Perceptual and articulatory factors in German fricative assimilation

Marianne Pouplier[1], Philip Hoole[1], Jona Cederbaum[2], Sonja Greven[2], Manfred Pastätter[1]

[1]*Insitute of Phonetics and Speech Processing, LMU Munich, Germany*
[2]*Department of Statistics, LMU Munich, Germany*

pouplier|hoole|manfred@phonetik.uni-muenchen.de
Jona.Cederbaum|sonja.greven@stat.uni-muenchen.de

Abstract

We present data from an EPG experiment on German fricative assimilation. It has been claimed that fricative sequences other than sibilant clusters do not assimilate due to perceptual constraints. We demonstrate that f#sibilant sequences show in principle the same kind of temporal overlap as sibilant clusters do, but due to the labial constriction dominating the acoustics, this temporal overlap is acoustically and perceptually less salient. Our data further reveal an order asymmetry: sibilant#f behave differently to f#sibilant clusters; there is no evidence in sibilant#f clusters for the labio-dental constriction overlapping the sibilant in our data. We consider perceptual and articulatory accounts of this asymmetry. We also investigate whether lexical stress affects assimilation patterns. To that effect we discuss a new statistical method for analyzing functional data with a mixed model allowing for multiple covariates and crossed randomeffects. We find that primarily stress of the word-final but not the word-initial syllable interacts with assimilation.

Keywords: fricative assimilation, German, EPG, acoustics, overlap, perception, mixed models, functional data

1. Introduction

Assimilation in fluent speech continues to be a main topic of research in the speech sciences since a complex interaction of articulatory, perceptual, and grammatical aspects conditions assimilatory patterns. Our current focus is on German fricative assimilation since this area is particularly well-suited to gauge the contribution of each of these factors to word-boundary assimilation. Fricatives have been hypothesized to be generally resistant to assimilation. For instance, in German /f#ʃ/ or /f#s/ will not assimilate to a sibilant sequence /ʃʃ/ or /ss/. This has been ascribed to perceptual constraints on assimilation: Final fricatives are perceptually salient therefore blocking assimilation, with the premise being that assimilation preferably occurs in perceptually low-salient environments (Byrd, 1996; Hura, Lindblom, & Diehl, 1992; Kohler, 1990). Thereby sibilants are an exception among the fricatives: an alveolar sibilant assimilates regressively to a following palatal sibilant (e.g., aus Schalke ('from Schalke') => auʃalke, but: auf Schalke => aufʃalke; not *auʃalke). Here reduction of articulatory complexity seems to override perceptual constraints, yet it is difficult to explain the difference between /f#ʃ/ and /s#ʃ/ sequences. Final labio-dental fricatives are arguably not very salient (Miller & Nicely, 1955), and thus other factors may be at play in assimilating /s#ʃ/ and apparently non-assimilating /f#ʃ/ sequences. To what extent final fricatives other than sibilants assimilate has, to our knowledge, not been tested systematically. In the present study, we pursue the possibility that /f#s/ and /f#ʃ/ fricative sequences may show articulatory overlap just like /s#ʃ/ sequences, yet due to the difference in articulators involved there may be few acoustic and auditory consequences. The absence of perceived assimilation in such /f/+sibilant sequences might then simply be due to nonlinearities in the articulatory-acoustic relationship rather than perceptual constraints on cross-word boundary coordination of fricatives. To test our hypothesis, we recorded electropalatography (EPG) and acoustic data for German with a variety of abutting fricatives. In order to investigate the further contribution of articulatory and perceptual factors to assimilation patterns, we also included two vowel and four stress conditions. Particularly stress should contribute to the relative perceptual and articulatory salience of a consonant. We investigate the possibility that in a C1#C2 sequence, an unstressed C1 should more likely be subject to assimilation whereas unstressed C2 should be less likely to trigger assimilation.

2. Methods

We recorded synchronized acoustic EPG data from 9 native speakers of German. All of the speakers were colleagues at the Institute of Phonetics, but naive as to the purposes of the experiment except for S1, the first author of this paper. EPG data were sampled at 200 Hz, acoustic data at 32768 Hz. Subjects were given plenty of time to practice with their palate and an accommodation phase preceded the actual recording.

2.1. Stimuli and experimental procedure

Stimuli consisted of noun-noun compound phrases embedded in a neutral carrier sentence. The stimuli combined the fricatives /f, s, ʃ/ as C1#C2 sequences in final and initial position rendering three cluster conditions: /f#s, s#f/; /f#ʃ, ʃ#f/; and /s#ʃ, ʃ#s/. Homorganic combinations served as controls (/f#f, s#s, ʃ#ʃ/). Note that in Standard High German, the initial alveolar sibilant is voiced, but voiceless in the Southern dialectal regions. Since all of our subjects realized the fricative as voiceless regardless of their dialectal background, we uniformly use the symbol /s/ for the alveolar sibilant here. In the following, we will refer to any analyses involving combinations of /s/ and /ʃ/ in either order as sibilant condition (s#ʃ, ʃ#s), and combinations of sibilant and /f/ in either order as f-condition (s#f, f#s, ʃ#f, f#ʃ). All target words were bisyllabic; we will refer to the syllables containing the fricatives as final and initial target syllable. Two different vowel contexts (i-a, a-i) were included. In the i-a condition, the vowel preceding the cluster was /i/ or /ɪ/ and the vowel following the cluster was /aː/ or /a/ (e.g., ['daː.tif#'ʃa.lə]), and correspondingly for the a-i condition (e.g., ['ku.kaf#'ʃɪm.məl]). Lexical stress of the target syllables was varied to be either stressed or unstressed, which we will abbreviate here as S (strong) and W (weak), rendering four stress conditions (SW, WS, SS, WW). The experimental variables were fully crossed. Stimuli were presented 5 times in blocks randomized differently per subject per block, while ensuring that there were no immediately consecutive trials of identical clusters in different stress conditions.

Subjects read the target sentences as presented to them on the screen at a self-chosen rate. They were familiarized with the sentences ahead of the experiment and completed a practice round containing all stimuli wearing their EPG palate. Targeted token total amounted to 9 consonant sequence x 2 vowel x 4 stress conditions x 5 repetitions x 9 subjects = 3240 tokens. Due to a coding error, S1 had no data for the sibilant control condition /ʃ#ʃ/, i-a, strong-strong, with the consequence that experimental conditions associated with these controls had to be excluded (i-a, strong-strong, /ʃ#f, f#ʃ, s#ʃ, ʃ#s/). Across subjects and conditions another 8 tokens were missing due to technical failure, leaving a total of 3207 tokens for analysis.

2.2. Data treatment

The acoustic data were downsampled to 24kHz and Thomson multi-taper spectra were computed with a 21.3ms window length, 75% overlap (Thomson, 1982). For each trial, the intervocalic fricative interval was segmented acoustically and scaled onto a time interval of [0, 1]. We used the 25% time point for statistical analyses. Both EPG and acoustic data were normalized following Pouplier et al. (2011): The temporal midpoint of the control condition served to created a reference pattern for each consonant relative to which all samples were normalized such that the data ranged between -1 and 1. We further corrected for inherent differences in how well the reference pattern mapped onto the control conditions.

For the sibilant condition, a value of 1 indicates close proximity to the /s/ reference pattern, while a value of -1 indicates close proximity to the /ʃ/ reference pattern. For the f-conditions, a value of 1 indicates close proximity to the /f/ reference pattern, while a value of -1 stands for reference /s/ or reference /ʃ/, respectively, depending on condition.

3. Results

3.1. Overall results

Our first step at data evaluation was to investigate the presence of assimilation for the sibilant and f-conditions overall, collapsing across all vowel and stress conditions. In order to identify whether assimilation had occurred at the 25% time point of the fricative interval, a classification algorithm in the form of a support vector machine (svm; Baayen (2008)) was trained in R on the extracted parameters of the control data (closed test). We then tested the heteroganic conditions against the controls. For example, for the sibilant cluster condition, the svm was trained on /s#s/ and /ʃ#ʃ/ control sequences. For a given experimental /s#ʃ/ token, at the 25% time point of the fricative interval, the svm algorithm should classify the token as /ʃ/ if regressively assimilated, otherwise the token should be classified as /s/. The training/classification was performed for each of the three cluster conditions ([sʃ/ʃs], [sf/fs], [ʃf/fʃ]) separately, once for the EPG and once for the acoustic data.

Table 1 gives the results of the svm classification for all conditions, for both EPG and acoustic data. For the sibilant condition, the table gives percent classified as /s/; a percentage of 100 means that all tokens were classified as /s/, a percentage of zero means that all tokens were classified as /ʃ/. The homorganic control conditions show excellent classification accuracy for both signal types (Table 1 cells A1-2; B1-2). For heteroganic sibilant sequences (Table 1 A3-4, B3-4), the pattern corresponds to what has been reported in recent studies for English (Niebuhr, Clayards, Meunier, & Lancia, 2011; Pouplier et al., 2011): there is a strong tendency for regressive assimilation for /s#ʃ/ (only 57 and 58% percent of data are classified as /s/ at the 25% time point, Table 1 cells A3, B3). For /ʃ#s/ clusters, there is some influence of /s/ on the palatal sibilant, yet only in about 10% of cases (9%, 12% in Table 1, cells A4, B4). Importantly, the results for the EPG and acoustic data are in close correspondence, the assimilation is evident in both the articulatory and acoustic domain.

We now turn to the f#sibilant sequences. For all sequences, Table 1 gives the percent classified as /f/. First we take note of the classification accuracy for the controls for both cluster conditions and both signal types (Table 1, rows C-F, columns 1-2). For the heteroganic conditions, first consider /f#ʃ/. The acoustic data (Table 1, C3) show a comparatively small tendency for assimilation to occur with 77% of tokens having been classified as /f/, the remaining 23% as /ʃ/. Yet there is a marked discrepancy between acoustic and EPG data (Table 1, D3): For the latter, only 47% of tokens were classified as /f/ at the 25% time point, i.e. 53% were classified as /ʃ/. This assimilation rate is comparable to the one of /s#ʃ/ sequences (58% classification as /s/, 42% as /ʃ/, cell B3). We interpret this discrepancy to the effect that while the /ʃ/ constriction is being formed behind the labio-dental constriction, the anterior labio-dental constriction dominates the acoustics.

We now turn to the opposite order /s,ʃ#f/ to find a third pattern of results. Note that there is no assimilation whatsoever in our data: the percentage of tokens classified as /f/ is on the same scale as the control conditions (Table 1, rows C-F, column 4); there is close agreement between acoustics and EPG classification results.

Table 1: Svm classification results for acoustic and EPG data for all cluster conditions.

		1	*2*	*3*	*4*
		homorg.		*heterog.*	
		% classification as /s/			
		s#s	*ʃ#ʃ*	*s#ʃ*	*ʃ#s*
A	*acoustic*	*100*	*0*	*57*	*9*
B	*EPG*	*100*	*0*	*58*	*12*
		% classification as /f/			
		f#f	*ʃ#ʃ*	f#*ʃ*	*ʃ#*f
C	*acoustic*	*100*	*0*	*77*	*2*
D	*EPG*	*99*	*0*	*47*	*1*
		% classification as /f/			
		f#f	*s#s*	f#*s*	*s#*f
E	*acoustic*	*99*	*0*	*84*	*1*
F	*EPG*	*100*	*0*	*53*	*0*

3.2. Evaluating the data using mixed modelling for functional data

We included two vowel and four lexical stress conditions in order to investigate whether assimilation would be more likely in certain vowel contexts and stress conditions. Evaluating such a complex experimental design with multiple covariates and crossed random effects statistically has been a long-standing problem in the speech sciences, particularly when aiming to take into account the articulatory/acoustic dynamics throughout the entire consonant interval. We present a new approach to functional data using mixed models based on an extension of Greven et al. (2010). Due to space limitations, we demonstrate for the sibilant cluster condition only (acoustic data) how the effects of order, vowel, and stress can be assessed within a single model with crossed random effects.

Instead of relying on a single 'magic moment' for analysis as we have done for the svm classification, the statistical model discussed now evaluates the evolution of the acoustic index over the entire fricative interval for each token. Only the heteroganic conditions are used for statistical modelling; the experimental conditions are evaluated relative to each other (N=707). In contrast to other functional data analysis methods (e.g., Ramsay & Silverman, 2005), the model is able to accommodate irregularly spaced functional data, i.e. while the time scale of each index curve was scaled to a [0, 1] interval, the data were not resampled to an across-tokens regular grid.

The stress conditions were dummy coded into two covariates: Stress1 (stress of the final target syllable (C1): 0 strong, 1 weak) and Stress2 (stress of the initial target syllable (C2): 0 strong, 1 weak). The other dummy coded covariates were Consonant Order (0: /s#ʃ/, 1: /ʃ#s/), and Vowel (0: i-a, 1: a-i). Recall that we calculated the index values such that they range between -1 and 1 with -1 being a reference /ʃ/ and 1 denoting a reference /s/ acoustic pattern. In order to be able to compare the index time series for the two consonant orders (/s#ʃ/, /ʃ#s/) directly, the index time series of /ʃ#s/ was mirrored along the time axis such that both C1#C2 conditions showed in principle an index dynamic ranging from 1 for C1 to -1 for C2. Figure 1 shows the resulting index curves for the two consonant orders by speaker, across vowel and stress conditions.

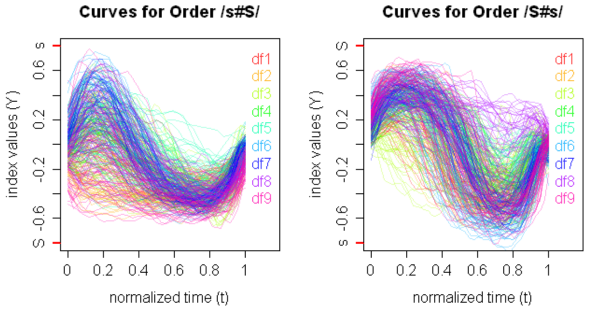

Figure 1: Index curves as they entered the statistics for /s#ʃ/ (left) and /ʃ#s/ (right) by speaker, across conditions.

The model used to analyze the data is given in (1):

$$Y_{ijht} = \mu(t, x_j) + B_i(t) + C_j(t) + E_{ijh}(t) + \varepsilon_{ijht} \tag{1}$$

with Y_{ijht} being the index over time for speaker i, item j, and repetition h observed at time $t \in T \subseteq [0, 1]$. $\mu(t, x_j)$ is a curve specific smooth mean function, x_j are known covariates and possible interactions of covariates. $B_i(t)$ and $C_j(t)$ are random functional intercepts for speaker and items, respectively. $E_{ijh}(t)$ is a speaker-, item-, and repetition-specific smooth random deviation and also includes the interaction between speaker and item. ε_{ijht} is white noise measurement error. The mean function $\mu(t, x_j)$ is specified as in (2):

$$\begin{aligned}\mu(t, x_j) &= \mu_0(t) + f_1(t)\cdot Order + f_2(t)\cdot Stress1 + f_3(t)\cdot Stress2\\ &+ f_4(t)\cdot Vowel + f_5(t)\cdot Order\cdot Stress1 + f_6(t)\cdot Order\cdot Stress2\\ &+ f_7(t)\cdot Order\cdot Vowel,\end{aligned} \tag{2}$$

with $\mu_0(t)$ + $f_1(t)$, ..., $f_7(t)$ as unknown fixed functions; functional random effects were modelled using functional principal components analysis (Ramsay & Silverman, 2005).

Point-wise confidence bands show a significant effect for all covariates. Of the interactions, significant effects were observed only for Order · Stress1. Figure 2 shows the estimated reference group mean function, $\mu_0(t)$, and the effects of covariates Order, Stress1, and the effect of their interaction, each with point-wise confidence bands. The reference group mean corresponds to dummy coding 0 of all covariates (see equation (2)), i.e. Order = /s#ʃ/, Vowel = i-a, Stress1 = strong, Stress2 = strong. The effect on the index trajectory is denoted by Δ Index values (ΔY). The effect of a covariate on the mean can be obtained by multiplying the covariate effect with the dummy coding (1 or 0) and adding it to the reference group mean. The covariates Vowel and Stress2 (not shown here) mostly affected the transition between the sibilants.

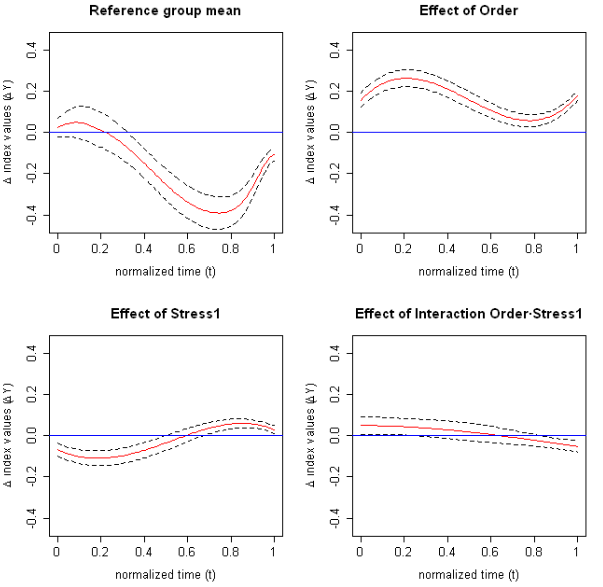

Figure 2: Effects (solid red) and point-wise confidence bands (dashed black) of reference group mean (upper left), covariate Order (upper right), covariate Stress1 (lower left), and interaction between Order and Stress1 (lower right). The reference group mean corresponds to dummy coding zero of all covariates.

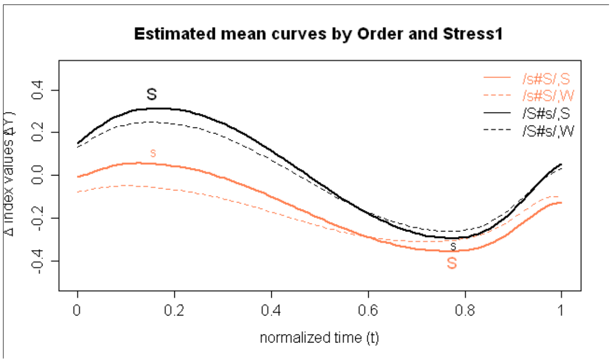

Figure 3: Estimated mean curves by Order and Stress1 with interaction. Solid red curve: Order /s#ʃ/ and Stress1=0 (S). Dashed red curve: Order /s#ʃ/ and Stress1=1 (W). Solid black curve: Order /ʃ#s/ and Stress1=0 (S). Dashed black curve: Order /ʃ#s/ and Stress1=1 (W). The other two covariates are set to their mean (0.5).

The effect of Order is positive over the whole sibilant interval (Fig 2 top right). Recall also that we mirrored the /ʃ#s/ curves such that both C1#C2 conditions have a reference index dynamic ranging from 1 for C1 to -1 for C2. Since Order /ʃ#s/ was dummy coded with 1, Figure 2 shows that the mean curve is pulled during C1 more towards the ideal reference pattern (index = 1) for /ʃ/ than for /s/; for C2, /ʃ#s/ is slightly further away from the ideal reference pattern (-1). This is also evident in the estimated means in Figure 3 in that the /ʃ#s/ curves consistently lie above the /s#ʃ/ curves. For unstressed final target syllables (Stress1=1; Fig 2 lower left), index values in the beginning of the sibilant interval are (on average) lower and values in the end of the interval are (on average) slightly higher compared to stressed final target syllables. The effect is stronger on the final syllable (C1) than on the initial syllable (C2). This may be considered to be an artefact of the stress grouping, since Stress1 groups according to the stress of the final target syllable. Yet the Stress2 pattern which groups by initial target syllable stress (not shown here) speaks against this interpretation: Stress2 has an entirely different effect on the index trajectory compared to Stress1 in that there are no significant effects at either the beginning or end of the sibilant interval, rather, the transition in the temporal mid-region of the interval shows slightly raised index values, i.e. a shallower transition. Overall, we see a pronounced difference in index values between stressed and unstressed final target syllables. The Order · Stress1 interaction can be interpreted with reference to Figure 3: At the beginning and end of the sibilant interval, Stress1 has a slightly different effect on the mean curve for Order /ʃ#s/ than for Order /s#ʃ/. Overall, there is a greater stress effect on C1 for /s#ʃ/ than for /ʃ#s/ from which we conclude that stress interacts with assimilation.

4. Discussion and conclusion

We have presented acoustic and EPG data on German fricative assimilation with the aim of uncovering the contribution of articulatory and perceptual factors to fluent speech phenomena. For one, we could show that in the case of word final labio-dental fricative /f#ʃ/ and /f#s/, the sibilant overlaps the labio-dental in time as has been shown to be the case for stop articulations (Browman & Goldstein, 1990). We found a pronounced asymmetry between articulation and acoustics: while the articulatory data revealed the overlapping fricative constrictions, the acoustic classification showed a much lesser effect of the sibilant: fewer tokens were classified as assimilated acoustically than articulatorily. Overall, our data support the assumption that there is a similar degree of gestural overlap for /s#ʃ/ and /f#s,ʃ/ sequences, yet due to the different articulators being involved, there are few acoustic consequences of the sibilant constriction formation during the /f/, since the sound source of /f/ is anterior to the overlapping sibilant. For sibilant sequences, both consonants call on the same articulators, leading to blended articulations or, presumably due to its relatively greater dorsal control, a dominance of /ʃ/ (Pouplier et al., 2011).

There was further an order asymmetry in the data which at first blush may point to perceptual constraints: For sibilant#f clusters, there was no evidence for the labial fricative encroaching on the sibilant. Several recent publications have looked into the order asymmetry of sibilant assimilation and have found that /ʃ/ is prone to dominate in case of gestural overlap (Niebuhr et al., 2011; Pouplier et al., 2011; Recasens & Mira, 2013). For English /ʃ#s/ sequences, the (near-)lack of regressive assimilation was attributed to /s/ overlapping /ʃ/ having almost no consequences for the articulation / acoustics of the palatal sibilant due to its tighter, more holistic tongue control. Our results for the sibilant condition are consistent with this interpretation; the German pattern is quite similar to the English one. Yet this scenario obviously cannot apply to the f-conditions, since in that case largely independent articulators are involved. For one, perceptual factors may be at play: assimilating a sibilant to /f/ might be perceptually salient, therefore the sibilant may be protected from temporal overlap by the labio-dental. For the reverse order this argument is void, since the labio-dental fricative due to its anterior constriction is protected from assimilation acoustically. This would support the role of perceptual factors in fricative assimilation, even though in a more differentiated manner than has been proposed previously. However, with EPG data we have no positive information about the constriction formation during /f/. If the labio-dental constriction is beginning to be formed during the sibilant, we will see no record of this in the EPG data. Also acoustically, this will have virtually no consequences because the acoustic properties of the frication noise will greatly depend on the point of biggest pressure drop-off which is the point of narrowest constriction. That is, while the sibilant constriction is at its maximum and /f/ is in the process of being formed, there will be little evidence for /f/ formation in the acoustics. It may thus very well be the case that /f/ overlaps a preceding sibilant to some degree, but we cannot tell from our data. In short, it is possible that the same kind of planned overlap has in our type of data observable articulatory consequences for f#sibilant, but not for sibilant#f sequences. This issue will have to be pursued using recording techniques which give information on jaw and lip movement or articulatory modelling. Finally, we have presented a novel statistical approach for mixed modelling of irregularly sampled functional data. This analysis revealed that sibilants are generally sensitive to lexical stress in final position, but less so in initial position. Stress also had a more pronounced effect of final /s/ than on final /ʃ/, supporting the assumption that stress interacts with assimilation.

5. Acknowledgements

Work supported by the ERC under the EU's 7th Framework Programme (FP/2007-2013)/Grant Agreement n. 283349-SCSPL and DFG Emmy Noether grants PO 1269/1-1 to M. Pouplier and GR 3793/1-1 to S. Greven. Thank you to our colleagues who took part in our experiment and to S. Waltl for help with acoustic segmentation.

6. References

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Browman, C., & Goldstein, L. (1990). Tiers in Articulatory Phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech.* (pp. 340-376). Cambridge: CUP.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24*, 209-244.

Greven, S., Crainiceanu, C., Caffo, B., & Reich, D. (2010). Longitudinal Functional Principal Component Analysis. *Electronic Journal of Statistics, 4*, 1022 - 1054.

Hura, S. L., Lindblom, B., & Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35*(1-2), 59-72.

Kohler, K. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 69-92). Dordrecht: Kluwer.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*(338-352).

Niebuhr, O., Clayards, M., Meunier, C., & Lancia, L. (2011). On place assimilation in sibilant sequences - comparing French and English. *Journal of Phonetics, 39*(3), 429-451.

Pouplier, M., Hoole, P., & Scobbie, J. (2011). Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. *Journal of Laboratory Phonology, 2*(1), 1-33.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis.* New York: Springer.

Recasens, D., & Mira, M. (2013). An articulatory and acoustic study of the fricative clusters /sS/ and /Ss/ in Catalan. *Phonetica, 70*, 298-322.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE, 70*(9), 1055-1096.