of L2 mastering is the decrease of the speech articulation rate, which lengthens segment average duration. In order to better apprehend the impact of categorization and fluency, we selected four series of consonants from the IFCASL corpus, i.e. voiced and unvoiced fricatives uttered by French native (FF) and German non-native (GF) speakers. The realizations of unvoiced consonants uttered by GF speakers are essentially dependent upon fluency, whereas the realizations of voiced consonants by the same speakers are dependent upon both fluency and categorization. We evaluated a set of acoustic cues related to the voicing distinction –including consonant duration and periodicity–, and submitted the data to a hierarchical clustering analysis (HCPC, from R, FactoMineR package). Results, discussed as a function of speaker's level and prosodic boundaries, confirmed the mutual importance of fluency and segmental categorization on non-native realizations.

Index Terms: L1/L2 interference, fluency, voicing neutralization, categorization, fricatives, German/French.

1. Introduction

The pronunciation of a non-native segmental category is influenced by the native phonological system of learners [1]. L1/L2 interferences at the segmental level have been the object of a number of models, including Best's PAM [2] and Flege's SLM [3]. The latter model predicts that a new L2 sound is easier to acquire by learners when it is dissimilar to L1 sounds than when it is similar to (at least) one of them. Difficulties for learners come also from phonological phenomena, such as neutralizations. In such cases, they might erroneously apply in L2 a rule present in their native language, e.g. German learners might devoice French voiced final obstruents [4].

Inside sentences or texts, the realizations of non-native segmental categories also depend upon L2 fluency. A well-known consequence of a lack of fluency is the decrease in articulation rate [7] [8]. This decrease in turn influences the duration of speech segments, which tends to be longer in L2 than in L1 speech. L2 speech is also characterized by the presence of more disfluencies, such as silent and filled pauses, or repairs, than L1 [5], probably due to a lack of mastering at various L2 levels (lexicon, syntax, intonation ...) [6]. There is a growing number of works on the influence of prosody on segments [9] [10] [11], and this influence is particularly strong

in L2 due to the lack of fluency. The lengthening of speech segments at phrase boundaries, often called "final lengthening", is an example of this influence that is of particular interest in this study. It has been observed in a number of works for various languages and increases with the importance of the prosodic domain [12] [13]. Since the degree of "lengthening" should be relatively important before pauses, the higher number of pauses in L2 speech, mentioned above, should lead to the presence of a higher number of long final segments in L2 than in L1. Hence, L2 segments are affected both by L1/L2 interferences at the segmental level and fluency at a more global level.

In order to investigate the impact of fluency and categorization on L2 segments in sentences, we analyzed the pronunciation of French voiced and unvoiced fricatives in final position of an accental group, but not in sentence final position, by French and German speakers. In German, the phonological opposition between the voiced and the unvoiced fricative categories is neutralized in final position in favor of the unvoiced series [14]. Therefore, German speakers tend to devoice French final voiced fricatives [4]. Voicing, on an articulatory point of view, corresponds to vocal fold vibration and generates periodicity. However, the distinction between voiced and unvoiced categories does not rely only upon the presence or absence of periodicity [15]. For fricatives, the duration of the consonant, the relative duration of the preceding vowel, as well as the intensity of the noise play also an important role [16]. An error in categorization, due to L1/L2 interference at the segmental level, would have the following acoustic consequences: the realizations of the final French voiced consonants by German non native speakers should be longer, more intense, less periodic, and the vowel preceding the consonant should be shorter than what is observed for native realizations.

Inside sentences, the most obvious impact of (a lack of) fluency on speech segments is the lengthening of the average segment durations, due to a decrease in articulation rate. With respect to our study, this should increase fricative duration, an effect that goes in the same direction than that of an error in categorization. In turn, the increase in fricative duration should lead to a decrease in the amount of periodicity during the noise since it is very hard to maintain periodicity in long fricative noises [17], an effect that also goes in the same direction than an error in categorization. In addition, we should observe more fricatives followed by a pause for non-native than for native speakers, which could again contribute to lengthen average fricative durations. To consider this factor, we have distinguished a number of prosodic contexts in the analysis, including the presence or absence of a pause after the fricative under analysis (section 3).

We selected four series of consonants from the IFCASL corpus [18], i.e. voiced and unvoiced fricatives uttered by French native (FF) and German non-native (GF) speakers. The realizations of unvoiced consonants uttered by GF speakers are dependent upon fluency, whereas the realizations of voiced consonants by the same speakers are dependent upon fluency and categorization. We evaluated a set of acoustic cues related to the voicing distinction, and submitted the data to a hierarchical clustering analysis.

2. Corpus and Protocol

Our corpus material is a subset of the bilingual IFCASL corpus [18]. We extracted three French sentences (sr01, sr03 and sr07) containing respectively the fricatives /z ʒ ʃ/ in final position of an accentual group but not in sentence final position, and produced by French and German speakers (hereafter the native and non-native speakers, respectively).

The subjects were seated in a quiet room and read the sentences from the screen of a Windows laptop, with a headset microphone (AKG C520) connected to an Audiobox (M Audio Fast track). Forty German learners of French (23 beginners - A1, A2 and B1 levels-, and 17 advanced -B2 and C1 levels) and fifty-four French native speakers recorded the sentences. Each sentence was produced once by each speaker, which makes a total amount of 282 sentences. Some data have been removed due to spelling mistakes and the acoustical analysis was based upon 270 sentences.

The sentences were made up of three noun phrases and a verb phrase in that order: NP1 VP NP2 NP3, where NP1 was a subject, NP2 a direct object and NP3 an adverbial phrase beginning with a preposition. The fricative was at the end of the last word of NP2, was preceded by the vowel /a/ and the preposition (i.e. the word following the fricative) began with the vowels /a/ or /ã/. Note that we do not find for the consonant /s/ a sentence with a structure identical to the one described above, which explains the absence of this consonant in our corpus. The sentences were « Mon ami a perdu ses bagages à la gare » (“My friend lost its luggage at the railway station”); « Les élèves doivent cocher la bonne case avec un feutre » (“Pupils have to tick the right box with a felt-tip pen”); and « La piscine est couverte d’une bâche en plastique », (“The swimming pool is covered with a plastic sheet”).

The broad transcription of the standard dialect form for the context surrounding the fricative consonant is either /aCV/, /aC#V/, or, although less frequent, /aCə#V/.

3. Segmentation and acoustic cues

The author has segmented the sentences by hand with Praat. In order to take into account the variation of fricative duration due to the boundary type realized by the speaker at the end of NP2, the fricatives have been classified into four groups. In the first group, (G1), fricatives were directly followed by the initial vowel of the last NP (/aCV/); there was no schwa and no pause between the accentual groups. In the second group, (G2), they were followed by a schwa and a pause (/aCə#V/). In the third one, (G3), they were followed by a pause (/aC#V/), and in the fourth one, (G4), essentially observed for German speakers, they were followed by a schwa and the initial vowel (/aCəV/).

We analyzed the values of five acoustic cues. Two cues were related to duration: (1) the duration of the fricative, in milliseconds (DurC), (2) the ratio of the duration of the

preceding vowel to that of the consonant (DurVC). The duration values were provided by segmentation. Two cues were related to the intensity of the consonant (1) Low-FreqE, in dB, which evaluates the intensity of the consonant in low frequencies (0-1000Hz), with respect to that of the preceding vowel in the same frequency region, and (2) High-FreqE, which evaluates the intensity of the consonant in high frequencies (above 2000Hz), with respect to that of the preceding vowel in the same frequency region. Low-FreqE should have higher values for voiced consonants whereas High-FreqE should have higher values for intense noises (unvoiced consonants). We used Praat for the evaluation of intensity. The fifth parameter, periodicity (VR), is estimated by the percentage of voiced frames over the duration of the consonant, (function “Voicing Report”, Praat software). The estimation of periodicity during fricatives, sounds with often intense noises, being especially difficult, we checked the results manually, and corrected them when necessary (there was few corrections, made for obvious mistakes).

In order to estimate the articulation rate, we calculate the average duration of each segment, which is the inverse of the articulation rate (number of speech units –here segments- per second). Note that the pauses are not taken into account to evaluate the articulation rate, contrary to what is done for the speech rate. We choose the segment as the unit to evaluate the articulation rate, and not the syllable, which is more frequently used, for at least two reasons. Firstly, the sentences had relatively different syllable structures (e.g. there is more clusters in sr03 than sr01). Secondly, a segmental unit is more convenient to make comparisons with the frication durations. To evaluate the average length of each segment in a given sentence, we subtract the duration of the various pauses from the duration of the sentence, and we divided the result by the number of segments constituting the sentence.

Table 1 Median values of consonant duration, in ms, as a function of language, boundary types, and category. Values for advanced speakers and beginners, in that order, are given in parenthesis

	G1	G3	All
L1sr01 (/z/)	61	73	64
L1sr03 (/ʒ/)	63	80	66
L2sr01 (/z/)	78 (74-90)	90 (78-97)	84
L2sr03 (/ʒ/)	88 (81-132)	137 (122-154)	109
L1sr07 (/ʃ/)	88	107	88
L2sr07 (/ʃ/)	98 (96-104)	133 (85-153)	122

4. Results

4.1. Consonant duration.

Table 1 shows the median values for consonant duration for voiced and unvoiced consonants produced by German and French speakers, as well as their variation as a function of sentences, speakers’ levels and boundary types (G1 and G3). For the sake of simplicity, we choose to present and discuss results for the first and the third boundary types only, since the other types, G2 and G4 (consonants followed by a schwa) contained relatively few data. GF speakers realized more pauses after the fricatives than FF speakers (twice more pauses, on the

average, for the four series of fricatives) and beginners more pauses than advanced speakers.

We made pairwise comparisons using Wilcoxon rank sum test for the statistical analysis, with Benjamini, Hochberg correction for multiple comparisons. The analysis shows that, for a same group (G1 or G3) and a same voicing feature, voiced or unvoiced, consonants pronounced by native speakers are always significantly shorter than those pronounced by non-native ones (we adopt a statistical significance level of 0.05). Overall medians show that the durations of the voiced consonants produced by German speakers (Median at 92 ms) are well above those of voiced consonants produced by French speakers (65 ms), and slightly higher than the unvoiced consonants (88 ms) produced by these native speakers, which seems to be in agreement with an error in categorization due to L1.

Table 2 Averaged segment length, in ms, as a function of language and sentence. Values for advanced speakers and beginners, in that order, are given between parenthesis

	sr01 (/z/)	sr03 (/ʒ/)	sr07 (/ʃ/)
L1	74	74	75
L2	94 (86-107)	112 (90-122)	98 (86-104)

However, we observe in Table 1 that the three modalities sr03, G3 and beginners (from the sentence, boundary type and level factors, respectively) substantially lengthen the duration of the noises. As a consequence, the values observed for voiced consonants uttered by GF beginners are in general higher than those obtained for FF unvoiced consonants in the same context. These differences are striking for the sentence sr03, both for the first and the third groups. This clearly shows that an error in categorization cannot explain alone the long noise durations observed here. In the same vein, results for GF unvoiced consonants (122 ms) are well above those for FF unvoiced consonants (88 ms), whereas there is no problem of categorization in this case.

Such results can be explained by a decrease in articulation rates, often observed for L2 speakers. Table 2 shows the averaged duration of speech segments, as a function of sentence, language, and GF speakers' level. Results show that 1) the averaged segment duration is substantially higher for German speakers than for French speakers, and for beginners than for advanced speakers, which is in agreement with previous work [20] and 2) concerning German speakers, there is a drastic difference between sentences, values for sr03 being higher than those for sr01 and sr07. Such a difference was not expected since sentences have similar structures. We believe this is due to the complexity of the sentence, which is somewhat higher in sr03 (there are more segments, the syllable structure is in general more complex...). Let us recall that we find the same kind of differences between the duration of the noises for GF sr01 and GF sr03, a difference which is too high to be explained by a difference in place of articulation (see values for French speakers in Table 1, for comparison). We thus made the hypothesis that the differences between the averaged segment length of the sentences, probably due to sentence complexity, have an impact on noise duration for German speakers.

Figure 1 represents the duration of the voiced and unvoiced consonants pronounced by French native and German non-native speakers as a function of the averaged segment length of

the sentence. We observe that in most cases, long fricative durations are associated with long averaged segment lengths. For German speakers, we find a positive correlation between the duration of the voiced consonants and the averaged segment length (Pearson coefficient of correlation: 0.62, adjusted R-squared: 0.38, F-statistic: 47.64 on 1 and 75 DF, p-value: 1.414e-09; results given by the R-software [20]). The increase in averaged segment duration, which corresponds to a decrease in articulation rate, and which varies with levels and sentences can thus explain a part of the increase of the noise duration for L2 speakers.

Table 3 Median values for Low-FreqE as a function of language (L1 or L2), sentence, and boundary types

sentence	sr01 (/z/)		sr03 (/ʒ/)		sr07 (/ʃ/)	
Group	G1	G3	G1	G3	G1	G3
L1	-7.7	-12.4	-7.8	-11	-13.7	-16.4
L2	-10.9	-14.1	-12.3	-16.3	-16.6	-17

4.2. Whole set of acoustic cues and clustering analysis.

We observe for all the cues corresponding to periodicity and noise intensity (VR, Low-FreqE, and High-FreqE) the same phenomena than those reported for the consonant duration (see section 4.1, and Table 1): 1) their values vary as a function of speakers' L1, GF levels, boundary types (G1 or G3) and sentences, and 2) concerning GF voiced consonants, values observed for beginners, G3 and (the sentence) sr03 tend to be more characteristic of an unvoiced fricative than in other contexts. The values for these modalities are sometimes more "extreme" than those of French unvoiced fricatives, which means that the realizations are typical of very long and very intense non-periodic noises. As an example, Table 3 displays the results obtained for the Low-FreqE (energy in low frequencies, see section 3). For GF unvoiced consonants, we also observe that the cue values are often different (more "extreme") than those of French unvoiced consonants.

All the realizations, including data from the boundary types G2 and G4, characterized by the values of all the acoustic cues considered in this study, have been submitted to a hierarchical clustering made on principal components (HCPC, from R, FactoMineR package). This function is an unsupervised method that combines principal components methods, hierarchical clustering and partitioning. The analysis splits the consonants into clusters without any a priori and reveals similitudes and differences between the different realisations. We have limited the number of classes to 4.

The results of the HCPC analysis can be visualized in Fig. 2, which displays the repartition of the four clusters on the axes of the principal component analysis. For each class, the analysis provides statistical data on the quantitative variables, i.e. here the acoustic cues.

The first class (in black in Fig.2) contains essentially non-native data, both voiced and unvoiced. The fourth of all GF voiced data are in this class, with the half of GF unvoiced data (55%). Almost all the FF unvoiced consonants (95%) are in the second class (in red in Fig.2), with the other half of GF unvoiced consonants (45%), and more than a third (38%) of the GF voiced consonants. Finally, the third and fourth class are only made up of voiced consonants: 86% of the French voiced

consonants are in these classes, and 36% of the German voiced consonants, essentially produced by advanced speakers.

In the first class, made up of GF unvoiced and voiced data, the percentage of voiced frames (VR, mean at 14%), the energy in low frequencies (Low-FreqE, mean at -19 dB) as well as the VC duration ratio (DurVC, mean at 1.07) have all very low values. The energy in high frequencies is high (High-FreqE, mean at 9 dB) and the duration of the consonant very important (mean at 162 ms). When we compared all the clusters, the means obtained for the first cluster represent either the lowest values (Low-FreqE, VR, DurVC), or the highest (duration of the consonant, High-FreqE). There are typical of very long (too long for fluent speech), intense, and non-periodic noises.

The clustering analysis confirms the impact of fluency on GF speakers' realizations. The first cluster contained only relatively few FF unvoiced French consonants, but more than half of the German unvoiced consonants, for which there is no problem in categorization with respect to voicing, and the fourth of GF voiced fricatives, all produced by beginners, and in general followed by a pause. In this group, the consonants have very long noises, partly due to the increase in the averaged segment length of the sentences, as shown in the preceding subsection.

5. Concluding remarks

The study explores the impact of fluency, in particular the decrease in articulation rate, and categorization in the production of French final voiced fricatives by German learners, in controlled prosodic contexts. Results shows that the values of the various acoustic cues observed for German voiced consonants, cannot be explained only by categorization, since a substantial number of realizations exhibits more "extreme" values than French unvoiced consonants. In particular the consonants are longer and have a more intense noise. A clustering analysis (HCPC) gathers in a same cluster both half of the unvoiced GF consonants and the fourth of GF voiced consonants (i.e. half of the beginners' realizations) which exhibit values typical of very long, intense and aperiodic noises. For German speakers, the impact of fluency at the sentence level is shown by the existence of a positive correlation between the averaged segment length (the inverse of the articulation rate), and the duration of consonants.

The study shows the mutual effects of an error in categorization and fluency in learners' segmental categories, which 1) raises the question of how to judge their performance (is it possible to separate effects due to fluency from effects due to a possible error in categorization?) and 2) leads, at least for beginners to the realization of "strong-accented" speech segments.

Acknowledgments. I am very grateful to Anne Gégout-Petit, Professor at IECL-UL-Nancy, expert in statistics, for her help with regard to the statistical analysis.

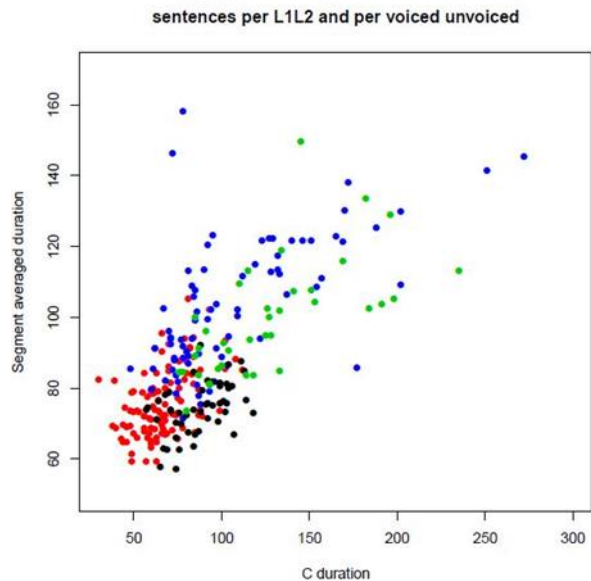


Figure 1 Consonant duration (ms) as a function of averaged segment length (ms). French voiced consonants in red, French unvoiced in black, German voiced in blue, German unvoiced in green

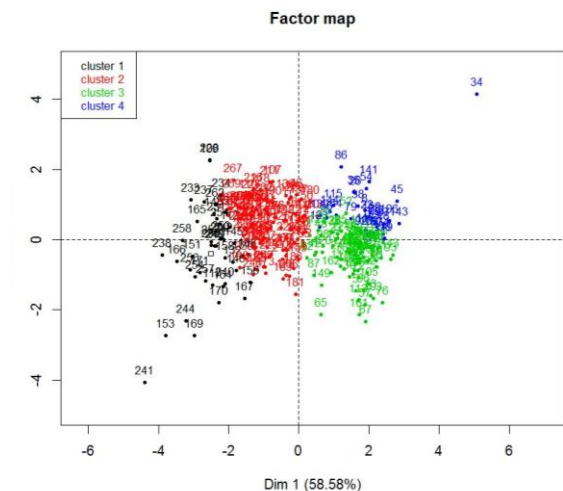


Figure 2 The four clusters of the clustering analysis (hcpc) on the principal component analysis.

6. References

- [1] J. E. Flege, « Production and perception of a novel, second-language phonetic contrast », *J. Acoust. Soc. Am.*, vol. 93, p. 1589-1607, 1993.
- [2] C. T. Best, G. W. McRoberts, et E. Goodell, « Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system », *J. Acoust. Soc. Am.*, vol. 109, n° 2, p. 775-794, 2001.
- [3] J. E. Flege, « Second-language speech learning: theory, findings, and problems », in *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research*, York Press., Timonium, MD, 1995.
- [4] A. Bonneau et M. Cadot, « German non-native realizations of French voiced fricatives in final position of a group of words », in *Interspeech 2015*, Dresde, Germany, 2015.
- [5] U. Gut, *Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt: Peter Lang, 2009.
- [6] M. Baese-Berk et T. Morrill, « Speaking rate consistency in native and non-native speakers of English », *J. Acoust. Soc. Am.*, vol. 138, n° 3, 2015.
- [7] C. Brand et S. Götz, « Fluency versus accuracy in advanced spoken learner language: A multi-method approach », *Int. J. Corpus Linguist.*, vol. 16, n° 2, 2011.
- [8] L. Temple, « Second language learner speech production », *Stud. Linguist.*, vol. 54, n° 2, p. 288-297, 2000.
- [9] K. J. Kohler, « Communicative Functions Integrate Segments in Prosodies and Prosodies in Segments », *Phonetica*, vol. 68, n° 1-2, p. 26-56, 2011.
- [10] C. Kuzla, T. Cho, et M. Ernestus, « Prosodic strengthening of German fricatives in duration and assimilatory devoicing », *J. Phon.*, vol. 35, n° 3, p. 301-320, juill. 2007.
- [11] H. Mitterer, T. Cho, et S. Kim, « How does prosody influence speech categorization? », *J. Phon.*, vol. 54, p. 68 - 79, 2016.
- [12] C. Wellmann, J. Holzgrefe, H. Truckenbrodt, I. Wartenburger, et B. Höhle, « How each prosodic boundary cue matters: Evidence from German infants », *Front. Psychol.*, vol. 3, n° 580, 2012.
- [13] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, et P. J. Price, « Segmental durations in the vicinity of prosodic phrase boundaries », *J. Acoust. Soc. Am.*, vol. 91, n° 3, p. 1707-1717, 1992.
- [14] R. Wiese, *The phonology of German*, Oxford University Press. Oxford, 2000.
- [15] L. Lisker et A. Abramson, « A cross-language study of voicing in initial stops: Acoustical measurements », *Word*, vol. 20, n° 3, p. 527-565, 1964.
- [16] M. Jessen, *Phonetics and phonology of tense and lax obstruents in German.*, John Benjamins. Amsterdam, 1998.
- [17] S. Fuchs, J. Brunner, et A. Busler, « Temporal and spatial aspects concerning the voicing contrast in alveolar and postalveolar fricatives », *Adv. Speech-Lang. Pathol.*, vol. 9, n° 1, p. 90-100, 2007.
- [18] « www.ifcasl.org ». .
- [19] M. Baese-Berk et T. Morrill, « Speaking rate consistency in native and non-native speakers of English », *J. Acoust. Soc. Am.*, vol. 138, n° 3, 2015.
- [20] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna (Austria): R Foundation for Statistical Computing, 2017.