Original Paper

Phonetica 2016;73:52–78
DOI: 10.1159/000442590

Received: May 4, 2015
Accepted: November 11, 2015
Published online: February 9, 2016

Articulatory and Acoustic Characteristics of German Fricative Clusters

Marianne Pouplier Philip Hoole

Institute of Phonetics and Speech Processing, LMU, Munich, Germany

Abstract

Background: We investigate the articulatory-acoustic relationship in German fricative sequences. We pursue the possibility that /f/#sibilant and /s#ʃ/ sequences are in principle subject to articulatory overlap in a similar fashion, yet due to independent articulators being involved, there is a significant difference in the acoustic consequences. We also investigate the role of vowel context and stress. ***Methods:*** We recorded electropalatographic and acoustic data from 9 native speakers of German. ***Results:*** Results are compatible with the hypothesis that the temporal organization of fricative clusters is globally independent of cluster type with differences between clusters appearing mainly in degree. Articulatory overlap may be obscured acoustically by a labiodental constriction, similarly to what has been reported for stops. ***Conclusion:*** Our data suggest that similar principles of articulatory coordination underlie German fricative clusters independently of their segmental composition. The general auditory-acoustic patterning of the fricative sequences can be predicted by taking into account that aerodynamic-acoustic consequences of gestural overlap may vary as a function of the articulators involved. We discuss possible sources for differences in degrees of overlap and place our results in the context of previously reported asymmetries among the fricatives in regressive place assimilation.

Introduction

The goal of the current paper is to shed light onto the articulatory-acoustic relationship in German fricative sequences involving /f, s, ʃ/. We pursue the hypothesis that the fricative clusters are articulatorily coordinated in a generally similar fashion independently of their exact segmental composition. However, due to nonlinearities in the acoustic-articulatory relationship, the same kind of articulatory overlap can have very different acoustic and auditory consequences, depending on the particular fricatives involved. Specifically, a labiodental fricative may acoustically obscure ongoing constriction formation of a following lingual fricative, similarly to what has been reported in the context of stops. We hypothesize that our results may serve as an explanation as

KARGER

E-Mail karger@karger.com
www.karger.com/pho

0031–8388/16/0731–0052
$39.50/0

Marianne Pouplier
Institute of Phonetics, LMU
Schellingstrasse 3
DE–80799 München (Germany)
E-Mail pouplier@phonetik.uni-muenchen.de

to why in German labiodental fricatives are seemingly exempt from regressive place assimilation.

Interactions between adjacent consonants have been a major topic in speech production research as much as in theoretical phonology, since grammatical, articulatory and perceptual factors combine to condition a variety of adaptation processes. The relative contribution of these factors to the range of patterns observed in any particular instance has been a matter of controversy, likewise the question of whether these phenomena are to be situated in the cognitive-symbolic or phonetic domain. Consonantal interactions have mostly been studied within the context of assimilation and coarticulation research. For a long time, assimilation in which a word-final consonant takes on properties of a following initial consonant was treated as a purely phonological feature-spreading formalism by which association lines are added and deleted independently of an utterance's phonetic characteristics. Coarticulation on the other hand was considered as a primarily physiologically rooted, contextual adaptation of neighboring sounds. A growing number of studies since the 1990s has discarded both views as too narrow, since they fall short of accounting for evidently gradient assimilatory patterns on the one hand, and systematic language- and speaker-specific variation in coarticulation on the other. Acoustic-articulatory work has shown that the concept of assimilation as a purely symbolic, categorical alteration of a phonological plan fails to predict many cases in which assimilated sequences show evidence of blending, meaning that both the trigger and the target unit of assimilation shape the vocal tract at the same time. This has led to the proposal that assimilation is similar, possibly tantamount to coarticulation, since gradient assimilatory patterns are the result of spatiotemporal overlap of successive consonants. For example the fricative cluster in a phrase like *claps Shaun* can be pronounced with articulatory and spectral characteristics intermediate between /s/ and /ʃ/ (Nolan et al., 1996; Zsiga, 1995). Relatedly, for cases in which the overlapping constrictions of stops are performed by relatively independent articulators (e.g. coronal-labial), perceived categorical deletion may be the result of gestural overlap leading to gestural hiding. For instance *perfect memory* can audibly be pronounced as perfec[km]emory, with the final /t/ of *perfect* being perceptually obscured by the following labial stop, yet articulatory records have revealed that the coronal can still be fully produced (Browman and Goldstein, 1990, p. 364, particularly fig. 19.13). Generally, instances of complete assimilation, consistent with a symbolic restructuring account, and instances of gradient gestural blending stand side by side and are observed in the same experiment, for the same stimulus items, within and across speakers. Whether the complete assimilations can adequately be accounted for as end points of a graded assimilation continuum or are qualitatively different from 'intermediate' productions remains a matter of controversy. For general discussions of these issues see, among others, Farnetani and Recasens (2010), Harrington et al. (2013), Iskarous et al. (2012), Niebuhr et al. (2011), Ohala (1993), Pouplier et al. (2011), Recasens and Pallarès (2001) and Zsiga (2006).

It is well known that there are systematic asymmetries in consonant interactions, such as final coronals in many languages being target but not trigger of place assimilation (e.g. /t#k/ > /k#k/, but /k#t/ > */t#t/), or temporal overlap differing in C1C2 sequences depending on the manner of C2 (Bombien et al., 2013). Various accounts of these kinds of asymmetries have been proposed from both phonological and phonetic standpoints, including phonological underspecification, markedness, coarticulation resistance, frequency and collocational probability as well as approaches emphasizing the impact of perceptual constraints (Byrd, 1992; Hura et al., 1992; Jaeger and Hoole, 2011; Jun, 2004; Kohler, 1990; Niebuhr et al., 2011; Pouplier et al., 2011; Recasens, 2006; Recasens and Mira, 2013; Son et al., 2007; Steriade, 1995; Zsiga, 1995, 2011). The argument that assimilation patterns are actively shaped by listener-oriented constraints is common to strands of research that recognize assimilation and coarticulation as phonetics-phonology interface phenomena, independently of the details of the assumed assimilatory mechanism, and has also been used to explain different patterns of coarticulation in consonant clusters.

For our current paper, we focus on C1C2 interactions in German fricative sequences, in particular clusters combining /s, ʃ/ versus clusters involving /s/ or /ʃ/ in combination with a preceding or following labiodental. We thereby subscribe to the view that assimilation is an inherently continuous phenomenon caused by varying degrees of temporal overlap and is thus in principle on a par with coarticulation, even though we acknowledge the open issue of the status of 'complete' assimilations mentioned above. Our main question and result do not hinge on resolving this issue, since we are primarily interested in whether the fricative clusters investigated here pattern the same. Nonetheless, we link a contrasting analysis of the articulatory-acoustic patterning of these clusters to previous reports of /f#ʃ, s/ but not /s#ʃ/ sequences being exempt from regressive place assimilation (Kohler, 1990). For example, *aus Schalke* ('from Schalke') can be pronounced as /aʊʃalkə/ but there will be no /f/ > /ʃ/ assimilation in *auf Schalke* (/aʊfʃalkə/, not */aʊʃalkə/). This has in previous publications been adduced as a case in point in which functional considerations, namely perceptual recoverability requirements, condition asymmetric patterns in consonantal interactions. Generally, it has repeatedly been proposed that assimilation will mostly be observed in cases in which the final consonant is of low perceptual salience, and speakers can afford to save articulatory effort without endangering successful communication (Byrd, 1992; Hura et al., 1992; Jun, 1995; Steriade, 2001). In this view, final coronal stops are considered perceptually weak compared to final velar or labial stops, hence the well-known order asymmetry in consonant stop assimilation.

For German fricative clusters, the absence of articulatory overlap in /f/-sibilant but not sibilant-sibilant sequences has to our knowledge not been tested systematically, and it is problematic to explain the lack of regressive place assimilation in /f#ʃ/ sequences on the basis of perception. Final labiodental fricatives are not very salient (Babel and McGuire, 2013; Miller and Nicely, 1955), and, indeed, the actually observed assimilation pattern for /f#ʃ/ and /s#ʃ/ is in fact the opposite to what is predicted based on the perceptual hypothesis. There is audibly very salient assimilation for the /s#ʃ/ sequence, and no assimilation in the less salient case. Thus, other factors may underlie the patterning of /s#ʃ/ and /f#ʃ/ sequences. In the present study, we pursue the possibility that /f#s/ and /f#ʃ/ sequences may overlap in the temporal domain just like /s#ʃ/ sequences, yet due to the difference in constriction locations with few acoustic and auditory consequences. The perceived absence of assimilation in /f#s/ and /f#ʃ/ sequences would in this case be due to the articulatory-acoustic relationship of consonant sequences involving labiodentals rather than to active constraints on articulatory timing driven by perception. Such a scenario is reminiscent of the assimilation patterns for stops detected in the *perfect memory* example mentioned above (Browman and Goldstein, 1990; Tiede et al., 2001) in which acoustic/perception-production asymmetries can arise when a labial and lingual consonant overlap. We argue for a similar scenario in the case of /f#s, f#ʃ/ sequences: we expect that production data will reveal the sibilant constriction formation to happen during labiodental /f/, yet acoustically, there will be few consequences as long as the labiodental is the narrower constriction and hence the point of biggest pressure drop-off.

There is to our knowledge no existing detailed empirical work on any of the German fricative sequences (for an overall picture of place assimilation in German based on the evaluation of a transcribed corpus, see Zimmerer et al., 2009). Judging from auditory impression, we would expect for /s#ʃ/ to see a similar pattern in German as has been described in English (Nolan et al., 1996; Pouplier et al., 2011): since /s, ʃ/ are produced with very similar articulatory synergies, their temporal overlap should lead to varying degrees of articulatory-acoustic blending of the underlying target articulations. While /s#ʃ/ sequences in English are known to assimilate, /ʃ#s/ shows very little, if any, regressive assimilation and even a slight tendency for progressive assimilation has been reported (Pouplier et al., 2011). In our work on English, we attributed this to asymmetries in coarticulation resistance and articulatory control between the two consonants: if /ʃ/ exerts holistically a greater control over tongue shape, /s/ overlapping a preceding /ʃ/ would have few articulatory consequences. At the same time, studies on French (Niebuhr et al., 2011) and Catalan (Recasens and Mira, 2013) underscore the language-specific nature of sibilant assimilation, since they report a dominance of /ʃ/ for either consonant order. Recasens and Mira (2013) hypothesize that regressive adaptation is exclusive to /s#ʃ/ suggesting that articulatorily /ʃ/ can be anticipated during /s/ but not vice versa. We are not aware of an existing investigation for German, but impressionistically we expect it to pattern with English also in the case of /ʃ#s/. For /s#f/ and /ʃ#f/ sequences, we should see in the acoustics no or only minor effects of /f/ formation during the sibilant due to fricative aerodynamics being predominantly determined by the narrowest constriction; for articulatory records we employ electropalatography (EPG; details below), and hence have no positive information on /f/. We note though that /f/, like labial stops, may in principle trigger place assimilation in German. For instance, pronunciations of *einfach* or *einverstanden* can be transcribed as [mf] (Kohler, 1990; Zimmerer et al., 2009), although the likelihood of regressive place assimilation is of course known to differ for syllable-final nasals and obstruents.

Our experiment also includes lexical stress of the word-final syllable as a covariate. We use the prosodic manipulation to vary the perceptual and articulatory strength of the word-final fricative. The both gradient and categorical nature of /s#ʃ/ assimilation in English has been debated in the literature for some time (Browman, 1995; Nolan, 1992; Nolan et al., 1996; Pouplier et al., 2011): for English /s#ʃ/ sequences, the articulatory-acoustic properties have repeatedly been shown to lie along a continuum between [sʃ] and [(ʃ)ʃ]. Intermediate pronunciations arising from blended production parameters of /s, ʃ/ have been observed alongside no assimilation at all, or complete dominance of /ʃ/ with no trace of /s/ by the given measure employed. There is no clear picture of when blended productions occur rather than productions consonant with a categorical feature replacement account; speaker-, possibly dialect-specific factors have been suggested (Ellis and Hardcastle, 2002). We investigate here whether stress may bias the productions towards one or the other end of the gradient assimilation continuum. We might see a relatively greater dominance of C2 when either final C1 is part of an unstressed syllable and/or C2 is part of a stressed syllable. While sibilants are generally less variable under prosodic variation than other consonants, previous work nonetheless leads us to expect some observable effects. Cho and McQueen (2005) report little change in gestural magnitude as a function of accent for Dutch fricatives, yet Iskarous et al. (2010) provided evidence that prosody affects vertical (rather than horizontal) tongue position in English sibilants. Silbert and De Jong (2008) found that focus mostly affects length and acoustic power differences in fricatives. There is little work investigating the interaction of stress and assimilation specifically: Recasens and Mira (2013) report that categorical /s#ʃ/ assimilation in Eastern Catalan is independent of phrasal stress. For /ʃ#s/ sequences which are generally implemented with a higher degree of articulatory variability in their data, stress causes a bias towards /ʃ/. The authors attribute this to a greater degree of dorsal raising for /ʃ/ under phrasal stress. In terms of onset obstruent-sonorant clusters, Bombien et al. (2010) found for German that stress decreases articulatory overlap, but their work looks at tautosyllabic clusters, meaning both consonants were part of the stressed syllable in their work.

In sum, we predict a greater dominance of C2 production parameters in the case of unstressed C1 as much as in the case of stressed C2. This part of our data will add to the knowledge of how prosodic factors affect the articulation of fricatives. Another covariate used in our experiment is vowel context which we included to have some degree of contextual generalization in the data. The cluster-abutting vowels are either high for the final and low for the following initial syllable or vice versa (termed i-a vs. a-i conditions, respectively, see the Methods). Conceivably, we will see a stronger adaptation effect in the i-a condition due to the greater articulatory compatibility of the postalveolar sibilant with /i/ compared to /s/.

Throughout the paper the term sibilant without further specification is used to refer to either /s/ or /ʃ/; we differentiate between the two as alveolar and postalveolar where necessary. For instance '/f/#sibilant' comprises both /f#s/ and /f#ʃ/.

Method

We recorded acoustic and EPG data using a Matlab interface to the Articulate Instruments EPG system. The articulatory data were sampled at 200 Hz. Audio data were captured with a Sennheiser MKH40 microphone and recorded on a SonyEx multichannel system at 32,768 Hz, together with a synchronization signal from the EPG hardware. Subjects wore custom-fit palates with 62 electrodes. They were given their palates a couple of weeks before the recordings with instructions for practicing articulating with the palate. A 30-min accommodation phase with the palate in place preceded the practice session. Subjects practiced the stimulus list once wearing their palates by reading the list out loud. After a further adaptation phase of 1 h, we started the actual recording.

Participants

Nine native speakers of German participated, all of them colleagues or students at the Institute of Phonetics. Eight were female, 1 male. They were naive as to the purposes of the experiment except for subject 1, the first author of the paper. Experimental participants came from various dialectal backgrounds. In Standard High German, in word-initial position the alveolar fricative is phonologically voiced, yet it is devoiced in Southern Standard German (for an overview of German dialects, see Russ, 1990). Our speakers were all standard speakers of either variety. Seven participants were from dialectal regions featuring a voiceless initial alveolar fricative (Bavaria, Baden-Württemberg, Hessen) and 2 from a northern region with a voiced initial alveolar fricative (Schleswig-Holstein; subjects D5, D8). Regardless of their dialectal background, in the present recordings all speakers realized the word-initial alveolar sibilant uniformly as voiceless, as we determined by visual inspection of the spectrogram and auditory analysis. Therefore we will in this paper not differentiate in transcription between /s/ and /z/ and instead use /s/ to denote the alveolar sibilant in all positions.

Stimuli

Stimuli consisted of noun-noun compound phrases with abutting medial fricatives; the compounds were embedded in the carrier sentence '*[Lotte|Anna|Otto|Peter] hat _______ gehört*', that is '[Lotte|Anna|Otto|Peter] heard ______'. The variation in the sentence-initial proper name was introduced to avoid monotony. The stimuli combined the fricatives /f, s, ʃ/ in word-final and initial position, rendering 3 *cluster conditions:* /f#s, s#f/; /f#ʃ, ʃ#f/; /s#ʃ, ʃ#s/. Homorganic combinations served as controls (/f#f, s#s, ʃ#ʃ/). Both nouns forming the compound were lexical items or names and disyllabic; we will refer to the syllables containing the abutting fricatives as *first and second target syllables*. In the following, we will refer to any analyses involving combinations of /s, ʃ/ in either order as *sibilant conditions* (/s#ʃ, ʃ#s/), and combinations of sibilant and /f/ in either order as */f/ conditions* (/s#f, f#s/; /ʃ#f, f#ʃ/).

Two different vowel contexts were included: in the i-a condition, the first target syllable contained a high front vowel, the second target syllable a low vowel (e.g. ['da.t**i**f#'ʃ**a**.lə]), and vice versa in the a-i condition (e.g. ['ku.k**ɐ**f#'ʃ**ɪ**m.məl]). Due to lexical constraints, there is some variability as to the identity of the low back/high front vowels. Our experiment further included a stress condition, depending on whether the target syllables were lexically stressed or unstressed, rendering 4 combinations of stressed/unstressed target syllables. Due to these covariates, the final stimulus set consisted of semantically nonsensical noun-noun compounds (even though each individual noun is a lexical item or a name). The motivation to use compounds was that compounding is a common and highly productive process in German (Wiese, 1996) which can be used in everyday language to create ad hoc novel words. The stimuli are given in table 1.

We recorded 5 repetitions per target phrase in pseudorandomized blocks. Blocks were pseudorandomized differently for each repetition. Pseudorandomization was employed to prevent immediately adjacent stress condition differences for a given vowel-cluster combination. This means that it was never just stress that varied between one trial and the next. Our token total amounted to 9 consonant sequences × 2 vowel × 4 stress conditions × 5 repetitions × 9 subjects = 3,240 tokens. Due to a coding error, speaker 1 had no data for the control condition /ʃ#ʃ/, i-a, both syllables stressed, with the consequence that experimental conditions associated with these controls had to be excluded for this one speaker (i-a, stressed-stressed, /ʃ#f, f#ʃ, s#ʃ, ʃ#s/). Across subjects and conditions another 9 tokens were missing due to technical failure, leaving a total of 3,206 tokens for analysis.

In our figures throughout the paper, we use upper-case <S> as symbol for /ʃ/.

Data Treatment: Similarity Index Calculation

The intervocalic fricative interval was manually segmented based on the acoustic signal; the fricative interval was defined as the time from vowel offset to vowel onset of the VC#CV sequence. Vowel offset and onset were identified on the basis of visual inspection of the waveform and spectrogram (see Appendix fig. 10 for an example). Subsequent to segmentation, all trials were mapped onto a time interval of [0, 1]. We normalized our acoustic and EPG data with the same procedure used in Pouplier et al. (2011) which was in turn inspired by Gusik and Harrington (2007). The method allows us to compute a similarity index that quantifies in a single number how similar a given articulation/acoustic pattern is to either one of the two homorganic control conditions, for example where along a continuum between control /s#s/ and /ʃ#ʃ/ a given spectral slice or palatogram from the /s#ʃ/ condition falls. We illustrate the procedure for the sibilant condition; it was performed analogously for the /f/ conditions. We emphasize that all analyses of the EPG and acoustic signals are based on the identical acoustic segmentation, i.e. the segmented time interval is the same independently of which signal is analyzed in any particular case.

Electropalatography

Each palatogram was normalized to a scale ranging from –1 to 1, with –1 being the most extreme value for a typical /ʃ/ and 1 being the most extreme value for a typical /s/. The procedure was carried out on a by-speaker, by-condition basis as follows: for each speaker and condition we calculated the across-repetition average contact per electrode at the acoustic midpoint of the control conditions /ʃ#ʃ/ and /s#s/, giving us a reference contact pattern for the alveolar and postalveolar sibilants for a given speaker and vowel/stress condition. Example reference patterns are given in figure 8 of the Appendix. A difference pattern was then created by subtracting for each electrode the value of the /ʃ/ pattern from the value of the

Table 1. Stimuli by condition

V	Stress	C1	Stimulus	C2					
				stressed			unstressed		
				#f-	#s-	#ʃ-	#f-[1]	#s-	#ʃ-
i-a	stressed	-f#	Tari[f]	[f]alle	[s]albe	[ʃ]ale	[f]erleih	[s]alat	[ʃ]almei
		-s#	Gebi[s]						
		-ʃ#	Gemi[ʃ]						
	unstr.	-f#	Dati[f] [2]						
		-s#	Kürbi[s]						
		-ʃ#	Garmi[ʃ]						
a-i	stressed	-f#	Okta[f][2]	[f]immel	[s]impel	[ʃ]immel	[f]ilet	[s]ymbol	[ʃ]ilette[3]
		-s#	Kolo[s]						
		-s#	Goua[s]						
	unstr.	-f#	Kuhka[f]						
		-s#	Calla[s]						
		-ʃ#	Gula[ʃ]						

[1] Note that German orthographic <v> maps phonologically onto /f/ as in *Verleih* or /v/ as in *Vase*, depending on the word. No /v/-initial words were employed in this experiment.
[2] Due to German featuring word final obstruent devoicing, these words with underlying final /v/ are pronounced as [f]. German final fricative devoicing has been shown to be a complete neutralization (in contrast to stops; Piroth and Janker, 2004).
[3] This is a well-known company name and can variably be pronounced with a German native initial [ʃ], or a voiced initial [ʒ] which appears in loan words only. Since the voicing realization may be variable within speaker, we do not consider voicing in our analyses.

/s/ pattern. For electrodes that display a high contact value for /s/ but not for /ʃ/, a large positive number resulted, and a large negative number in cases in which the contact for /ʃ/ substantially exceeded the one for /s/. Contacts that are very similar or the same had a small positive/negative number or zero. This difference pattern was then normalized by dividing each value in the pattern by the sum of the absolute values of the pattern (see Appendix fig. 9). The difference pattern then served to weight the electrodes for each EPG sample. For each EPG sample, each contacted electrode was assigned its value in the difference pattern, each uncontacted electrode contributed the negative of its value to the difference pattern. These positive and negative values of the given EPG sample were then summed over the whole palate. For each EPG sample, the result is a similarity index value between –1 and 1 with –1 being maximally ʃ-like and 1 being maximally s-like. For an /s#ʃ/ cluster, we therefore expect an index dynamic from 1 to –1 over the course of the fricative interval; for an /ʃ#s/ cluster we expect a dynamic from –1 to 1. For the other cluster conditions, the index was calculated correspondingly. The ideal reference values for each of the conditions are as follows: /s#s/ +1 versus /ʃ#ʃ/ –1 for all /s, ʃ/ combinations; /f#f/ +1 versus /ʃ#ʃ/ –1 for all /f, ʃ/ combinations; /s#s/ +1 versus /f#f/ –1 for all /s, f/ combinations.

For the sibilant condition, an EPG index value of zero can theoretically mean two things: an articulation intermediate between /s/ and /ʃ/ or a palatogram with no contacts at all. Our index calculation for the sibilant condition cannot distinguish between these cases; thus, we ascertained that this did not arise. The minimum number of contacts we had across all tokens, all subjects, and all sibilant combinations (/s#s, s#ʃ, ʃ#s, ʃ#ʃ/) in a given sample was 10 contacts. Furthermore, sibilant turbulence cannot be generated without any tongue-palate contact, since bracing is required to achieve appropriate air channeling. A failure to produce turbulence would have been caught as a pronunciation error during acoustic segmentation, but this case did not occur. For the /f/ conditions, palatograms with no contact are to be expected and can indeed be found, but these render an index value very close to the reference pattern for /f#f/ (see Appendix fig. 8).

Acoustics

The audio data were downsampled to 24,576 Hz. We calculated Thomson multitaper spectra for each cluster (Thomson, 1982) with a window length of 20.8 ms (512 points at 24,576 Hz) being moved across the entire utterance with 75% overlap. The quantification of the spectral differences between the sibilants proceeded in analogy to the calculation of the normalized EPG index (each EPG frame can be considered as a vector of 62 values, just as each spectrum consists of a vector of e.g. 256 values), but differed in the arithmetic details because for spectra it is not possible to exploit the restriction of EPG patterns to values of 0 and 1. First, reference patterns for each fricative were computed on the basis of spectra taken at the temporal midpoint of the homorganic control conditions on a by-speaker, by-condition basis. For the following index calculation, the frequency region from 1 to 11 kHz was used. For each sample the Euclidean distance of a given spectral slice from each of the two reference control patterns was calculated. To reduce the influence of overall changes in the signal level, the average value over the given frequency region was subtracted from the measured and reference spectra before computing these Euclidean distances. The final index value was then computed by dividing the Euclidean distance from the /ʃ/ reference by the sum of the two Euclidean distances. This division operation renders values between 0 and 1. For comparability with the EPG index, these values were mapped onto a range from –1 (here: most /ʃ/-like) to 1 (here: most /s/-like) by subtracting 0.5 and multiplying by 2. The cluster condition-specific reference values are the same as for the EPG index: /s#s/ +1 versus /ʃ#ʃ/ –1 for all /s, ʃ/ combinations; /f#f/ +1 versus /ʃ#ʃ/ –1 for all /f, ʃ/ combinations; /s#s/ +1 versus /f#f/ –1 for all /s, f/ combinations.

The similarity index calculation assumes that the resulting data are, for the control conditions, symmetrical around the zero line between –1 and 1. We noticed a difference in how well the homorganic control conditions approached the ideal reference pattern, i.e. in how close the controls came to the values of –1 and 1, respectively. In particular /s/ had a greater distance from the ideal reference pattern than /ʃ/. We corrected for this for all conditions by calculating the index symmetry point for each sample of the controls, allowing us to interpolate an empirical symmetry line. This empirical symmetry line was then subtracted from all data points. This may lead to index values slightly outside the –1 to 1 range. The correction was performed for both indices independently on a by-subject, by-condition basis.

Figure 10 in the Appendix gives oscillogram, spectrogram, segmentation lines, and corresponding EPG and acoustic indices for an illustrative example from one speaker.

Results

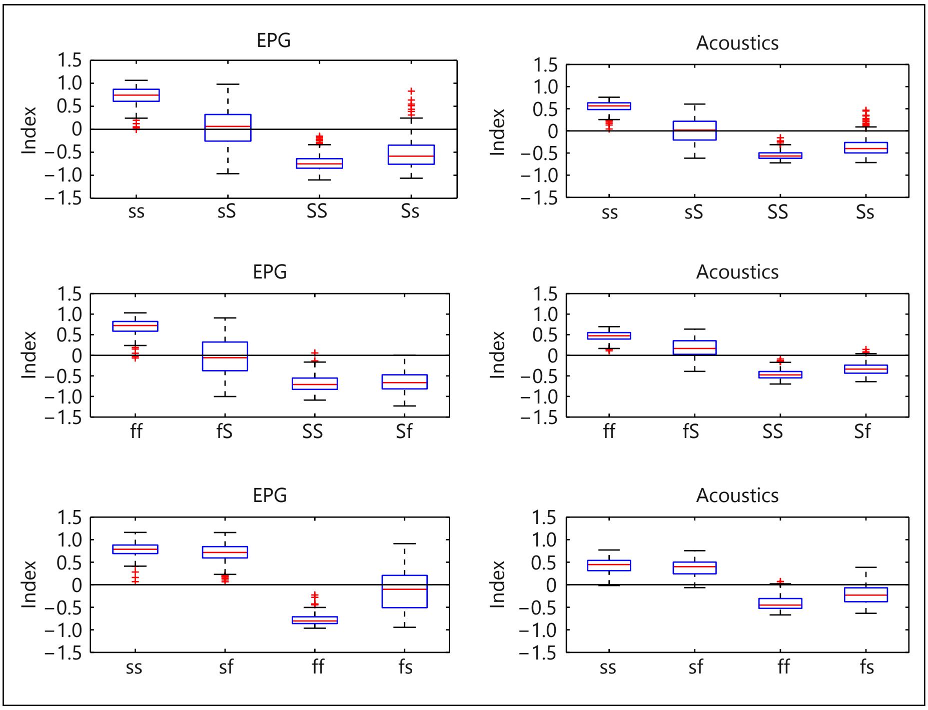
We analyzed the index values at the 25% time point of the normalized cluster interval, reasoning that this time point should approximate the temporal midpoint of C1. For our previous work on English (Pouplier et al., 2011), this time point successfully captured various degrees of blending in /s#ʃ/ productions. The main aspect for our current predictions is that we should observe, at this analysis time point, some evidence for blending for the /s#ʃ/ sequences in both the acoustic and articulatory domain, whereas for the /f#s, ʃ/ sequences, we should find positive evidence for the sibilant in the EPG but not in the acoustic signal. The choice of this analysis time point is of course only a working criterion, but it is important to keep in mind that we evaluate relative differences between conditions to which identical measurement procedures have been employed. The main emphasis of our analyses is on the 25% time point, but we will also refer to other time points (5%, 75%) as appropriate, and show example index trajectories over the entire interval below (fig. 4, 7; see also particularly fig. 10 of the Appendix). The first part of our Results section will be devoted to identifying global differences between the 3 cluster conditions as well as order effects; the second part will be concentrated on the role of stress and vowel context.

Cluster Condition and Order Effects

In our similarity index calculation, assimilation (defined as evidence of the gradient or complete presence of C2 production parameters at the 25% time point, by working hypothesis during underlying C1) will be manifest in the distance from the ideal reference value. To the extent that production parameters of both consonants are present in an overlapping fashion, the index values of the heterorganic conditions should deviate from 1 (or –1, depending on condition). Figure 1 gives box plots for the index values for the controls and heterorganic conditions by cluster for EPG and acoustics, collapsed across all speakers, vowel, and stress conditions. Note that the ideal reference value for /f#f/ is +1 in the /f#f/ versus /ʃ#ʃ/ condition, but –1 in the /f#f/ versus /s#s/ condition. This stems from the similarity index being calculated on a by-condition basis (see Methods).

For the EPG data (left column of fig. 1), particularly the index values for /s#ʃ, f#s/ and /f#ʃ/, assimilation is evident both in terms of median location and spread of the data (whiskers). The median of the EPG pattern for /s#ʃ/ is around zero, i.e. it falls midway between /s/-typical and /ʃ/-typical patterns. This provides evidence that production parameters of both consonants are blended due to their spatiotemporal overlap. The spread of the whiskers underscores that there is considerable variation in the similarity index with values occurring along the entire /s-ʃ/ similarity continuum, including cases in which the similarity index for /s/ in /s#ʃ/ is fully in the range of control /ʃ#ʃ/. This corresponds to previous reports on alveolar sibilant assimilation for other languages, mainly English, which report a coexistence of graded, blended productions and, by the respective measurements, fully assimilated cases (Niebuhr et al., 2011; Pouplier et al., 2011; Recasens and Mira, 2013). For the /f/#sibilant conditions, the EPG index values show a similar spread to that of the /s#ʃ/ condition, giving evidence for the constriction formation for C2 occurring during C1. For the acoustic index, the medians for /f/#sibilant conditions are closer to their respective controls. While the acoustic /s#ʃ/ median is approximately zero (fig. 1, second box of right column, top row graph), the /f#ʃ/ median is closer to reference +1 (first and second boxes of right column, middle row graph) and the /f#s/ median is closer to reference –1 (third and fourth boxes of right column bottom row graph; recall that /f/ has a reference value of +1 for the /f#f/ vs. /ʃ#ʃ/ conditions, but a reference value of –1 for the /f#f/ vs. /s#s/ conditions). There is an order effect in that /s#ʃ/, /f#s/ and /f#ʃ/ have a greater distance from their reference index value compared to /ʃ#s/, /s#f/ and /ʃ#f/. Also for /ʃ#s/, there is a visible difference to the control /ʃ#ʃ/ median in both EPG and acoustics, and the spread of the data as evident in the whiskers and outliers gives evidence of blended production parameters as to be expected from temporally overlapping articulations. For /ʃ#f/ and /s#f/, the EPG index shows little difference to the /ʃ#ʃ/ and /s#s/ controls, respectively. On the one hand, this is to be expected, since there is no positive information on /f/ constriction formation in EPG. On the other hand, our similarity index continuum could provide information on constriction formation of /f/ if temporal overlap were concomitant to a reduced (and thus less control sibilant-similar) lingual articulation. The box plots suggest that there is no reduction of the sibilant articulation; whether there is an overlapping constriction formation of the labiodental can only be assessed from the acoustics. The acoustic similarity index suggests in turn, at least for /ʃ#f/, that there is indeed some influence of a labiodental constriction on the spectrum (difference in median location between /ʃ#ʃ/ and /ʃ#f/).

In order to quantify our observations, we employ a classification approach. In a second step, we follow up on the difference between acoustics and EPG indices using multivariate analyses.



Color version available online

Fig. 1. Box plots of 25% time point similarity index values for the controls and the heterorganic conditions. Note that the meaning of the +1 and –1 reference values is condition specific: /s/ = +1, /ʃ/ = –1 for the /ʃ#ʃ/ versus /s#s/ conditions, /f/ = +1, /ʃ/ = –1 for the /f#f/ versus /ʃ#ʃ/ conditions, and /s/ = +1, /f/ = –1 for the /f#f/ versus /s#s/ conditions.

Classification

A support vector machine (SVM; Baayen, 2008) was trained across subjects on the extracted parameters of the control data as a closed test. We then tested the heterorganic conditions against the control conditions. The classification is by its nature binary; therefore we will refer to all tokens classified as the ‘other’ category (C2) as assimilated. For example, for the /s#ʃ/ condition, the SVM was trained on /s#s/ and /ʃ#ʃ/ sequences. For a given /s#ʃ/ token, at the 25% time point of the fricative interval, the SVM algorithm should classify the token as /ʃ/ if C2 production characteristics dominate, otherwise the token should be classified as /s/. Of main interest in our present context it is whether we obtain a comparable classification for the EPG and acoustic signals for each cluster. We predict a close alignment in the case of sibilants (covariation in the articulatory and acoustic domains), but a divergence for clusters involving the labiodental (articulatory overlap with small acoustic consequences). The training/classification was performed for each of the 3 cluster conditions (/s#ʃ, ʃ#s/, /s#f, f#s/, /ʃ#f, f#ʃ/) separately, once for the EPG and once for the acoustic data. Since this analysis concerned the overall presence of assimilation for the sibilant and /f/ conditions in interaction with consonant order effects, we collapsed across all vowel and stress conditions.

Table 2 gives the results of the SVM classification for all conditions, for both EPG and acoustic data. We first consider the sibilant condition. For this condition, the table

Table 2. SVM percent classification results for acoustic and EPG data for all 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

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 controls show unsurprisingly excellent classification accuracy for both signal types (table 2, cells A1-2, B1-2). For heterorganic sibilant sequences (cells A3-4, B3-4), the pattern corresponds to our predictions in that the acoustic and EPG percentages are in close agreement (cells A3-B3, A4-B4). This is also in line with what has been reported in other studies for English in terms of there being a strong tendency for regressive assimilation for /s#ʃ/ since only 57 and 58% of the data were classified as /s/. For /ʃ#s/ clusters, there is, consonant with our qualitative descriptions above, some influence of /s/ on the postalveolar sibilant, yet this leads to a category switch only in about 10% of cases (9 and 12% in cells A4 and B4).

We now turn to the /f/#sibilant sequences. For all sequences, table 2 gives the percent classified as /f/. Classification accuracy for the controls for both cluster conditions and signal types can be found in table 2, rows C–F, columns 1–2. For the heterorganic conditions, first consider /f#ʃ/. The acoustic data (cell C3) show some influence of C2 with only 77% of tokens having been classified as /f/, the remaining 23% as /ʃ/. Importantly, there is a marked discrepancy between acoustic and EPG data (cells C3 vs. D3): for the latter, only 47% of tokens were classified as /f/, i.e. 53% were classified as /ʃ/. The results for /f#s/ are similar in this respect, with 47% of tokens in the EPG signal being classified as the C2 category, i.e. /s/ (cell F3). Yet this is true for only 16% of tokens when considering the acoustic signal (cell E3). Overall, the assimilation rate for the /f/#sibilant sequences identified in the EPG data is comparable to the one of /s#ʃ/ (58% classification as /s/, 42% as /ʃ/, cell B3). Yet only the /f/#sibilant, not the sibilant sequences show a marked discrepancy between acoustic and EPG signals. This confirms our hypothesis that while an alveolar or postalveolar sibilant constriction is being formed behind the labiodental constriction, the anterior labiodental constriction dominates the acoustics as long as it forms the point of narrower constriction and hence the largest pressure drop-off. Note though that the dominance on the acoustics is not complete. Both figure 1 and the percentages in table 2 (C3, E3) give witness to some effect on the aerodynamics which we infer are due to two overlapping fricative constrictions.

We finally turn to /s, ʃ#f/. Although qualitatively the box plots in figure 1 show slight effects on the acoustic index mainly for /ʃ#f/, there is no binary category switch; the percentage of tokens classified as /f/ is on the same scale as the control conditions (table 2, rows C–F, column 4). There is close agreement between acoustics and EPG classification results. This is not unexpected since we have no positive information in the EPG signal on /f/ constriction formation. For /f#ʃ, s/ we saw that to some extent the interaction of the consonants was reflected in the acoustic index, even though to a much lesser degree than in the EPG signal. For /ʃ, s#f/ we see hardly any trace of an interaction in the acoustics. In the light of the results for /f#ʃ, s/ this may be taken to mean that there is actually less overlap.

Multivariate Analyses

We now turn to multivariate analyses in order to examine the articulatory-acoustic relationship of these clusters in more detail. Significance for mixed linear models was evaluated through a likelihood ratio test in which the full model was compared to a model without the factor in question. For ease of comparison, we multiplied the index values for these conditions such that they always range from an ideal reference value of 1 for C1 to –1 for C2. For the 25% time point, all conditions therefore have an ideal reference value of around 1 if no overlap occurs; to the extent that C2 characteristics are present at the 25% time point, the index values should be biased towards –1.

It became evident in figure 1 that the acoustic index has an inherently narrower range than the EPG index. This is due to the acoustic signal being much richer than the EPG signal – 256 continuously valued points per spectrum enter the index calculation, while for the EPG signal it is only 62 binary points. For the following analyses which directly compare the EPG and acoustic indices, we corrected for the difference in spread between EPG and acoustics. At the 25% time point, we normalized the index value for each token by subtracting it from the control mean and dividing it by the spread of the controls. The normalized index value obtained expresses as a percentage the distance a given measurement point has from the ideal reference value (+1 at the 25% time point). The closer the value is to zero, the closer the index is to the reference. We performed the normalization for both the EPG and acoustic indices.

Figure 2 compares histograms of the normalized index values for /s#ʃ/, /f#ʃ/, and /f#s/ conditions. The data were partitioned into 20 equally spaced bins. The histograms underpin the results of the SVM classification in terms of the asymmetry in the articulatory-acoustic relationship for /f#ʃ/ and /f#s/, but not for /s#ʃ/. While for the sibilant condition the normalized EPG and acoustics indices are more or less on top of each other, there is some discrepancy between the two indices for /f#s/, and a more pronounced discrepancy for /f#ʃ/. Thereby the EPG data display a greater distance from the controls (larger values) compared to the acoustic data. For /ʃ#s, ʃ#f, s#f/ (right hand column of fig. 2), there is some indication of the acoustics being more sensitive to the presence of overlapping constrictions than EPG.

For statistical testing, we obtained a Δ index value by subtracting the normalized acoustic index from the normalized EPG index on a token-by-token basis, setting acoustic and EPG signals in a direct relationship to each other for each token. Figure

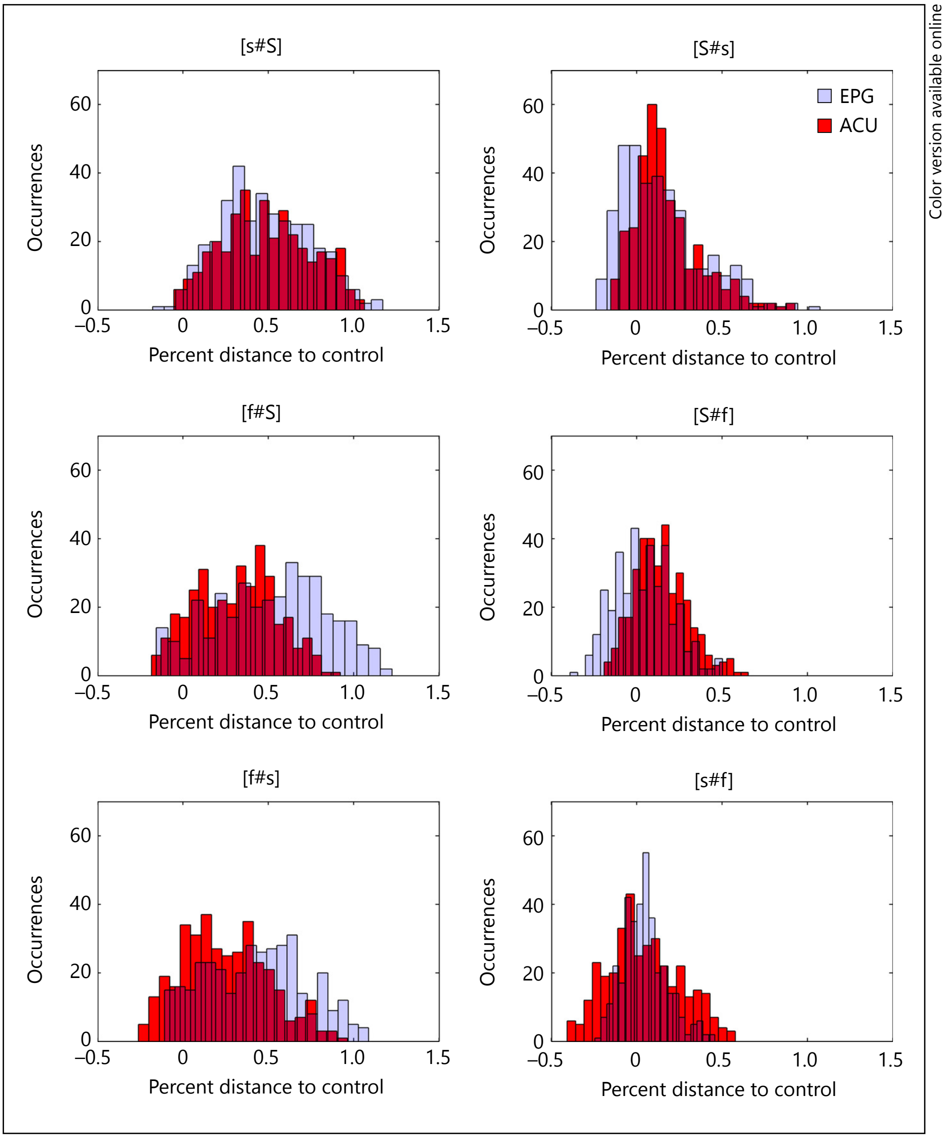


Fig. 2. Histograms of the normalized EPG and acoustic signals. The x-axis shows the percent distance from the control condition, i.e. smaller values mean greater similarity to the control.

3 gives the data as a box plot. Positive values mean that the EPG index has a greater distance to the control condition compared to the acoustic index, negative values mean that the acoustic index deviates to a relatively greater degree from the control. For /s#ʃ, ʃ#s/ we confirm global covariation in the acoustic and EPG signals; the median for /s#ʃ/ falls close to the zero line, although /ʃ#s/ has a slight bias towards negative values. For /f#ʃ, f#s/ EPG values mostly show a greater distance to the controls than the acoustic index (positive median locations). Interestingly, this is the opposite for

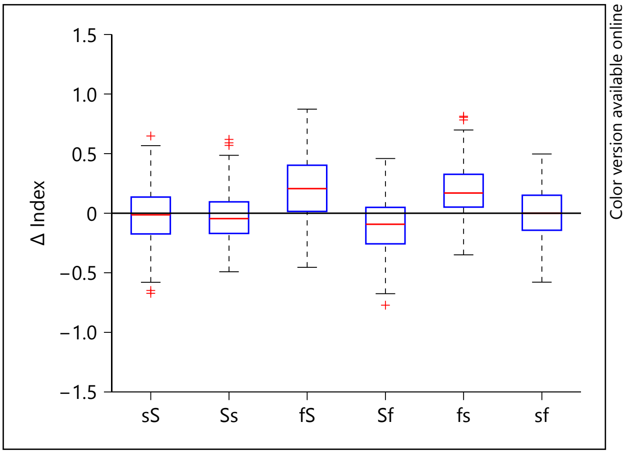


Fig. 3. EPG acoustic Δ index for all clusters.

order /ʃ#f/ (negative median), consistent with our initial qualitative observations that an overlapping C2 /f/ has some effect on the acoustics, yet not enough to cause a category switch in the SVM analysis. For /s#f/, in contrast to /ʃ#f/, EPG and acoustic effects are centered on the zero line. Note though that the median of close to zero here has a different meaning from a near-zero median for /s#ʃ/. While in the latter both acoustic and EPG indexes shift to intermediate production values, for the former there is hardly any shift, neither in the acoustics nor in the EPG signal (fig. 1). The latter fact points once more to the possibility that there is very little overlap in /s#f/ sequences. Alternatively it would be possible that the acoustic effect of an overlapping labiodental constriction is quite different for /ʃ/ compared to /s/, but we cannot test this with the data at hand.

For statistical data quantification of the articulatory-acoustic asymmetry, we ran two separate models, one for cluster group /s#ʃ, f#ʃ, f#s/ and one for cluster group /ʃ#s, ʃ#f, s#f/. The grouping of the clusters was done so as to avoid having 6 levels of the main factor 'cluster'. We first turn to group /s#ʃ, f#ʃ, f#s/. A mixed linear model was run on the Δ index values as dependent variable with fixed factor 'cluster' (/s#ʃ, f#ʃ, f#s/), random intercept and slope for 'speaker' and random intercept for 'word'. Model comparison did not yield a significant effect for random factor 'repetition', therefore it was not included in the final model. A significant main effect was found [$\chi^2(7) = 307$, $p < 0.001$]. Under a Tukey post hoc analysis /f#s/ and /f#ʃ/ did not differ significantly from each other ($p = 0.45$), but both conditions differed significantly at $p < 0.01$ from /s#ʃ/. Overall, these results confirm the patterns revealed by the SVM analysis: there is a discrepancy between EPG and acoustics for the /f#s, ʃ/ conditions, which is significantly less pronounced for the /s#ʃ/ condition. The discrepancy for the /f#s, ʃ/ conditions stems from the acoustic index values having a lesser distance from the reference value than the EPG index values, meaning the overlap evident in the EPG signal has only a small impact on the acoustics. An identical statistical model was run for cluster group /ʃ#s, ʃ#f, s#f/. The main effect cluster was again significant [$\chi^2(7) = 161.2$, $p < 0.001$]. A Tukey post hoc test reported a significant difference for /ʃ#f/ versus /s#f/ ($p = 0.007$) only. In sum, our results show that there is articulatory and concomitantly acoustic blending for /s#ʃ/ and, to a lesser degree, for /ʃ#s/, arising from the temporal overlap of segments calling on the same articulators. At the same analysis time point there is an acoustic-articulatory discrepancy for the /f/#sibilant sequences due to the articulatory overlap having small consequences in the acoustics. There is some evidence from the acoustic signal for articulatory overlap in the case of /ʃ#f/, but not in the case of /s#f/.

The data so far suggest an overall stronger effect for /f#ʃ/ compared to /f#s/. To follow up on this, we show in figure 4 example EPG index time series from one speaker for each of the 3 cluster conditions in which the Δ index value at the 25% time point for at least 1 repetition had a close to 100% distance from the control condition (maximal dominance of C2). For the sibilant condition, we see a pattern familiar from English: for some tokens, the postalveolar sibilant dominates the EPG signal entirely in that there is no longer a pronounced index dynamic from 1 to –1. For /f#ʃ/, this is similar. The index time series is for many tokens more or less on top of the control condition index for /ʃ#ʃ/. For /f#s/, however, for all tokens, the index is initially similar to /f#f/, but very quickly changes to an /s/-typical index value. This difference between /f#s/ and /f#ʃ/ is, by visual inspection, rather systematic. We thus computed, across all speakers, the median EPG index value for each cluster condition at the 5% time point. For /s#ʃ/, the median across subjects and conditions was 0.13, for /f#ʃ/ it was 0.25, and for /f#s/ 0.39. This supports our assumption that there is overall less temporal overlap for /f#s/ than for /f#ʃ/ and /s#ʃ/.

The final part of the Results section now considers stress and vowel context effects.

Vowel and Stress Effects

The goal of the following analyses is to test for the role of vowel context and stress for enhancing or attenuating interactions between abutting fricatives. In order to limit the number of statistical tests, we will concentrate on the orders /s#ʃ, f#ʃ, f#s/ and the EPG index values only, since these are the clusters and the data channel in which the interaction of the fricatives could be observed most clearly. For the stress analyses, we summarized the 4 levels of our stress factor in 2 groups: stress 1 stands for a stressed/unstressed word-final target syllable (C1), stress 2 codes a stressed/unstressed word-initial target syllable (C2). We assessed the effects of stress 1 and stress 2 in separate models.

A mixed linear model on the factors stress 1 and vowel with random factors speaker and word rendered no vowel effect [$\chi^2(1) < 1$, p = 0.67], a marginally significant main effect of stress 1 [$\chi^2(1) = 3.4$, p = 0.06] and a significant interaction [$\chi^2(1) = 13.01$, p < 0.01]. As above, random factor repetition was excluded from the final model due to the lack of an effect in model comparison. The means are given in table 3. The interaction arises due to there being a stress effect for the i-a, but not the a-i condition. For the i-a condition, the unstressed condition displays a greater departure from the reference mean (lower index value).

Based on this result, we looked for an interaction between cluster condition and stress 1 for the i-a condition only (n = 530; fig. 5). The interaction was significant at $\chi^2(2) = 10.8$, p = 0.004. The interaction was due to a weaker stress effect for /f#s/ compared to the clusters with word-initial /#ʃ/ (/s#ʃ, f#ʃ/). This is consistent with our conjecture above that /f#s/ is possibly characterized by a lesser degree of temporal overlap than the other 2 cluster conditions, and hence weaker stress effects would be observed.

We now turn to stress 2. Again we first look for a main effect of stress 2 and a possible interaction with vowel; the means are presented in table 4. There were no significant main effects nor was there a significant interaction (F < 1 for all factors). Also

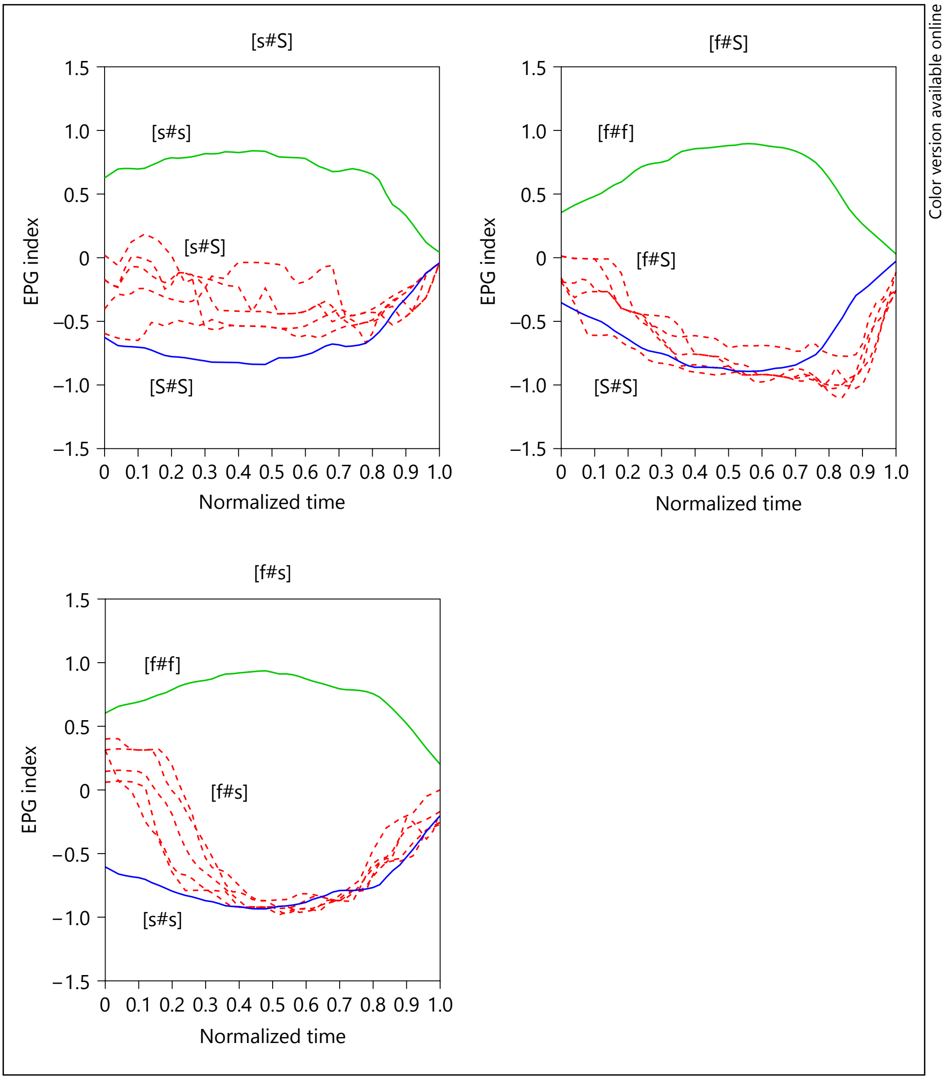


Fig. 4. EPG index dynamics for speaker 6. Shown are example data from the i-a, weak-weak condition. The top and bottom lines show the control average. The reference value for C1 is 1 (top continuous line), for C2 it is –1 (bottom continuous line). The dashed lines are all 5 individual repetitions for the given heterorganic condition.

a mixed model testing for an interaction between stress 2 and cluster rendered F < 1. Due to the low F values, no model comparisons were run.

We also investigated the effects of stress 2 at the 75% time point for the same subset of data, in order to assess the effect of initial lexical stress (fig. 6). The data show neither a main effect of stress 2 (F = 1) nor a significant interaction with cluster (F < 1); the box plots are given in figure 6. Note that the index values are negative because the reference value for C2 is –1.

Table 3. EPG index means by stress 1 and vowel

Vowel context	Stress 1	
	stressed	unstressed
a-i	0.35	0.35
i-a	0.39	0.23

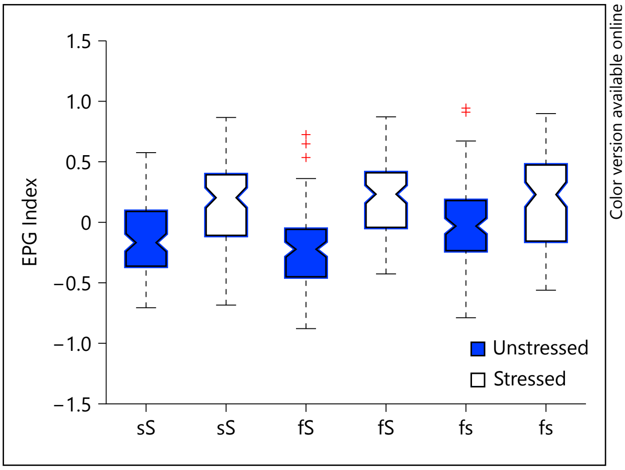


Fig. 5. Stress 1 by cluster interaction; EPG index, vowel condition i-a only.

Overall the analyses support neither a vowel context nor a stress effect of the word-initial target syllable (stress 2). Stress of the word-final target syllable (stress 1) rendered a marginally significant main effect as well as, for the i-a vowel condition, a significant interaction with cluster condition. Overall, stronger effects were seen in unstressed final syllables (stress 1). The stress of the initial target syllable (the trigger of assimilation; stress 2), however, did not condition any differences in index values. The stress effects were specific to the i-a condition.

Discussion

We have presented acoustic and EPG data on German fricative clusters focusing on variation in the articulatory-acoustic relationship as a function of cluster composition. For one, we could show that there is evidence for blending for /s#ʃ/ as well as, in an attenuated fashion, in /ʃ#s/ clusters. In the case of a word-final labiodental fricative (/f#ʃ/, /f#s/), the sibilants seem to overlap temporally with the labiodental constriction similarly to the overlap seen for /s#ʃ/ sequences. For the latter cluster, temporal overlap renders a blended articulatory-acoustic output, confirming the simultaneous presence of the production parameters of both consonants. This could yield intermediate productions along the /s#ʃ/ similarity continuum, or a complete dominance of C2. There was close agreement between the articulatory and acoustic signals; the median of the Δ

Phonetica 2016;73:52–78
DOI: 10.1159/000442590

Table 4. EPG index values for stress 2 by vowel

Vowel context	Stress 1	
	stressed	unstressed
a-i	0.028	0.073
i-a	0.031	0.018

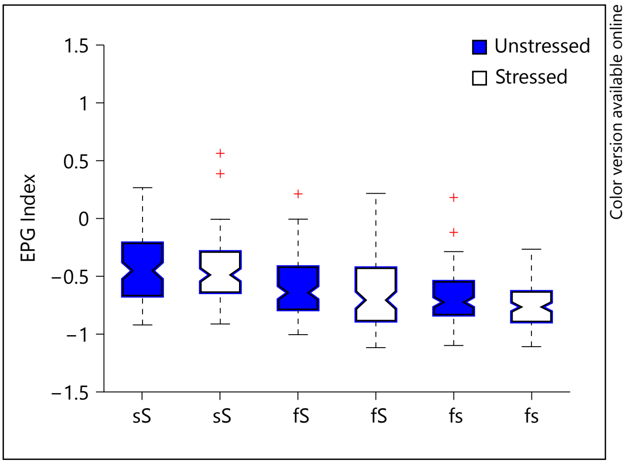


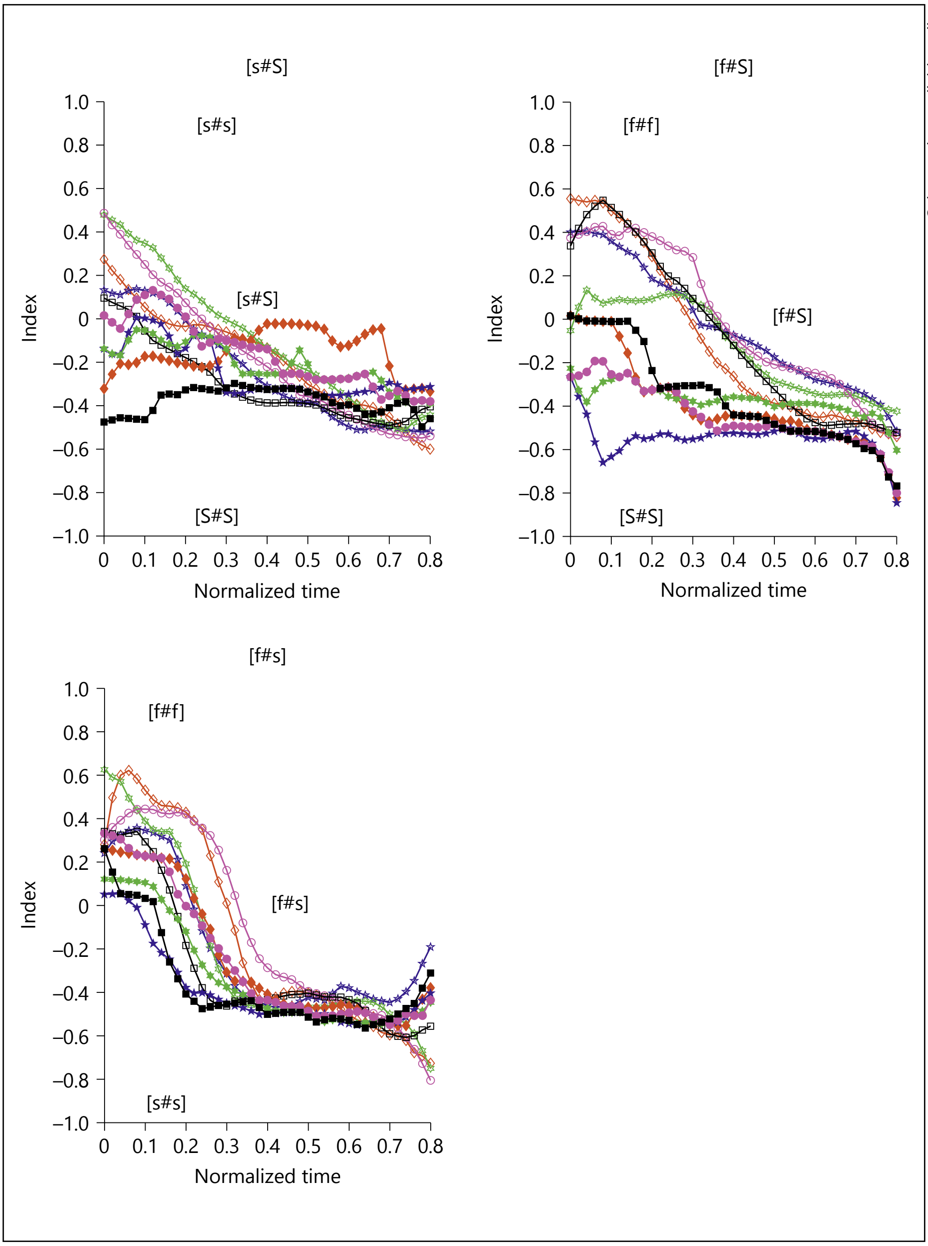
Fig. 6. Stress 2 at the 75% time point by cluster; EPG index, i-a condition only.

index value in figure 3, which quantifies the degree of divergence of the acoustic and EPG signals, was very close to zero. Interestingly, there was a relatively greater deviation from zero for the Δ index value for /ʃ#s/ with the median being slightly negative. This points to a greater sensitivity of the acoustic signal for detecting traces of /ʃ#s/ overlap, and may mean that the tongue-shaping effects of C2 /s/ on preceding /ʃ/ may happen in parts of the tongue not touching the palate or being too posterior for EPG. Conceivably, cross-sectional shape variations in posterior parts of the tongue would not be captured by EPG but could still have some influence on the acoustics even though posterior to the source (Perkell et al., 1979). Carefully controlled modeling studies will have to be employed to follow up on these speculations. 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, albeit mediated by language-specific effects (Niebuhr et al., 2011; Pouplier et al., 2011; Recasens and Mira, 2013). For English /ʃ#s/ sequences, the (near-)lack of regressive assimilation was attributed to /s/ overlapping a preceding /ʃ/ having almost no consequences for the articulation (nor the acoustics) of the postalveolar sibilant due to the latter's tighter, more holistic tongue control. Our results for the sibilant cluster condition are to some degree consistent with this interpretation; the German pattern is similar to the English one, although in our data there was evidence for increasing similarity between C1 and C2 also in the case of /ʃ#s/. We uncovered this effect by setting articulatory and acoustic indices in direct relationship to each other on a token-by-token basis. This was not done in the same way in the other studies cited here, so it may also be that a deeper probing of the articulatory-acoustic relationship would have unraveled more subtle effects in those earlier studies.

For the /f/#sibilant conditions there was, as we predicted, a pronounced asymmetry between articulation and acoustics due to the different constriction locations of the interacting consonants. While the articulatory data revealed overlapping fricative constrictions, the acoustic classification showed a much lesser effect of the sibilant on /f/ in that fewer tokens were classified as assimilated acoustically than articulatorily. Also the multivariate analyses underscored how the relative covariation in articulation and acoustics for the sibilant clusters contrasts with a relative divergence between the two signals in the case of /f/#sibilant clusters. Overall, our data support the assumption that there is a similar pattern of temporal overlap for /s#ʃ/ and /f#ʃ/, to some extent also for /f#s/ sequences, yet due to the different articulators being involved, there are relatively few acoustic consequences of the sibilant constriction formation during the /f/. We propose that this articulatory-acoustic asymmetry may be the reason why /f#s, ʃ/ clusters have been reported to be exempt from regressive place assimilation in German. Due to the articulatory-acoustic covariation in the case of /s#ʃ/, the gradual or complete spatiotemporal overlap of these consonants is clearly audible. In the case of C1 being a labiodental, the ongoing constriction formation of a following sibilant has only a small impact on the acoustics, giving rise to the impression that there is no regressive place assimilation. This is also in line with our SVM classification results which showed that the impact of the sibilant constriction on the acoustics at the 25% time point evident in figure 1 is not enough to cause a category switch/classification as C2, in contrast to /s#ʃ/ where this was the case for over 40% of tokens.

The already cited literature on English /s#ʃ/ assimilation has debated rather vigorously how to interpret the simultaneous presence of both intermediate fricative patterns and instances consistent with a scenario in which the /ʃ/ has replaced /s/ in a complete, possibly categorical fashion. Whichever view one wants to take, the most interesting question in the current context is whether such a complete dominance of C2 is seen in the case of /f#ʃ/, /f#s/. The SVM classification results for the acoustic index showed a discrepancy to the EPG index for these conditions, but there was still a drop in /f/ classification compared to the controls even in these cases (cells C3 and E3 of table 2). In figure 4 we saw for the /f#ʃ/ condition that the EPG index can be very close to, or for some repetitions virtually identical to, the C2 control index throughout the entire interval. Yet the acoustic Δ index values for these repetitions (which quantify the distance to the control in a spread-normalized fashion) are a lot lower, meaning a lesser distance to C1 control. For instance, repetition 2 has an EPG Δ index value of 1, meaning a maximal dominance of C2 (100% distance to the /f/ control value), but the acoustic Δ index is at 0.3, showing that acoustically the labiodental fricative is still strongly present at that time point. To underscore this point once more, and to emphasize that this pattern generalizes beyond our single time point, 'magic moment' analyses, we give in figure 7 the spread-normalized EPG and acoustic indices for the same data we show in figure 4 (without controls). Spread normalization was performed to correct for the acoustic and EPG indices differing in their range, as discussed above. Each index point was divided by the absolute range of the controls at that time point. The EPG (filled symbols) and acoustic (empty symbols) indices for a given repetition can be matched on the basis of symbol type and color. For the moment we can confirm – based on our data – that there are instances of a full presence of C2 in /f#ʃ/ sequences, but acoustically there nonetheless remains substantial, if not dominant evidence for C1. This leads us to conclude that

DOI: 10.1159/000442590



Color version available online

Fig. 7. Spread-normalized EPG and acoustic indices for the same data as presented in figure 4, without the controls. Filled symbols show the EPG index, empty symbols the acoustic index. The indices belonging to the same repetition have the same symbol type and color. The time interval has been cut off at 0.8 for presentation purposes since converging acoustic control indices lead to extremely large spread-normalized values at the edge.

articulatorily, what is called complete regressive place assimilation in the case of /s#ʃ/ may occur in a similar fashion for /f#ʃ/ in that there is near-complete temporal overlap of the two constrictions. What differentiates these cases is the independence of articulators in the latter, but not the former case. For /f#s/, as demonstrated in an exemplary fashion in figures 4 and 7, there seems to be less overlap, and extreme overlap cases do not seem to arise. How exactly overlapping fricative constrictions can play out aerodynamically is anything but straightforward to predict, and modeling work is necessary to make headway on these questions. Also perceptual experiments would add valuable knowledge here as to how sensitive listeners are to the different temporal patterns and different acoustic manifestations of articulatory overlap in these fricative clusters.

Turning to /s#f/ and /ʃ#f/, we likewise found evidence for differences in degree of overlap. There was some effect of the following labiodental constriction on the postalveolar sibilant, but not on the alveolar sibilant. Such a pattern may arise for two reasons: either there is less articulatory overlap, or the aerodynamics of the same degree of articulatory overlap play out differently to condition this effect. Given our qualitative observations and the difference in medians reported in the context of figures 4 and 7, it seems plausible to assume that there is less overlap in case of /s#f/ compared to /ʃ#f/. Likewise, the difference in median index values between /f#s/ and /f#ʃ/ at the 5% time point is suggestive of there being an actual difference in the degree of temporal overlap. Generally, word-initial /#ʃ/ (/s#ʃ, f#ʃ/) displayed stronger effects in terms of a greater C2 dominance compared to word-initial /#s/ (/ʃ#s, f#s/).

At this point we have to ask whether inherent durational differences between consonants might be a factor conditioning our results. For instance, if word-initial /#ʃ/ were systematically longer than word-initial /#s/ (assuming for the moment that cluster duration is the sum of the duration of its members), the 25% analysis time point would capture an earlier absolute time point in the cluster (possibly more C1-like) in the case of word-initial /#s/, but a later time point (possibly during the C1-C2 transition) in case of /#ʃ/, independently of any consonant interactions. Of course, durations vary with context but can, within any given context, not be obtained independently from the consonants' possible interactions. There is to our knowledge very little published work with reference to durational patterns for German fricatives in different syllable positions and contexts. Möbius and van Santen (1996) report mean durations for a single male speaker from the Kiel Corpus, and we will take their data as a lead to gauge the possibility of durational confounds. Onset durations are about equal for all 3 fricatives with /f/ 90 ms, /s/ 91 ms, and /ʃ/ 92 ms. Durations for coda fricatives are /f/ 96 ms, /s/ 116 ms, and /ʃ/ 132 ms. To the extent that the means for their single speaker generalize, these durational patterns cannot underlie the stronger effects we observe for word-initial /#ʃ/ compared to word-initial /#s/. For instance, for /f#s/ and /f#ʃ/ we would, from their results, expect about equal consonant durations for the two cluster members (91#96 ms and 91#92 ms), but it is precisely here that we saw effect differences in our data. For /s#ʃ/ we would get durations of 116#92ms, but despite C1 possibly being longer than C2 in this case, this was the condition in which we saw the expected articulatory-acoustic blending results consonant with previous research. Furthermore, the durational data would lead us to predict a stronger effect for /f#s/ and /f#ʃ/ compared to /s#ʃ/, due to the shorter duration of coda /f/ (96 ms) compared to coda /s/ (116 ms). This again is not the case in our data. For /ʃ#f/, /s#f/ and /ʃ#s/ durational differences may condition some effect strength differences especially for the sibilant clusters: /s#ʃ/ with 116#92 ms, /ʃ#s/ with 132#91 ms. While in both cases C1 is longer than C2, final /s/ is

DOI: 10.1159/000442590

Table 5. Bigram statistics of the Bavarian Archive for Speech Signals

C1	C2				
	[f]	[v]	[s]	[z]	[ʃ]
[v]			0	0.00006	0.00002
[f]			0.00326	0.00277	0.00249
[s]	0.0246	0.11164			0.0102
[z]	0.00047	0.0013	0.00476		0.00038
[ʃ]	0.00181	0.0232	0.00051	0.00149	

shorter than final /ʃ/. Yet the magnitude of these durational differences does not make a strong case for them being the main factor conditioning our results.

Another group of factors which has been known to affect speech production are frequency related, such as phonotactic probability, neighborhood density, as well as word and collocational frequency (Goldrick and Larson, 2008; Jaeger and Hoole, 2011; Scarborough, 2013; Snoeren et al., 2006; Vitevitch, 2002). For our highly controlled, semantically artificial compounds, word frequency cannot be assessed, but we use the published statistics of the Bavarian Archive for Speech Signals to take an initial look at whether phonotactic probability might be an influencing factor. Phoneme bigram statistics of the Bavarian Archive for Speech Signals (Schiel, 2010) are given in table 5.[1] The bigram statistic includes any sequence of C1C2, irrespective of an intervening boundary (syllable, word, phrase, etc.).

There are several reasons why these probabilities have to be interpreted with care for our present purposes. For one, although Schiel refers to these statistics as 'phoneme' based, they are actually based on Munich Automated Segmentation (MAUS) system phonetic transcriptions of spontaneous speech (Schiel, 1999). With respect to voicing this means that the statistic gives phonetic, not canonical voicing status, which is why we give all voicing variants here. Note for instance that phonologically a voiced alveolar sibilant followed by another fricative is phonotactically illegal in German, but due to contextual voicing assimilation this may, according to MAUS results, be encountered in spontaneous speech. Bigram statistics based on phonological, canonical pronunciations are not available.

Note also that MAUS weights the recognition probability of a transcription by several thousand rewrite rules for fluent speech phenomena which are in turn associated with a probability of occurrence (Schiel, 1999). This means regressive place assimilation may also be part of a given MAUS phonetic transcription and enter the phoneme bigram statistic. Because of these issues the bigram statistics are in detail not entirely suitable for our current purposes, but we can still take an initial look at whether phonotactic probability aligns with our results. We adopt the working criterion that there is a substantial difference in bigram probability if differences are of an order of magnitude.

The dissimilarity in the degree of articulatory blending in the two sibilant clusters /s#ʃ/ and /ʃ#s/ seems to be directly reflected in bigram probability: 0.0102 for

[1] These and other statistics are available online at http://www.bas.uni-muenchen.de/forschung/Bas/BasPHONSTATeng.html.

/s#ʃ/ versus 0.00051 for /ʃ#s/ or 0.00149 for /ʃ#z/, respectively. This would point to a lesser articulatory integration between consonants with a lesser phonotactic probability. However, for the other clusters the bigram statistics are more ambiguous. /s#v/ has the highest co-occurrence probability (0.111) of all currently relevant combinations. /s#f/ also has a bigram probability on an order of magnitude greater than /f#ʃ, ʃ#f, ʃ#v/, even though not as high as /s#v/. Hence overall /s#f, v/ is minimally on the same order of magnitude as /s#ʃ/, but we see the weakest effects for the former and the strongest effects for the latter. Overall, phonotactic probability (at least in the form available to us) does not help interpret why we would see a stronger effect for clusters involving the postalveolar rather than the alveolar sibilant as C2 nor as to why we would see less overlap in /s#f/ than /ʃ#f/ in our study. Of course our work was not designed to test for these effects and was based on carefully controlled laboratory speech, employing semantically nonsensical compounds. It may thus be the case that frequency and probability effects may arise in other circumstances (see such effects in French voicing assimilation, in Snoeren et al., 2006). Future research will have to address this issue.

As to stress and vowel effects, only the i-a context was significantly impacted by the prosodic conditions. There was no main effect of vowel. Stone et al. (1992) traced vowel coarticulation effects during sibilants, but this was mainly evident in the cross-sectional shape of the tongue (depth of the groove) which can only be captured by EPG to the extent that this correlates with lateral bracing differences. That stress 1 effects were limited to the i-a context may nonetheless be related to /i/ being known among the vowels to exert the greatest coarticulatory effects on surrounding consonants (Recasens and Espinosa, 2009). It is only in combination that the contact-enhancing/attenuating effects of stress and /i/ became apparent in our data. The stress of the word-final target syllable (stress 1) played a role in assimilation in that more assimilation occurred in the unaccented condition. The strength of the word-initial target syllable (stress 2), however, had no statistically significant effects.

Overall, our results are compatible with the hypothesis that the temporal organization of fricative clusters is globally independent of cluster type, with differences between clusters appearing mainly in degree. Articulatory overlap may be obscured acoustically by a labiodental constriction, similarly to what has been reported for stops. We hypothesize that previously reported asymmetries in regressive place assimilation among the fricative clusters can by and large be predicted by taking the asymmetries in the articulatory-acoustic relationship into account, in particular when labiodentals are involved. Future modeling work is called for in order to gain a better understanding of the aerodynamic consequences of overlapping fricative constrictions.

Acknowledgments

This work was supported by the ERC under the European Union's 7th Framework Programme (FP/2007-2013)/Grant Agreement No. 283349-SCSPL and DFG grant PO 1269/1-1 to M. Pouplier. We are indebted to the speakers who took part in our experiment, to Susanne Waltl for help with acoustic segmentation and to Manfred Pastätter for running the experiments and helping with data analysis. We also thank two anonymous reviewers and the Editor for their comments and suggestions.

Appendix

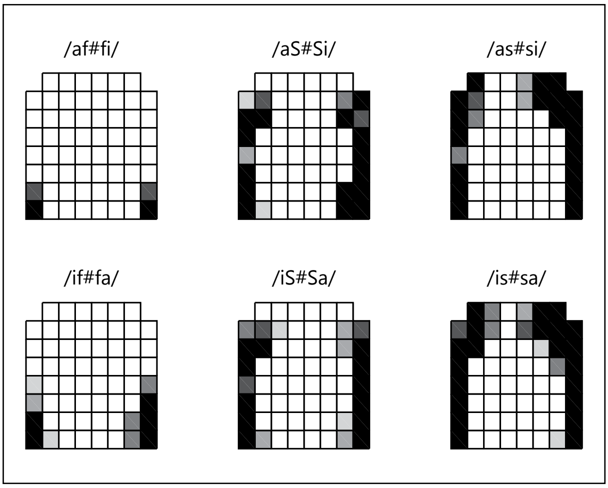


Fig. 8. Palatograms for speaker 9. Shown are the reference patterns for the 3 homorganic control conditions for 2 vowel contexts, unstressed-stressed condition. The reference patterns served as the basis for the by-speaker, by-condition index calculation; see Methods for details. Shading is equivalent to percent average contact over repetitions with black being 100% and white being 0. The anterior is oriented towards the top in each graph.

	100	0	0	40	100	100	
80	0	0	0	40	100	40	0
0	–40	0	0	0	0	0	20
0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	–100	0
0	–20	0	0	0	0	–100	0

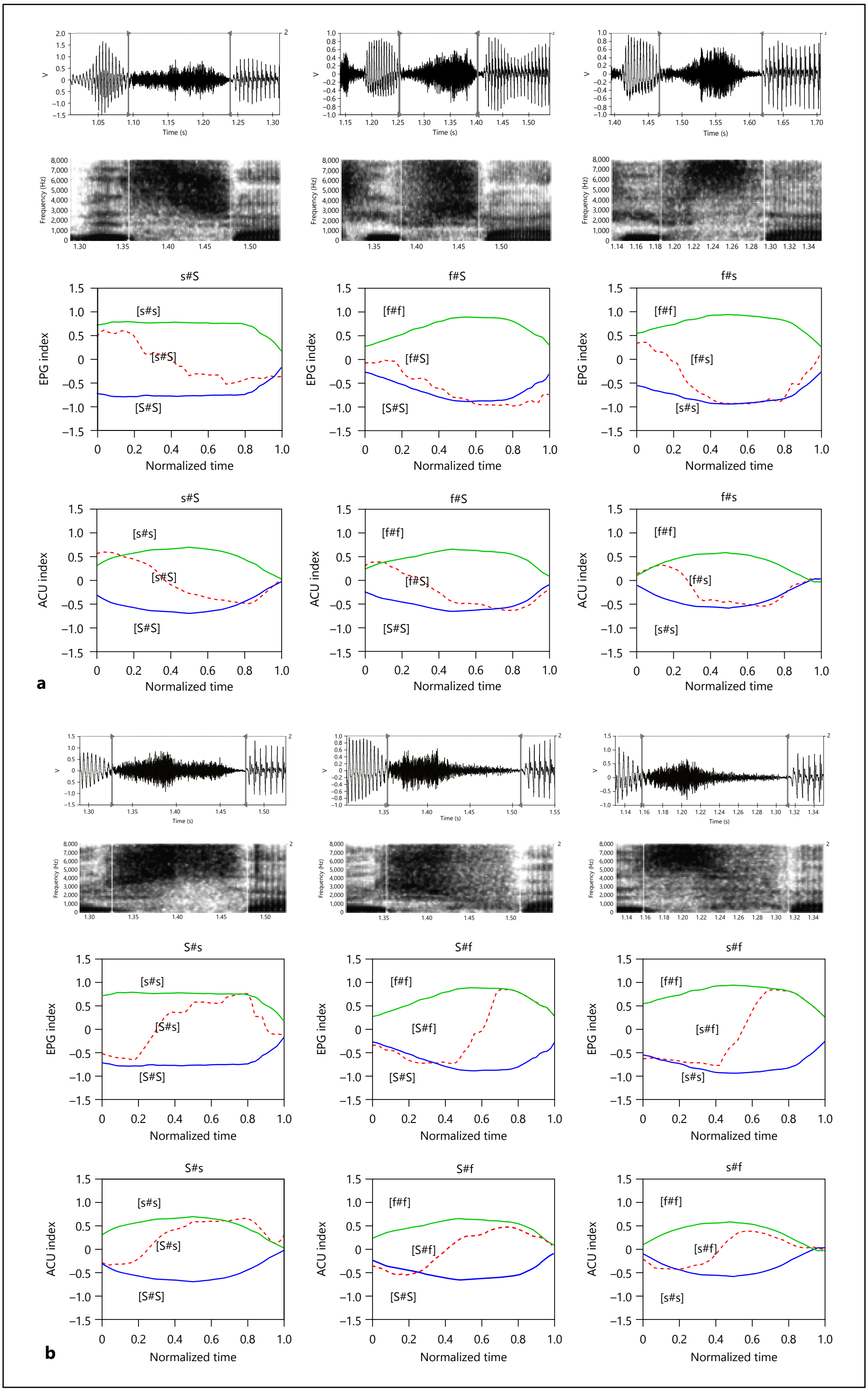
a

	0.11	0	0	0.04	0.11	0.11	
0.09	0	0	0	0.04	0.11	0.04	0
0	–0.04	0	0	0	0	0	0.02
0	0	0	0	0	0	0	0
0.02	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	–0.11	0
0	–0.02	0	0	0	0	–0.11	0

b

Fig. 9. Illustration of the difference pattern by which each EPG sample was weighted. The difference pattern was created by subtracting – on a by-condition, by-subject, by-electrode basis – the relevant reference conditions from each other (here: /s#s/ – /ʃ#ʃ/, i-a, unstressed-stressed, speaker 9, i.e. the bottom row mid and rightmost palatograms of fig. 8) and dividing each electrode's difference value by the sum of the absolute values of all the electrodes in the raw difference pattern; see Methods for more detail. **a** Raw difference pattern (/ss/ – /ʃʃ/). **b** Normalized difference pattern [(/ss – ʃʃ/)/Σ abs(difference pattern)].

Fig. 10. Waveform, sonogram, EPG, and acoustic (ACU) index of speaker 6 (vowel condition: i-a, stress 1: unstressed, stress 2: stressed). The temporal interval over which the EPG and acoustic indices were computed is indicated by vertical lines in the waveform and spectrogram. In each index plot, the top and bottom continuous lines represent the index of the averaged homorganic controls. The dashed line corresponds to the index of the single repetition for which also waveform and spectrogram are displayed (repetition 4 for all conditions). **a** /s#ʃ, f#ʃ, f#s/. **b** /ʃ#s, ʃ#f, s#f/. (*For figure see next page.*)



Color version available online

10

References

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

Babel M, McGuire G (2013): Listener expectations and gender bias in nonsibilant fricative perception. Phonetica 70:117–151.

Bombien L, Mooshammer C, Hoole P (2013): Articulatory coordination in word-initial clusters of German. J Phonet 41:546–561.

Bombien L, Mooshammer C, Hoole P, Kühnert B (2010): Prosodic segmental effects on EPG contact patterns of word-initial German clusters. J Phonet 38:388–403.

Browman C (1995): Assimilation as gestural overlap: comments on Holst and Nolan; in Connell B, Arvaniti A (eds): Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV. Cambridge, Cambridge University Press, pp 334–342.

Browman C, Goldstein L (1990): Tiers in articulatory phonology, with some implications for casual speech; in Kingston J, Beckman ME (eds): Papers in Laboratory Phonology. I. Between the Grammar and Physics of Speech. Cambridge, Cambridge University Press, pp 340–376.

Byrd D (1992): Perception of assimilation in consonant clusters: a gestural model. Phonetica 49:1–24.

Cho T, McQueen JM (2005): Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress. J Phonet 33:121–157.

Ellis L, Hardcastle W (2002): Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. J Phonet 30:373–396.

Farnetani E, Recasens D (2010): Coarticulation and connected speech processes; in Hardcastle J, Laver FE, Gibbon WJ (eds): The Handbook of Phonetic Sciences, ed 2. Hoboken, Wiley-Blackwell, pp 316–352.

Goldrick M, Larson M (2008): Phonotactic probability influences speech production. Cognition 107:1155–1164.

Gusik KM, Harrington J (2007): The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI). Adv Speech Lang Pathol 9:109–119.

Harrington J, Hoole P, Pouplier M (2013): Future directions in speech production; in Knight R-A, Jones M (eds): The Bloomsbury Companion to Phonetics. London, Bloomsbury, pp 242–259.

Hura SL, Lindblom B, Diehl RL (1992): On the role of perception in shaping phonological assimilation rules. Lang Speech 35:59–72.

Iskarous K, McDonough J, Whalen D (2012): A gestural account of the velar fricative in Navajo. Lab Phonol 3:195–210.

Iskarous K, Pouplier M, Marin S, Harrington J (2010): The interaction between prosodic boundaries and accent in the production of sibilants. Proceedings of the 5th International Conference on Speech Prosody, Chicago.

Jaeger M, Hoole P (2011): Articulatory factors influencing regressive place assimilation across word-boundaries in German. J Phonet 39:413–428.

Jun J (1995): Perceptual and Articulatory Factors in Place Assimilation: An Optimality Theoretic Approach; PhD dissertation, UCLA, Los Angeles.

Jun J (2004): Place assimilation; in Hayes B, Kirchner R, Steriade D (eds): Phonetically Based Phonology. Cambridge, Cambridge University Press, pp 58–86.

Kohler K (1990): Segmental reduction in connected speech in German: phonological facts and phonetic explanations; in Hardcastle WJ, Marchal A (eds): Speech Production and Speech Modelling. Dordrecht, Kluwer, pp 69–92.

Miller GA, Nicely PE (1955): An analysis of perceptual confusions among some English consonants. J Acoust Soc Am 27:338–352.

Möbius B, van Santen J (1996): Modeling segmental duration in German text-to-speech synthesis. Proceedings of the International Conference on Spoken Language Processing, Philadelphia, vol 4, pp 2395–2398.

Niebuhr O, Clayards M, Meunier C, Lancia L (2011): On place assimilation in sibilant sequences – comparing French and English. J Phonet 39:429–451.

Nolan F (1992): The descriptive role of segments: evidence from assimilation; in Docherty GJ, Ladd DR (eds): Laboratory Phonology. II. Gesture, Segment, Prosody. Cambridge, Cambridge University Press, pp 261–280.

Nolan F, Holst T, Kühnert B (1996): Modelling [s] to [sh] accomodation in English. J Phonet 24:113–137.

Ohala J (1993): Coarticulation and phonology. Lang Speech 36:155–170.

Perkell JS, Boyce SE, Stevens K (1979): Articulatory and acoustic correlations of the [s-š] distinction. Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America, Cambridge, pp 109–113.

Piroth HG, Janker PM (2004): Speaker-dependent differences in voicing and devoicing of German obstruents. J Phonet 32:81–109.

Pouplier M, Hoole P, Scobbie J (2011): Investigating the asymmetry of English sibilant assimilation: acoustic and EPG data. J Lab Phonol 2:1–33.

Recasens D (2006): Integrating coarticulation, blending and assimilation into a model of articulatory constraints; in Goldstein L, Whalen D, Best C (eds): Laboratory Phonology 8. Berlin, Mouton de Gruyter, pp 611–634.

Recasens D, Espinosa A (2009): An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. J Acoust Soc Am 125:2288–2298.

Recasens D, Mira M (2013): An articulatory and acoustic study of the fricative clusters /sS/ and /Ss/ in Catalan. Phonetica 70:298–322.
Recasens D, Pallarès MD (2001): Coarticulation, blending and assimilation in Catalan consonant clusters. J Phonet 29:273–301.
Russ CVJ (ed) (1990): The Dialects of Modern German. A Linguistic Survey. London, Routledge.
Scarborough R (2013): Neighborhood-conditioned patterns in phonetic detail: relating coarticulation and hyperarticulation. J Phonet 41:491–508.
Schiel F (1999): Automatic phonetic transcription of non-prompted speech. Proceedings of the International Congress of Phonetic Sciences, San Francisco, pp 607–610.
Schiel F (2010): BAStat: new statistical resources at the Bavarian Archive for Speech Signals. Proceedings of the International Conference on Language Resources and Evaluation 2010, La Valletta, p 277.
Silbert N, De Jong K (2008): Focus, prosodic context, and phonological feature specification: patterns of variation in fricative production. J Acoust Soc Am 123:2769–2779.
Snoeren N, Hallé P, Segui J (2006): A voice for the voiceless: production and perception of assimilated stops in French. J Phonet 34:241–268.
Son M, Kochetov A, Pouplier M (2007): The role of gestural overlap in perceptual place assimilation in Korean; in Cole J, Hualde JI (eds): Papers in Laboratory Phonology IX. Berlin, Mouton de Gruyter, pp 507–534.
Steriade D (1995): Underspecification and markedness; in Goldsmith J (ed): The Handbook of Phonological Theory. Cambridge, Blackwell, pp 114–174.
Steriade D (2001): Directional asymmetries in place assimilation: a perceptual account; in Hume E, Johnson K (eds): The Role of Speech Perception in Phonology. San Diego, Academic Press, pp 219–250.
Stone M, Faber A, Raphael LJ, Shawker T (1992): Cross-sectional tongue shape and linguopalatal contact patterns in [s], [sh], and [l]. J Phonet 20:253–270.
Thomson DJ (1982): Spectrum estimation and harmonic analysis. Proc IEEE 70:1055–1096.
Tiede M, Perkell J, Zandipour M, Matthies M (2001): Gestural timing effects in the 'perfect memory' sequence observed under three rates by electromagnetometry (abstract). J Acoust Soc Am 110:2657.
Vitevitch MS (2002): The influence of phonological similarity neighborhoods on speech production. J Exp Psychol Learning Mem Cogn 28:735–747.
Wiese R (1996): The Phonology of German. Oxford, Clarendon Press.
Zimmerer F, Reetz H, Lahiri A (2009): Place assimilation across words in running speech: corpus analysis and perception. J Acoust Soc Am 125:2307–2322.
Zsiga E (1995): An acoustic and electropalatographic study of lexical and postlexical palatalization in American English; in Connell B, Arvaniti A (eds): Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV. Cambridge, Cambridge University Press, pp 282–302.
Zsiga E (2006): Assimilation; in Brown K (ed): Encyclopedia of Language and Linguistics. Amsterdam, Elsevier, pp 553–558.
Zsiga E (2011): Local assimilation; in Van Oostendorp M, Rice K, Hume E, Ewen CJ (eds): The Blackwell Companion to Phonology. Hoboken, Wiley-Blackwell, pp 1919–1944.