Voicing assimilations by French Speakers of German in stop-fricative sequences

Anne Bonneau

Voicing assimilations by French speakers of German in stop-fricative sequences

Anne Bonneau

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, FRANCE

anne.bonneau@loria.fr

Abstract

Voicing assimilations inside groups of obstruents occur in opposite directions in French and German, where they are respectively regressive and progressive. The aim of the study is to investigate (1) whether non native speakers (here French learners of German) are apt to acquire subtle L2 specificities like assimilation direction, although they are not aware of their very existence, or (2) whether their productions depend essentially upon other factors, in particular consonant place of articulation. To that purpose, a corpus made up of groups of obstruents (/t/ followed by /z/, /v/ or /f/) embedded into sentences has been recorded by 16 French learners of German (beginners and advanced speakers). The consonants are separated by a word or a syllable boundary. Results, derived from the analysis of consonant periodicity and duration, do not stand for an acquisition of progressive assimilation, even by advanced speakers, and do not show differences between the productions of advanced speakers and beginners. On the contrary the boundary type and the consonant place of articulation play an important role in the presence or absence of voicing inside obstruent groups. The role of phonetic, universal mechanisms against linguistic specific rules is discussed to interpret the data.

Index Terms: voicing assimilation, L2, obstruents, French/German interferences, speech rate.

1. Introduction

The phonological contrast opposing fortis to lenis consonants differs a lot in French and German, both at the phonological and phonetic levels. The presence of fortis aspirated stops [1] and the neutralization of the fortis/lenis opposition in final position [2] [3], which characterize German and not French, are well known. It is probably less known that the direction of voicing assimilations occurring between two obstruents C1 and C2, with C1 in final and C2 in initial positions, are different in both languages. Note that, in this paper, the term "voicing" refers to the presence of vocal fold vibration, not to the phonological contrast. In French, in this context, voicing assimilations are in general regressive, and occur from the second consonant to the first one [4] [5]. In German, assimilations between obstruents are progressive and occur from the first consonant, in final position and thus fortis, to a following lenis fricative in initial position [6]. The differences in assimilation directions raise the question as to whether non native speakers modify their L1 direction. We consider here the case of French learners of German uttering C1C2 sequences that elicit progressive assimilation in German (i.e. made up of a fortis consonant followed by a lenis fricative).

The realization of assimilation phenomena, their frequency (the number of occurrences) and magnitude (such as the percentage of segment duration affected by the phenomenon), depend upon a number of factors [7], including notably the prosodic boundaries which separate the potentially assimilated and assimilating segments, and speech rate. The frequency and the extent of assimilations should indeed be more important at weaker boundaries than at higher ones, and should decrease with decreasing speech rates.

Other factors have a direct impact on voicing and affect in turn assimilations, not only their frequency and magnitude, but also the conditions necessary for their emergence. Let us first mention the German phonological system. In word initial position, there is a voicing contrast between /f/ and /v/, but not between /s/ and /z/ (/s/ is not present at word onsets apart from loanwords) [1]. This means that, in this context, /z/ can be devoiced with no or minor impact on intelligibility. Thus, we can expect relatively frequent and (sometimes) important devoicing of /z/, especially when the consonant is preceded by a fortis consonant. Kuzla *et al.* [6] have shown that German speakers indeed devoiced more /z/ than /v/, which can be partly due to the German phonological system. A major source of pronunciation errors is due to orthographic interferences [8]. In word initial position, the letter "s" corresponds to /z/ in German but to /s/ in French, which is a source of errors for beginners. Finally, aerodynamic phenomena also play a role in the devoicing of lenis fricatives. It is indeed difficult to maintain voicing during the production of fricatives [9]. We can thus expect that /z/ (and, to a lesser extend, /v/, which is produced with less frication) will be at least slightly devoiced during their productions due to aerodynamic phenomena.

Assimilations are not taught in foreign language classes and speakers are not aware of their very existence, even in their own language. They might thus be difficult to acquire. The few studies on L2 assimilation acquisition yield conflicting results. On the one hand, the results of Schmidt [10] concerning the acquisition of regressive assimilation of voicing by English learners of Spanish showed that few learners, even the most advanced ones, exhibited L2 assimilation in their productions. On the other hand, in a series of perceptual experiments about assimilation compensation, Darcy *et al*. [11] showed that, whereas beginners applied their native compensation pattern to L2, advanced speakers were able to compensate for L2 assimilation processes. In a preliminary experiment investigating German C1C2 sequences made up, in particular, of a fortis cluster followed by /z/ and pronounced by French learners, we found that about half of the French speakers devoice the alveolar fricative [12]. In the light of what was previously said, we can wonder whether such results can be attributed to the acquisition of German assimilation or to the characteristics of /z/.

The aim of the study, which extends the previous one to all the fricatives present in German at word boundary (/v/, /z/ and /f/) and to different boundary types, is to investigate whether (1) French learners of German are apt to modify such subtle L2 specificities as assimilation direction, or, on the contrary (2) keep their L1 habits, and, finally (3) whether their productions

http://dx.doi.org/10.21437/Interspeech.2021-601

depend upon other factors, in particular prosodic boundary, or consonant place of articulation.

2. Experimental protocol

2.1. Corpus

The corpus is made up of groups of C1C2 consonants, where the first consonant, C1 is the fortis stop /t/ and the second consonant is the fricative /z/, /v/ or /f/. Groups of consonants appear at two different linguistic boundaries 1) inside a carrier sentence, at word final and initial positions and 2) inside words, at syllable final and initial positions. In the first context, the carrier sentence has the following structure « X hat Y gemalt » (« X has painted Y »), where X is a first name and Y is either Wälder, Senken, or Felder, which represents three different sentences. We have taken here a structure very similar to that of a series of sentences chosen by Kuzla *et al.* [6], so as to allow comparisons between results. In the second context, C1 and C2 are respectively in syllable final and initial position inside a word. The three words chosen were «Tatsache » (/tz/ sequence), «etwas » (/tv/ sequence) and «Bootfarhen » (/tf/ sequence).

Sixteen speakers, seven advanced (C1-C2) and nine beginners (A2-B1) recorded the corpus. There was approximately the same number of male and female speakers in each category. There were five repetitions for each sentence of the corpus in the first context (word boundary) and three repetitions in the second context (syllable boundary), for a total amount of 384 groups (4 repetitions on the average x 2 contexts x 16 speakers x 3 consonant categories).

During the recording session, 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). The subjects uttered, in the following order, (1) a series of six sentences to get used to the recording procedure, (2) the sentences with obstruents at word boundary, (3) the words with obstruents at syllable boundary. The sentences and the words were presented in a random order. We used the software Corpus Recorder for calibration and randomization [13].

2.2. Acoustical analysis

Segmentation have been done by hand by the author, using Praat. We have evaluated two acoustic cues: duration and periodicity. Duration, given by segmentation, has been estimated for C1 and C2, their surrounding vowels, and the whole sentence. We analysed the periodicity (glottal pulses) of the signal for each consonant of the consonantal sequence. It has been estimated by Praat "Voicing Report" function which provides the fraction of locally unvoiced frames during a segment, from which we deduced the fraction (the percentages) of voiced frames during the obstruents. We checked the results manually, and corrected them when necessary (there was few corrections, made for obvious mistakes).

3. Results

At word boundary only, there was sometimes a silent pause between the end of the first word, after C1, and the beginning of the second one, after C2. This concerns about 9% of all the realizations (both boundary types confounded). We separated these realizations from the others, and considered three conditions: C1C2 in word condition, with the consonants separated by pause (first condition), C1C2 in word condition, without pause (second condition), and C1C2 in syllable position (third condition).

3.1. Results per consonant.

We build four linear mixed models, one for each numeric factor (C1 and C2 % of voiced frames, C1 and C2 durations), with the numeric factor as the dependent variable; three fixed factors corresponding to the consonant category, the condition and the level, as well as a random factor (the speaker), as the predictors. Note that, for C1 (/t/), the factor "consonant category" corresponds to the category of the following consonant (C2). We then submitted each model to an anova analysis (Satterthwaite's method). For post hoc tests, we used Kenward-Roger's method to evaluate the degree of freedom. All the statistical procedures were made with R[14].

We begin by a short presentation of duration's data and discuss important differences between duration values.

C2 consonant duration. There is a main effect of the category ($F(2, 352)= 60, p < 2.2e-16$) and a two-way interaction between the category and the condition ($F(4, 352) = 6.5, p < 3.8e-05$). As expected, the fortis consonants (median at 130 ms) are significantly longer than others [1] (medians at 95 for /z/ and 79 ms for /v/), this is true in each condition ($p < 0.0001$, in ecah case), with only one exception (/f/ and /z/ in the first condition). Concerning the lenis consonants, we find that /z/ is significantly longer than /v/ in the first ($p < 0.003$) and the third condition ($p < 0.001$), but not in the second.

C1 consonant duration. There is a main effect of the condition ($F(2,352) = 128.7, p <2e-16$), and a two-way interaction involving the level and the condition ($F(2, 352) = 4.2, p < 0.015$). Post hoc analyses determining the effect of the condition show that, for each level, the realizations of /t/ are highly significantly longer in the first condition (presence of a pause) than in each other condition ($p <0.001$, for each comparison). The difference between the second and the third condition is also significant for beginners ($p <0.001$), these speakers producing slightly longer dentals at word boundary than at syllable boundary.

The differences due to the presence or absence of a pause are striking : the median for the three consonants is at 174 ms, in the presence of a pause, and 83 ms in its absence, and there is a very small overlap area between both contexts. This is mainly due to the nature of the boundary, the duration is longer at important boundaries [15], and the nature of the sound since for stops, the burst can be fully released and very long in this context. The importance of durations in the first condition, which are far above those observed by Kuzla *et al.*[6] when /t/ is preceded by a pause, is probaby also specific to L2 realizations.

Percent of voiced frames for C2. There is a significant main effect of the category of the consonant ($F(2,352) = 53.5, p <2.2e-16$), the condition ($F(2,352) = 12.2, p <7.31e-06$) as well as a significant interaction ($F(4,352) = 5.3, p <0.0004$) between both factors. The speakers' level has no significant effect on C2 values. Figure 1 represents, for C2, the % of voicing for each consonant in each condition.

Post hoc analyses examine the effect of the category (a main effect) as a function of the condition, because of the interaction between both factors, and for the same reasons, the effect of the condition as a function of the category. Concerning the effect of the category on voicing values, the analyses reveal that, for a given condition, the differences between the three consonants, /f/, /v/ and /z/, are significant ($p<0.01$, for all significant cases) but in two cases among nine: (1) /v/ and /z/ in the first condition, and, (2) /f/ and /z/ in the third condition.

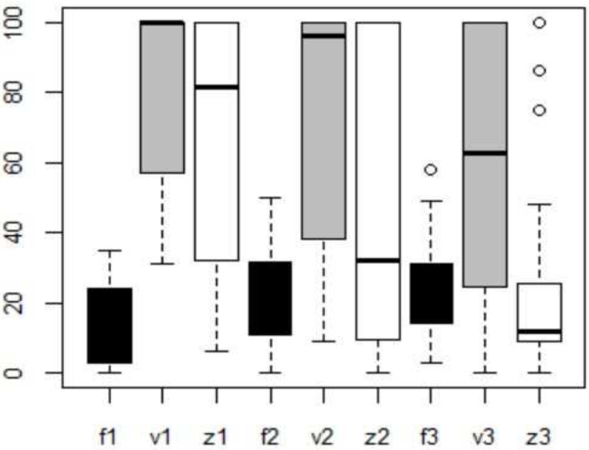

Figure 1. *Boxplots for C2 % of voiced frames as a function of the category and the condition (e.g. "v1" represents /v/ in the first condition).*

We examine now the effect of the condition for each given consonant, which amounts to examining whether the values for each consonant vary significantly as a function of the nature of the boundary. There is no significant difference between conditions for the consonant /f/, and there is also no significant difference between the first and the second conditions for /v/ and /for /z/. On the contrary, the differences between the third condition and each other condition for /v/ and for /z/ are all significant ($p<0.002$, in each case). In these cases, the consonant which is at the lower boundary (third condition) has a lower amount of voicing.

The rank order of the median values calculated for the three consonants is in agreement with what is expected since, for a given condition: (1) /f/, always unvoiced, has a lower median than /v/ and /z/ (but for the third condition, where /f/ and /z/ do not differ significantly, and the median for /z/ is slightly lower than that of /f/) and (2) /v/ is more voiced than /z/ (the difference is not significant in the first condition).

Whereas the medians for /v/ are always above 50% and those for /f/ under 50 %, we remark that the medians for /z/ in the second, and, especially, the third conditions are under 50 % (see Figure 1). Furthermore, there is a very important dispersion of values for lenis consonants with a number of data under 50%, especially for /z/ in the second condition, but also for /v/ in the second and the third condition (as shown in Figure 1). These low values for lenis consonants can be attributed to a progressive assimilation of devoicing. We will come back below to the very low values observed for /z/ in the third condition.

The differences observed between the conditions for the lenis consonants, showing a lower amount of voicing at the lower boundary, reinforces the hypothesis of the presence of a devoicing assimilation phenomenon.

The result obtained for /z/ in the third condition seems strange (at syllable boundary the amounts of voicing for /z/ and /f/ are not significantly different, and the median for /z/, at 17%, is very low). Although this could be interpreted in terms of a very strong devoicing assimilation from /t/ to /z/ at syllable boundary, we tend to think that this is mostly due to an error due to the spelling of /z/, which is written "s". Most learners probably consider "s" as an unvoiced consonant when "s" appears inside a word and is preceded by a fortis consonant.

Percent of voiced frames for C1. As it is the case for C2 voicing, there is no effect of the speakers' level, but a main effect of the category of the following consonant ($F(2,352) = 7.3$, $p<0.0007$), the condition ($F(2,352) = 13.7$, $p<1.7e\text{-}06$) as well as a significant interaction between both factors ($F(4,352) = 6.9$, $p<2.1e\text{-}05$). Figure 2 represents, for C1, the % of voicing for each consonant in each condition.

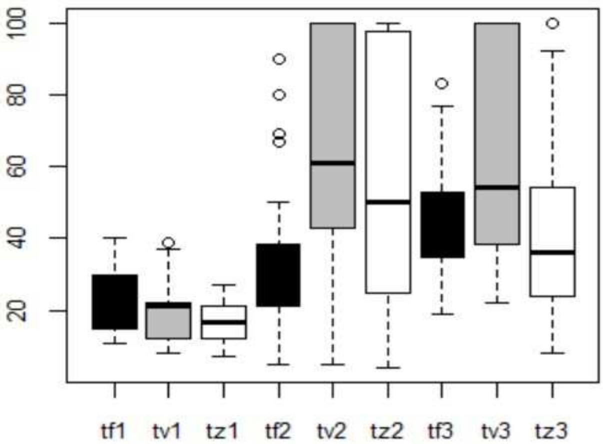

Figure 2. *Boxplots for C1 % of voiced frames as a function of the context and the condition (e.g. "tv1" represents /t/ followed by /v/ in the first condition)*

Let us first examine the effect of the following context (C2 category) on /t/ realizations, per each condition. In the first condition, there is no difference due to C2 category. In the second condition, there are highly significant differences between the realizations of /t/ followed by the fortis /f/, and those followed by each lenis consonant, ($p<0.001$, in each case), and no significant difference between the lenis contexts. In the third condition, there are significant differences between the context /v/ and each other context ($p<0.002$, in each case), and no significant differences between /f/ and /z/ contexts.

What is the impact of the condition on /t/ realizations, for a given consonantal context ? When /t/ is followed by /f/, the differences due to the conditions are either non significant or marginal. When /t/ is followed by /v/, the differences between the first condition and each other condition are very highly significant (all $p<0.001$), whereas there is no difference between the second and the third condition. When /t/ is followed by /z/, all the differences due to conditions are significant (all $p<0.01$).

These results are consistent with the presence of a regressive assimilation of voicing from the lenis consonants to the fortis /t/, when the context is favourable to the emergence of this phenomenon, i.e. when the consonant /t/ is followed by a lenis consonant, and when the consonants C1 and C2 directly follow each other. As already noted, a pause strongly reduces or eliminates assimilation.

In the first condition indeed, where C1 and C2 are separated by a pause, the realizations of /t/ are not affected by the nature (fortis or lenis) of the following consonant. The median values estimated for each context are around 20%, which is characteristic of an unvoiced consonant. In the second condition, whereas /t/ stays unvoiced when followed by /f/, equally unvoiced (median at 32%), the median values observed in lenis contexts (61 and 50%, for /v/ and /z/) clearly indicate that, in most cases, /t/ undergoes a regressive assimilation of voicing. This assimilation is not systematic, as shown by the large dispersion of values in lenis contexts (see Figure 2). In the third condition, we observed the same kind of realizations than in the second condition for /t/ followed by /f/ and /v/, but not for /t/ followed by /z/. In this context, the median (36%) indicated that most realizations of /t/ stays unvoiced. This observation, as well as the highly significant difference between /z/ contexts in word and syllable conditions strengthen our hypothesis of a possible spelling error for /z/ in the third condition.

Finally the highly significant and visible differences between the first condition (presence of a pause) and the other ones for /t/ followed by /v/, and between the first and the second condition for /t/ followed by /z/, is probably due to the absence of assimilation when a pause separates the consonants, and also to the differences in consonant length (/t/ being significantly longer before a pause, as underlined above).

3.2. Results for C1C2 voicing patterns.

After having analysed the voicing of each consonant separately, this section examine the whole C1C2 sequences, inside which assimilation occurs. Before the analysis of the results, it is important to underline here that 1) the extent of assimilation is variable, and only a part of the assimilated consonant might be affected by the phenomenon; studies on lenis fricatives preceded by fortis consonants and pronounced by German speakers [6] [16] indeed show that /v/ and /z/ are not entirely assimilated, 2) assimilation is not systematic, and its occurrence depends upon a number of factors, as mentioned in the introduction.

Table 1 displays, for the sequences /tv/ and /tz/ in condition 2 and 3, favourable to the appearance of assimilation phenomena, the percentages of C1C2 groups for which 1) each consonant is mainly voiced ("VV" groups), 2) each consonant is mainly unvoiced ("UU" groups), 3) the first is mainly unvoiced and the second mainly voiced ("UV") and 4) the first is mainly voiced and the second mainly unvoiced ("VU"). We have considered that a consonant is mainly voiced when its percent of voicing is above 50%, and mainly unvoiced in the other case. We add two configurations, "FU" (fully unvoiced), for which both consonants percentages of voicing are under 20%, and "FV" (fully voiced), for which both consonants percentages of voicing are above 80%.

Table 1 *Percentages of voicing configurations (UU,FU, VV,FV,UV,VU) as a function of C1C2 sequences and condition. See text for more explanations.*

	UU (FU)	VU	UV	VV (FV)
tv cond. 2	17 (2)	10	23	50 (37)
tz cond. 2	43 (17)	8	12	37 (32)
tv3 cond. 3	27 (0)	16	16	41 (39)
tz3 cond.3	68 (29)	19	2	11 (11)

Let us first give results for sequences not mentioned in Table 1. As expected, all /tf/ sequences are mainly unvoiced, since we obtained 100%, 95% and 80% of UU sequences, from condition 1 to condition 3, respectively. As expected too, in the first condition, /tv/ and /tz/ sequences are mainly "unvoiced" for /t/ and "voiced" for C2 (89% of UV sequences for /tv/ and 66% for /tz/).

Concerning /tv/ sequences in condition 2 (see Table 1), speakers produced mainly VV configurations (50%), with relatively few UU sequences (17%). The consonant /v/ stays essentially voiced in the large majority of cases (VV and UV sequences represent 73% of the cases). In condition 3, the balance between voiced and unvoiced sequences is slightly modified to the benefit of unvoiced sequences. Concerning /tz/ sequences in condition 2, we remark that there are less voiced and more unvoiced sequences (37% and 43%, respectively) than in /tv/. In the third condition, in agreement with what was remarked in the previous section, there is a large preponderance of unvoiced UU sequences (68%) against only 11% of VV sequences. The number of cases where the sequences are fully voiced reach about a third of C1C2 sequences for /v/ in condition 2 and 3 and /z/ in condition 2, with a maximum of 39% for /v/ in condition 2.

4. Discussion and concluding remarks

Results show that there is no systematic voicing assimilations in either direction, progressive or regressive, and that there are no significant differences between advanced speakers and beginners. This tends to corroborate the results obtained by Schmidt [10], in her study about L2 assimilation production, and does not correspond to the perceptual results obtained by Darcy *et al.* [11], presented in the introduction. The absence of systematic assimilations in one direction is clearly shown by the results obtained for /v/ and /z/ since, in the second condition, favourable to the presence of assimilation and inside which both consonants are considered by learners as lenis, the ratios between VV (both consonants are mainly voiced) and UU (both consonants are mainly unvoiced) sequences are of 50/17 for /v/ against 37/43 for /z/ (see Table 1). In French, assimilation in obstruent groups is mostly regressive in /tv/ *and* /tz/ sequences. In German, results obtained by Kuzla *et al.* [6], in word condition (more precisely in the "WORD" prosodic boundary, relatively similar to condition 2), exhibited a very important degree of devoicing assimilation for /z/ and, to a lesser extent, also for /v/ (median for the % of voicing frames around 20%), which is often devoiced by native speakers. The degree of devoicing assimilation is more important in Kuzla's study than in ours.

The absence of clear L1 or L2 patterns could be attributed to a progressive acquisition of L2 patterns. But such an explanation is contradicted by the absence of difference between advanced speakers and beginners. We can also consider that the results are explained by more phonetic, universal mechanisms, and less language-specific rules, notably put into evidence, for L1, by Gow [17]. This explanation would be compatible with the differences observed between /tv/ and /tz/ sequences, the first being more voiced and the second more unvoiced, which corresponds to the differences between /v/ and /z/, due to aerodynamics (see introduction), and with the absence of difference between advanced speakers and beginners.

5. Acknowledgements

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, and to all the speakers.

6. References

[1] M. Jessen, *Phonetics and phonology of tense and lax obstruents in German.*, John Benjamins. Amsterdam, 1998.

[2] R. Wiese, *The phonology of German*, Oxford University Press. Oxford, 2000.

[3] R. Port F. and M. O'Dell L., 'Neutralization of syllable-final voicing in German.', *J. Phon.*, vol. 13, pp. 455–471, 1985.

[4] N. Snoeren, J. Segui, and P. Hallé, 'Perceptual processing of partially and fully assimilated words in French', *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 34, no. 1, pp. 193–204.

[5] A. Marchal, *From speech physiology to linguistics phonetics*. Wiley, 2010.

[6] C. Kuzla, T. Cho, and M. Ernestus, 'Prosodic strengthening of German fricatives in duration and assimilatory devoicing', *J. Phon.*, vol. 35, pp. 301–320, 2007.

[7] M. Nespor and I. Vogel, *Prosodic phonology*. Dordrecht: Foris, 1986.

[8] K. Koda, 'Effects of L1 orthographic representation on L2 phonological coding strategies', *J. Psycholinguist. Res.*, vol. 18, no. 2, pp. 201–222, Mar. 1989, doi: 10.1007/BF01067782.

[9] S. Fuchs, J. Brunner, and A. Busler, 'Temporal and spatial aspects concerning the voicing contrast in alveolar and postalveolar fricatives', *Adv. Speech-Lang. Pathol.*, vol. 9, no. 1, pp. 90–100, 2007.

[10] L. Schmidt, 'Contextual Variation in L2 Spanish: Voicing Assimilation in Advanced Learner Speech', *Stud. Hisp. Lusoph. Linguist.*, vol. 7, 2014, doi: 10.1515/shll-2014-1159.

[11] I. Darcy, S. Peperkamp, and E. Dupoux, 'Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners', *Pap. Lab. Phonol.*, no. 9, pp. 411–442, 2007.

[12] A. Bonneau, 'German obstruent sequences by French L2 leaners', Melbourne, 2019.

[13] V. Colotte, *Corpus Recorder*. Nancy: LORIA, 2013.

[14] R. C. Team, *R: A Language and Environment for Statistical Computing*. Vienna (Austria): R Foundation for Statistical Computing, 2017.

[15] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, 'Segmental durations in the vicinity of prosodic phrase boundaries', *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1707–1717, 1992, doi: http://dx.doi.org/10.1121/1.402450.

[16] C. Shih, B. Möbius, and B. Narasimhan, 'Contextual effect on consonant voicing profiles: a cross-linguistic study', San Fransisco, 1999.

[17] D. Gow W., 'Feature parsing: Feature cue mapping in spoken word recognition', *Percept. Psychophys.*, no. 65, 2003.