On place assimilation in sibilant sequences—Comparing French and English

Oliver Niebuhr a,⁎, Meghan Clayards b, Christine Meunier a, Leonardo Lancia a

a Laboratoire Parole & Langage, CNRS UMR 6057, Université de Provence, Aix-en-Provence, 5 Avenue Pasteur, 13100 Aix-en-Provence, France
b Department of Psychology, University of York, York YO10 5DD, UK

1. Introduction

1.1. Aim of the study in light of assimilation and co-articulation

While assimilation and co-articulation both extend the phonetic properties of speech segments, the demarcation line between them is traditionally parallel to the demarcation of phonology and phonetics. Assimilation is traditionally anchored in the phonology and is hence a cognitive phenomenon, in which a phonemic feature is substituted by one of a (typically) adjacent segment in a categorical way (cf. McCarthy, 1988; Spencer, 1996). That is, if the feature substitution takes place it leaves no phonetic trace of the original feature. On the other hand, co-articulation traditionally refers to a phonetic phenomenon that results from the biomechanical properties of the articulatory apparatus, leading to gradual variation within phonological categories whose degree changes, for example, with speaking rate (cf. Amerman, Daniloff, & Moll, 1970; Daniloff & Hammarberg, 1973).

Assimilation and co-articulation are still treated as separate phenomena in the literature in some more recent studies (cf. Recasens, 1991; Recasens & Pallarès, 2001). However, a growing number of investigations discard this distinction. For example, Whalen (1990) and Wood (1991, 1996) provide evidence that co-articulation can also be pre-planned and hence rooted in a cognitive level. Moreover, several studies have shown that assimilation can create gradual phonetic variation (cf. Hardcastle, 1994; Nolan, 1992; Wright & Kerswill, 1989), while co-articulation can be categorical (cf. Ambrazaitis & John, 2005).

Furthermore, identical segmental and prosodic contexts can create both categorical and gradual assimilations, depending on the strategies of the individual speakers (Ellis & Hardcastle, 2002; Heuvel, van den, Cranen, & Rietveld, 1996; Kühnert & Hoole, 2004). Consequently, some authors have started to use the two terms synonymously (cf. Clark & Yallop, 1990; Ohala, 1993).

It is not the primary aim of the present study to look for empirical evidence for or against a separate modelling of assimilation and co-articulation, although our findings will have implications for the modelling of assimilation. Instead, the aim of this paper is to describe the acoustic interactions of alveolar and postalveolar sibilants across word boundaries with regards to place of articulation using segmentally controlled read-speech material. We will compare these interactions for French and English in the context of the following factors: the order of the sibilants in the sequence, the surrounding vowel contexts, simultaneous voice assimilations, and...
the individual speakers. Based on previous studies, we expect that for both languages the interactions can cross phonemic boundaries. For example, in a sequence like /sf/ the /s/ could be realized as a [ʃ], which is spectrally indistinguishable from an allophone of [ʃ] in a comparable context. For this reason, we prefer to use the term ‘assimilation’ to refer to the investigated effects. Moreover, the term ‘sequence’ is used to indicate that the two adjacent sibilants are structurally separated by a word boundary.

1.2. Research background

1.2.1. Place assimilation in French sibilant sequences

Both English and French have regressive voice assimilation across word boundaries (cf. Gimson, 1994; Roach, 1983; Snoeren, Hallé, & Segui, 2006; van Dommelen, 1985). However, while English additionally shows assimilation of place of articulation, which affects plosives, fricatives, and nasals (cf. Roach, 1983), this type of assimilation “is thought to be non-existent in French” (Fagyal, Kibbee, & Jenkins, 2006, p. 49; cf. also Gow, 2003; Ramus, 2001). In accord with this predominant view, the literature on French assimilation patterns deals almost exclusively with voice assimilation (cf. Armstrong, 1932; Malmberg, 1969; Price, 1991).

An apparent counter example is provided by Carton (1974), who observed that the /z/ at the end of “quinze” in “quinze juin” (‘June 15th’) can become a [ʒ]. However, this case of regressive alveolar-to-postalveolar place assimilation requires elision of the schwa that is located in between the two sibilants at the end of the first word, i.e. “quinze”.

Word-final /s/ elision is very common in the varieties in and around Paris but less common in Southern varieties (cf. Fagyal et al., 2006; Gadet, 1992; Pustka, 2008), suggesting that if place assimilation occurs after /a/ deletion we might expect it to be less prevalent in Southern varieties.

This expectation is inconsistent with the productions of “je suis” (/ʃwε, ‘I am’) and “je sais” (/ʃæs, ‘I know’) as [ʃ ʌɪ] or [ʃ ɛ], respectively. These reductions include alveolar-to-postalveolar sibilant assimilations after schwa elision, and they are very widespread across all regional varieties of French. Moreover, unlike in the “quinze juin” example, the place assimilation is not regressive, but progressive. There are two possibilities to reconcile all examples. Firstly, the supra-regional progressive place assimilations in “je suis” and “je sais” are exceptional in so far as they are somehow linked with the co-occurring regressive voiced-to-voiceless assimilation between /s/ and /ʃ/. Secondly, one could argue that “je suis” and “je sais” are fixed expressions, not subject to productive assimilation rules. This argument is supported by the orthographic representations <chai> and <cha> that have evolved.

However, the two possibilities are again inconsistent with the general claim of Kohler (2002, p. 19) that “the succession post-alveolar and alveolar fricatives [...] with Schwa elision [...] may result in [...] [ʃʃ]”. We found several examples supporting Kohler’s claim in the French ‘Corpus of Interaction Data’, CID (cf. Bertrand, Blache, & Espesser, 2007). Two of these examples with the same wording, i.e. “Scotch sur la bouche” (‘Scotch tape across the mouth’), are illustrated in Fig. 1(a) and (b). In both cases the alveolar sibilant /s/ at the onset of “sur” was assimilated progressively to the preceding postalveolar sibilant /ʃ/ at the offset of “Scotch”. The assimilations could not be triggered by co-occurring voice assimilations, and they were produced by a speaker of a Southern French variety. However, the examples are again exceptional in the sense that there was not /a/ to elide at the end of “sur”. The two sibilants were already adjacent.

Closer inspections show that (a) and (b) differ in the degree of the assimilation. In (a), the assimilation resulted in a spectrally constant sibilant quality, which is clearly different from the initial alveolar /s/ of “Scotch”, but comparable with the final postalveolar /ʃ/ of “bouche”. This suggests that the /ʃs/ sequence of (a) is realized with complete assimilation, which is supported by the auditory impressions of the authors who listened to the /ʃs/ sequence and the final /ʃʃ/ of “bouche” and found it hard to discriminate between the two sibilant sections. The assimilation in the /ʃʃ/ sequence of example (b) also created a roughly constant sibilant section. However, according to the spectral energy distribution of the noise in the spectrogram and the auditory impressions of the authors, its phonetic quality is distinct from both the /ʃs/ of “Scotch” and the /ʃʃ/ of “bouche”. The intermediate quality of /ʃʃ/ in terms of the spectral energy distribution is additionally illustrated in the three slice sections of Fig. 1(c), which were calculated at the centres of the initial /ʃ/, the /ʃʃ/ sequence, and the final /ʃʃ/. As in the case of /ʃʃ/, the spectral slice of /ʃʃ/ shows an energy maximum in the mid-frequency region of the noise, which is, however, slightly higher than that of /ʃʃ/. At the same time, the energy of the /ʃs/ slice remains at a high level towards the end of the displayed frequency range, which is a characteristic of /ʃ/, but not of /ʃʃ/. So, in the /ʃʃ/ sequence of example (b) the /ʃʃ/ only induced a partial progressive place assimilation of the following /ʃ/.

In summary, given the examples discussed above, it may be supposed that – contrary to the predominant view in the literature – assimilation of place of articulation does exist in French, at least in sequences of alveolar and postalveolar sibilants. Moreover, it may be a gradual process that changes the alveolar sibilant towards post-alveolar, and that also allows for complete assimilations at the
extreme end of the continuum. However, as far as the details of this assimilatory process including its direction, its regional dissemination, and its interaction with simultaneous voice assimilation are concerned, the examples raise more questions than they provide answers. Therefore, it is a major aim of the present study to advance this picture by adding qualitative and quantitative empirical details on the basis of the first systematic and comprehensive acoustic–phonetic analysis on place assimilation in French.

### 1.2.2. Place assimilation in English sibilant sequences

In contrast to French, place assimilation in the sibilant sequences of English has already been addressed in a number of studies that dealt with acoustic and articulatory patterns. For example, Holst and Nolan (1995) presented an acoustic analysis of the /s/j sequences across word boundaries in British English. As in other consonant sequences across languages (cf. Section 1.1), they found a gradual assimilation effect. At the one end of the continuum, there were examples of non-assimilations represented by two individual spectrally stable sibilant sections. These examples were called ‘type A’. At the opposite end, referred to as ‘type D’, there were sibilant sections which were spectrally indistinguishable from /j/ in a comparable environment, and which may thus be regarded as examples of complete regressive assimilations. Additionally, Holst and Nolan report different intermediate forms (subsumed under types B–C). They comprise successive spectral changes from more /s/-like to more /j/-like sibilant qualities and constant frications with spectra in between /s/ and /j/. Parallel to this continuum of different spectral patterns, Holst and Nolan found a duration continuum in which the non-assimilated sibilant sequences had the longest overall durations, whereas the completely assimilated sequences had the shortest overall durations. The spectrally intermediate forms were also marked by intermediate durations (cf. Browman, 1995; Nolan, Holst, & Kühnert, 1996). The EPG-based investigation of Zsiga (1995) revealed similar gradual regressive assimilation patterns within the /s/j/ sequences of American English that fit in with previous studies by Catford (1977) as well as by Zue and Shattuck-Hufnagel (1980). Moreover, the data of Zsiga (1995) points to considerable inter-speaker variation, not only for the realization of the /s/j/ sequences, but also for the production of single /s/ and /j/ sounds in similar environments.

In summary, the picture of place assimilation in sibilant sequences of English is much clearer than in French. However, the role of many contextual factors is still unknown, most importantly, the order of the sibilants in the sequence. That is, what do English speakers do in postalveolar–alveolar sequences like /s/, however, this would mean assimilation towards alveolar, which is scarcely to be expected, since assimilations typically lead away from alveolar articulations. This strong cross-linguistic tendency to give up these complex tongue-tip articulations, which are saddled up on the global, vowel-related tongue-body gestures, gave rise to the claim that assimilatory processes aim at reducing effort in speech production, while maintaining salient acoustic–perceptual properties (cf. Jun, 2004; Kohler, 1990; Steriade, 2001). So it is most likely that place assimilation in English sibilant sequences is an asymmetric phenomenon, i.e. it occurs solely for alveolar–postalveolar sequences and the alveolar sibilant is assimilated regressive.

The present study fleshes out the English picture further by addressing sibilant order, along with a number of other contextual factors such as the surrounding vowel contexts, the frequencies of the words involved, simultaneous voice assimilations, and the speakers themselves. Moreover, comparing the findings for French and English in light of these contextual factors and with reference to the language-specific structures can provide further insights into the process of assimilation in general and where it is located in the act of speech production (cf. Section 1.1). Therefore, we have developed, elicited, and analysed the French and English speech data in a parallel way.

### 1.2.3. A note on voice assimilation

Although place assimilation and voice assimilation have each been thoroughly investigated across languages, possible interactions between these two processes have received little attention. One obvious reason is that they are thought not to co-occur in the same language—at least not for the same groups of sounds or in the same morphosyntactic environments. For example, English is well known for having place assimilation across word boundaries (e.g. “right berries” as “ripe berries” cf. Gow, 2003) while voice assimilation is thought to occur word-internally (e.g. the plural suffixes in “walk-s” vs. “dog-s”, cf. Roach, 1983). Some have argued that voice assimilation can also occur across word boundaries but is restricted to function words (Lass, 1984; Selkirk, 1980) or fricatives in general (Gimson, 1994). Roach (1983) does not point to any restriction, but agrees with all other researchers that regressive voice assimilation in English goes from voiced to voiceless. In French, the situation is reversed. Assimilation occurs across word boundaries for voice (cf. Snoeren et al., 2006; van Dommelen, 1985), but not place. However, the single-case studies described above suggest that place assimilation exists in French sibilant sequences and that sequences like [z j] or [z s] can be the subject of simultaneous place and voice assimilations which may even go in opposite directions.

Thus in the present study we took the opportunity to include voice assimilation as a supplementary issue. The main focus in English is on providing more detailed evidence for voice assimilation. Based on Gimson (1994) and Roach (1983) above, we expected English sibilant sequences to show regressive voiced-to-voiceless assimilation, even across word boundaries and for non-function words. In French we focussed on interactions between voice and place assimilation processes. As far as we are aware, there is no evidence for such interactions in the literature.

### 1.3. Research hypotheses

Based on the findings from the single-case studies that are available for French, the following three hypotheses are put forward:

- (F1) Assimilation of place of articulation does occur within French sibilant sequences.
- (F2) The assimilation is pervasive across a wide range of lexical items, across French varieties, and independently of the phonological voicing features of the sibilants.
- (F3) The assimilation can be progressive as well as regressive, depending on whether the postalveolar sibilant is the first or the second element in the sequence. In both cases, the assimilation is gradual, ranging from non-assimilations over – temporally and spectrally – intermediate forms to complete assimilations.

As for the English sibilant sequences, we hypothesize that

- (E1) Based on Holst and Nolan (1995), place assimilation patterns are exclusively regressive and gradual in the time and frequency domains.
(E2) Assimilation patterns are asymmetrical. That is, alveolar sibilants can be assimilated by following postalveolar ones, whereas there is no comparable assimilation towards alveolar in postalveolar–alveolar sequences.

(E3) Voiced–voiceless sibilant sequences are subject to regressive voice assimilation and will thus become voiceless.

2. Method

The investigation is based on acoustic measurements in comparably structured corpora of read, segmentally controlled sentences. Given the great diversity in the temporal dynamics of spectral changes that can occur in gradual assimilations, our analysis integrates both time and frequency, which has rarely been done in previous (acoustic) studies of place assimilation.

2.1. Sentence material

The speech corpus for the investigation of French sibilant sequences includes 72 sentences made up of three different subsets. The primary subset contains word pairs with all the eight possible sibilant sequences across word boundaries that result from the cross-combination of the features ‘alveolar’, ‘postalveolar’ and ‘voiced’, ‘voiceless’, i.e. (a) /s/, (b) /z/, (c) /s/, (d) /z/, (e) /s/, (f) /s/, (g) /s/, and (h) /z/. They are inserted into the three symmetrical vowel contexts /i/_/_i/, /a/_/_a/, and /u/_/_u/. This yields 24 test sentences. The three non-nasalized vowels are articulatorily and acoustically very distinct (cf. Meunier, French–Mestre, and Lelekov-Boissard (2003) for characteristic formant values of the French vowels) in order to determine whether the vowel contexts affect the degree of assimilation (cf. Hughes & Halle, 1956; Kühnert & Hoole, 2004; Narrey, 1982; Soli, 1981). For the same reason, the secondary subset of test sentences combines the four phonologically voiced or voiceless sibilant sequences (a)–(d) with all six asymmetrical vowel contexts, which leads to another 24 sentences.

The remaining 24 sentences form a complementary subset, in which each of the four individual sibilants /s/, /z/, /z/, and /s/ is paired across word boundaries with a labial consonant (C) like /p/ or /v/ in the two possible orders __C and C__. The labial consonants do not involve the tongue (tip) as an active articulator. So, articulatory interference with the sibilants is minimized. As in the primary subset, the eight sequences of the complementary subset were combined with the three symmetrical vowel contexts.

The individual sibilants function as reference sounds providing baseline acoustic values for spectral and temporal measures. Importantly, temporal measurements of single reference sibilants are invariantly and acoustically very distinct (cf. Fagyal et al., 2006; Kohler, 1990, 2002; Roach, 1983). Including sequences of sibilants and labial consonants allows us to observe whether the vowel contexts affect the degree of assimilation (cf. Hughes & Halle, 1956; Kühnert & Hoole, 2004; Narrey, 1982; Soli, 1981). For the same reason, the secondary subset of test sentences combines the four phonologically voiced or voiceless sibilant sequences (a)–(d) with all six asymmetrical vowel contexts, which leads to another 24 sentences.

The remaining 24 sentences form a complementary subset, in which each of the four individual sibilants /s/, /z/, /z/, and /s/ is paired across word boundaries with a labial consonant (C) like /p/ or /v/ in the two possible orders __C and C__. The labial consonants do not involve the tongue (tip) as an active articulator. So, articulatory interference with the sibilants is minimized. As in the primary subset, the eight sequences of the complementary subset were combined with the three symmetrical vowel contexts.

The individual sibilants function as reference sounds providing baseline acoustic values for spectral and temporal measures. Importantly, temporal measurements of single reference sibilants rather than geminates allows us to test whether we are observing assimilation or elision, particularly since geminates can also undergo elision (cf. Fagyal et al., 2006; Kohler, 1990, 2002; Roach, 1983). Including sequences of sibilants and labial consonants should also help making the aim of the study less obvious for the subjects.

Since the French part of the study presupposes that word-final schwa elision takes place, we forced this processes (which is anticipated in the sibilant conditions (a)–(h) above) in the reading material by using deliberately colloquial orthography that omitted word-final schwas in function words (e.g. "si j’veux", “sous l’drap”). Further means are described in Section 2.2. However, some target word pairs like “Miss chignon” ([iʃiʃ/]) in the sentence material could never have a word-final schwa. But as there were only a few of these target word pairs, and as all other target words pairs produced with schwa in between the two sibilants will be disregarded, the two types of pairs were not analysed separately.

Overall, the structure of the English sentence list is comparable to the French list. However, since English does not have /s/ at word edges, the primary subset only includes the four sibilant sequences (a), (b), (g), (h) above. For the same reason, the secondary subset contains just the two voiceless sequences (a)–(b). Thus, the two subsets are reduced to 12 sentences each. The complementary subset again combines the individual sibilants with labial consonants in the two possible orders __C and C__ across word boundaries. Due to the lack of /s/ at word edges, it consists of 18 sentences. So, in total, the English list comprises 42 sentences. The vowel contexts are also similar to the French contexts. However, the number of vowels had to be extended in order to find sensible English word pairs. Three pairs of vowel phonemes are used instead of three single phonemes. That is, in some cases /i/, /u/, and /a/ are substituted by /i/, /u/, or /a/, respectively. Although there are differences in the second formant frequency for these substitutes, the vowel phonemes in each pair differ mainly in the frequency of the first formant (cf. Wells, 1962). Since the spectra of sibilants vary primarily according to the F2 of the surrounding vowels (e.g. Soli, 1981), the different vowels in each pair should not have substantial influences on the acoustic manifestations and hence on the spectral measurements of the sibilants. Moreover, in view of Jongman (1998) it may be assumed that the quantity differences within each pair do not significantly affect the sibilant durations.

Finally, in both the French and the English sentences, the word pairs which contain the target sequences of vowels and consonants consist of two to six syllables and are placed towards the end of the sentences. The word boundaries between the sibilants are mainly located within a syntactic noun or verb phrase. The lists of French and English sentences are given in the Appendix A and B.

2.2. Recording procedure

The recording procedure was the same for French and English. That is, for each of the two languages four female subjects were recruited. All subjects produced the sentence list four times, and in each round the sentences in the list were given a differently randomized order. The French subjects were 20–46 years old. Following from hypothesis (F2), all four speakers can be assigned to Southern French varieties as they were born in Southern France and/or lived there for a long time. However, none of them had a clearly locatable regional accent. The latter was also true for the four English subjects who were between 22 and 29 years old. The restriction to female speakers was to exclude gender-specific effects on the durational and spectral measurements (cf. Gordon, Barthmainer, & Sands, 2002; Simpson, 2009).

Prior to the individual recordings, the speakers were instructed to read the sentences displayed on a screen in front of them as naturally as possible with a constant loudness suitable for normal conversation. Moreover, each sentence had to be said twice with different speaking styles: first, with a slow and careful pronunciation, as if talking in front of an audience; then with a fast, casual style, as if in an informal conversation with a good friend. The direct repetition and the resulting stylistic contrast was intended to elicit the fast, casual style in a consistent and clear way in the laboratory situation, as only the sentences of this style were to be analysed in terms of assimilation.

The French subjects produced the sentences at the Laboratoire Parole et Language of the Université de Provence. The English subjects were recorded at the Psychology Department of the University of York. The recordings were made digitally in anechoic chambers at a sampling rate of 44.1 kHz and with a 16 bit amplitude quantization. The first of the four repetitions of the sentence list was preceded by 10 dummy sentences, which

resembled the sentences on the list, and which were to familiarize the subjects with their task. The subjects were given a short break after each repetition of the sentence list. Overall, the recording sessions took about 40 min. Since the French list comprised 72 sentences, produced in four rounds with two different speaking styles, each of the four subjects produced 576 sentences, which resulted in a total number of 2304 sentences. The English recordings yielded 1344 sentences. A short interview after each session showed that none of the four French or four English speakers guessed the actual reason for the recording. Most of the subjects speculated that the study’s aim was phonetic effects of speaking-rate differences; others assumed that it was concerned with speech–melody patterns or with consonant–vowel interactions in general. The interviews were also recorded to get some spontaneous speech (around 3 min) from each subject.

2.3. Acoustic analyses

In preparation for the temporal and spectral measurements, the French and English sentences were labelled by means of PRAAT (cf. Boersma, 2001) with the boundaries of a number of linguistic units. The labelling was done manually by a trained phonetician, i.e. the first author (ON). Most importantly, the onsets and offsets of the single sibilants and sibilant sequences were marked. Furthermore, the boundaries of silent intervals within the sibilant sections were marked. As it is very likely that such interrupted articulations are different from those articulations in which the (phonological patterns of the) sibilant sections are produced continuously, the corresponding tokens were excluded from the temporal and spectral measurements.

2.3.1. Spectral measurements: CoG means and ranges

The labelled sibilant boundaries guided the spectral measurements that were done automatically by a MATLAB script that was written by Leonardo Lancia. The script determined the centre-of-gravity (CoG) measure. The CoG represents the average value of the frequencies in the spectrum, weighted by their amplitudes. In previous studies on a number of languages, the CoG values or similar measures turned out to be good representatives of the acoustic and perceptual differences between alveolar and postalveolar sibilants (cf. Goodacre & Nakajima, 2005; Gordon et al., 2002; Heinz & Strevens, 1961; Hoole, Nguyen-Trong, & Hardcastle, 1993; Hughes & Halle, 1956; Jassem, 1968; Jongman, Wayland, & Wong, 2000; Maniwa, Jongman, & Wade, 2009; Narrey, 1982; Tabain, 2001; Toda, 2007). The CoGs of postalveolar sibilants were found between 2 and 4 kHz, while the alveolar ones were substantially higher, typically between 4 and 8 kHz, and they showed more variation due to the speaker or the phonetic context. 1

The CoG measurements of the present study were based on spectral analyses with a 30 ms Hamming window (512 FFT points). It was shifted throughout the sibilant or the sibilant sequence in steps of 7 ms. Hence, the number of spectral slices varied according to the overall duration of the sibilant section. For each slice, the CoG was calculated within a frequency range between 1.5 and 15 kHz. This band-pass filter covers the spectral peaks that were found to be decisive for the distinction between alveolar and postalveolar sibilants (cf. also above). At the same time, the exclusion of frequencies lower than 1.5 kHz limits the risk that voicing, i.e. F0 and the adjacent high-energy harmonics, can pull down the CoG substantially. This band-pass filtering is a major difference to the CoG measurements implemented in PRAAT. In a following step, two CoG values were determined for each sibilant or sibilant sequence: the mean CoG and the CoG range. For the former, the CoGs of the individual slices of the respective sibilant or sibilant sequence were summed up and divided by its total number of slices. The CoG range represents the difference between the largest and the smallest CoG value within the slices of a sibilant or sibilant sequence.

The mean CoG conveys a holistic impression of the sibilant quality, or, in the case of the sequences, of the degree of assimilation. For instance, the more the /s/ in a /s]/ sequence is assimilated regressively by the postalveolar sibilant, the lower the mean CoG of the sequence will be. Taking assimilation patterns into account that were observed in previous studies (cf. the sibilant-sequence patterns A–D of Holst & Nolan, 1995, Section 1.2.2.), two different (but not mutually exclusive) assimilatory effects are conceivable. Firstly, the duration of the /s/ is shortened in favour of the duration of the /]. Secondly, the sibilant quality and hence the spectral-energy distribution of the /s/ is shifted towards postalveolar []. In order to distinguish between the two assimilatory effects, the CoG range was determined in addition to the mean CoG. If the assimilatory effect manifests itself primarily in terms of durational changes, then solely the mean CoG will change (i.e. decrease in the case of the /s]/ example), whereas the CoG range will remain constantly high. However, if the assimilation concerns a shift in sibilant quality, as, for example, in the case of the /s/ that is shifted towards [j], then the CoG range will decrease and approximate the CoG ranges of the corresponding single sibilants. In summary, although the mean CoG is the primary measure of the degree of assimilation, it must be interpreted in the context of the CoG ranges.

2.3.2. Temporal measurements: sibilant durations and the duration ratio

As outlined in Section 2.3.1, the integrated interpretation of the CoG means and ranges already provides some insights into assimilation-driven durational changes of the alveolar and postalveolar elements in the sibilant sequences. However, these insights are indirect. Therefore, the CoG measurements were complemented by direct measurements of the durations of the individual sibilant elements in the sequences. From these absolute durations we also calculated a duration ratio by dividing the duration of the first element in the sequence by the overall sequence duration. The result was then multiplied by 100 in order to arrive at percentages (i.e. Sibilant1/[Sibilant1 + Sibilant2] × 100). This ratio was able to represent the internal temporal structure of the sibilant sequences in a single value, independent of their overall durations. The ratios were used in the statistical tests.

Determining the duration ratios required that additional segmental boundaries be set within the sibilant sequences, referred to as the ‘B’ (=boundary) labels. Two precautions were taken in order to make the ‘B’ labelling as objective and reliable as possible. First, setting ‘B’ labels presupposed that there was a clear change in sibilant quality. So, all those sibilant sequences that showed a constant spectral pattern, as for example, when a /s]/ sequence was assimilated to [j], had to be excluded from the ‘B’ labelling. We used the CoG range as an objective indicator of spectral (in-)consistency. That is, for each speaker the maximum CoG range within the single sibilants was determined and this was compared with the CoG ranges of the sibilant sequences. Then, the ‘B’ labelling was restricted to only those sibilant sequences whose CoG ranges

---

1 Although this is a clear empirical foundation, we are aware that the performance of CoG measurements is limited. For example, Jones and Munhall (2003) suggest that small sibilant changes are not always detected by CoGs, particularly when they concern disordered productions. However, these are limitations of the sensitivity and the extended use of the measure. They do not concern the internal validity of CoG differences with respect to our hypotheses. Moreover, using CoGs allows for parametric comparisons with previous studies. Alternative and less frequently used measures (cf. Evers, Reetz, & Lahiri, 1998; L. Edwards, & Beckman, 2007) lack this scientific value and have other limitations. We accounted for the limitations of the CoG by refraining from conclusions from the completeness of the assimilation, at least unless judgments of human observers are also taken into account (cf. Section 2.3.2).
exceeded the maximum range of the single sibilants. Secondly, the ‘B’ labelling was done independently by two phoneticians. One of them came from the Institute of Phonetics and Digital Speech Processing of the University of Kiel (TO). The other trained phonetician was the first author (ON). Unlike the first author, the other phonetician was naive with respect to the aims and hypotheses of our production study. Moreover, both phoneticians were naive insofar as their native language was neither French nor English. The labelling was done by integrating visual inspections of spectrograms and close auditory assessments of the corresponding sound sections. After the labelling was completed, the temporal locations of the ‘B’ labels were compared, and all labels were disregarded that differed by more than 20% of the overall duration of the sibilant sequences (inter-observer agreement criterion). According to empirical studies, this percentage threshold (mostly corresponding to around 20–30 ms) should roughly represent the just noticeable difference of duration changes in speech (cf. Bochner, Snell, & MacKenzie, 1988).

In all sibilant sequences in which the temporal distances between the two independently set ‘B’ labels complied with the 20% criterion, the ‘B’ label of the first author was used to calculate the duration ratios. Due to the selection procedure described above, sibilant durations and duration ratios were only determined for a subset of the French and English data, whereas the CoG measurements covered the entire samples. The sizes of the duration subsamples are specified in the results section (cf. Figs. 3 and 7).

Fig. 2 shows characteristic examples of sibilant sequences that counted as spectrally inconsistent (a–b) or spectrally consistent (c) according to the CoG-range criterion. The ‘B’ labels in the voiceless or voiced alveolar–postalveolar sibilant sequences of examples (a) and (b) were set by the first author. Their temporal distances to the ‘B’ labels of the other phonetician met the interobserver agreement criterion. Thus, the durations of the sibilant sections to both sides of the ‘B’ labels entered the duration-ratio calculations. The excerpts of the speech signals shown in (a) and (b) are part of the French utterances “C’est une trousse chargée” (a) and “C’est une phrase japonaise” (b). Example (c) comes from the French utterance “Tu te caches si mal”.

2.3.3. Additional measurements on the general production framework

In addition to the spectral and temporal measurements of the sibilant sections themselves, several other measurements were calculated:

- the overall sentence duration without pauses,
- the durations of the vowels that surrounded the sibilant sequence,
- the proportion of voicelessness in the overall sibilant sequence for those sequences with heterogeneous phonological voice features.

While the overall sentence durations served to determine speaking rate differences between the speakers as well as between the fast and the slow speaking rate conditions, the vowel durations were used as a measure of the perceptual salience of the corresponding syllables or words. In perception experiments with speech or speech-like stimuli Gay (1978) and Kohler (2008) demonstrated that a syllable can be perceived as standing out against its neighbours by increasing, for example, its fundamental frequency, duration, and intensity levels. An increase in syllable duration, which is mainly done cross-linguistically by means of the vowel portion (cf. Barry, Andreeva, & Steiner, 2007; Hirst, Astésano, & Di Christo, 1998; Klatt, 1976), is one of the most powerful acoustic triggers of such perceptual salience. Duration-based variation in perceptual salience is involved in marking lexical-stress and/or pitch-accent positions, as well as information structure. On the basis of the voicelessness proportions, we tested previous statements in the literature on voice assimilation in French and English (cf. Section 1.2.3). Moreover, the proportions were used as a measure to determine possible interactions between voice and place assimilation patterns, mainly with regard to French (cf. hypothesis F2).

Finally, we also estimated the frequencies of the words whose initial or final sibilants formed the sibilant sequences. The French word frequencies were estimated on the basis of a corpus derived from movie subtitles (cf. New, Brysbaert, Veronis, and Pallier (2007), about 52 million words) and the newspaper ‘Le Monde’ (cf. van Rullen et al. (2005), about 10 million words). The English estimations were based on the British National Corpus (cf. Leech, Rayson, and Wilson (2001), about 18 million words).

2.4. Statistical analyses

Since our primary aim is to describe and compare the sibilant assimilation patterns in French and English, the analysis involved only the sentences of the fast, casual condition. Compared with the slow, careful conditions, they should evoke clearer assimilation patterns. Moreover, it is more likely that the fast, casual condition triggered word-final /s/ elision and hence created the basic condition required for sibilant assimilation by the Southern French speakers.
Based on the sub-sample of the fast, casual sentences, three analogous mixed-model, linear-regression analyses were performed (cf. Baayen, Davidson, & Bates, 2008) on the sibilant sequences for each language to examine each of the dependent variables separately, i.e. duration ratio, mean CoG, and CoG range. All three analyses included three groups of independent variables: (a) order of place of articulation within the sibilant sequence and phonological voice features within the sibilant sequence, (b) following and preceding vowel context, relative word frequency of each word, and duration of the two vowels surrounding the sibilant sequence, and (c) speaker as well as sentence. The two variables in the group-(c) were treated as random effects. The remaining variables were treated as fixed effects that represented either control variables (group-(b)) or experimental variables (group-(a)).

In supplement to individual target-word frequencies, some studies suggest that the co-occurrence frequency of the two target words can be an informative complementary measure of assimilation-determining factors. For example, extreme degrees of reduction (i.e. assimilations/elisions) could be more likely, if the two target words not only have high individual frequencies, but also have high co-occurrence frequency, as in the case of sequences of function words (cf. Jäger and Hoole, 2007). However, for the present study we preferred using just the individual frequencies as they show clear differences across items, whereas the co-occurrence frequencies are all similarly low. As regards the adjacent vowel contexts, the statistical analyses were based on two contrasts: /i,ʊ/ vs. /a,ʌ/, /u,ɨ/, which correspond to the dimension back, rounded vs. non-back, unrounded, and /i, ʊ/ vs. /a,ʌ/, which represents the closed vs. open dimension (the second vowel in the pairs only applies to the English material). For the CoG measures, four additional mixed-model, linear-regression analyses were performed for each language to compare the sibilant sequences with the single alveolar and postalveolar sibilants.

In all analyses variables with coefficients more than two standard deviations from zero \((T > 2)\) were regarded as reliable (cf. Baayen et al., 2008). Log-likelihood ratio tests also confirmed that removing those factors significantly impacted the fit of the data to the model and removing factors with unreliable coefficients \((T < 2)\) did not have a significant effect on the fit. The results of the analyses are summarized for each language in Sections 3.2 and 3.3. Section 3.1 briefly presents some general characteristics of the sentence and sibilant realizations within the French and English corpora.

3. Results

3.1. Introductory remarks

Before conducting the main analyses on the temporal and spectral patterns in the French and English sibilant sequences, two preliminary analyses were conducted. First, careful comparisons of the speaking-style conditions were conducted to ensure that the speakers were (a) speaking more quickly in the fast, casual style than the slow, careful style and (b) that the fast casual style was similar to spontaneous speech.

For each of the 8 speakers, overall sentence duration (without pauses) was compared in the two speaking style conditions using t-tests. The results showed that each speaker produced the fast, casual sentences highly significantly shorter than the slow, careful sentences \((19.02 \leq t \leq 27.45; df = 287; p < 0.0001)\). For most of the speakers, this shortening amounted to 30–35%. The French sentences in the two speaking-style conditions were also shorter than the English sentences, although they were no less complex \((t = -52.48; df = 2599; p < 0.001)\). This additional observation is consistent with known cross-linguistic differences in speaking rate (Dellwo, Ferragne, & Pellegrino, 2006).

A 10-second sample of spontaneous speech from each speaker was also randomly selected. Syllables per second were calculated (without pauses) for the spontaneous sample and compared to syllables per second in the two speaking-style conditions for each speaker. In both French and English the speaking rates of the spontaneous sections \((7.2–10.3 \text{ syl/s})\) were clearly more similar to the rates of the fast, casual sentences \((7.5–10.0 \text{ syl/s})\) than to the rates of the slow, careful sentences \((5.0–6.5 \text{ syl/s})\). One exception was the French speaker CDL. CDL produced the slow, careful speech at roughly the same rate as the fast, casual speech of the other French speakers, and she produced considerably more syllables per second in the fast, casual sentences (about 11.8 syl/s) than in her own spontaneous-speech section (9.5 syl/s). Overall, the syllable-per-second differences between the slow, careful sentences and the fast, casual or spontaneous sentences are in line with empirical values of previous studies (cf. Koreman, 2006).

In summary, while the French sentences had generally shorter overall durations and had higher speaking rates than the English sentences, all French and English speakers managed to produce the speaking-style difference in consistent and clear ways. Moreover, given the similarity in speaking rates between the fast, casual conditions and the actual spontaneous speech, the production of the sibilant conditions may be regarded as representative of everyday communication (the French speaker CDL being the only exception).

The second preliminary analysis compared the durations of the reference sibilants and sibilant sequences in the French and English fast, casual sentences. The t-tests yielded highly significant differences between single sibilants and the sibilant sequences of each speaker (French: \(-29.55 \leq t \leq 33.67; 276 \leq df \leq 282; p < 0.0001\); English: \(17.41 \leq t \leq 21.84; 156 \leq df \leq 164; p < 0.0001\)). For most of the speakers, the sibilant sequences were between 140 and 180 ms long, whereas the single sibilants had overall durations of 60–90 ms, making them roughly half as long as the sibilant sequences. The English speaker CMO and the French speaker CDL deviated from this pattern, showing single sibilants just 25% shorter than their sibilant sequences. Nevertheless, there were overall just 11 cases (out of 1700) in which a speaker produced a single sibilant that was longer than her shortest sibilant sequence. Separate assessments revealed that in all of these cases the overlap was due to exceptionally long single sibilants rather than to exceptionally short sibilant sequences and these long single sibilants could be accounted for by hesitational, pre-pausal, or accent-clash-related lengthening (cf. Delattre, 1966; Klatt, 1976). Thus the overall durations of the single sibilants and the sibilant sequences fell into discrete distributions, suggesting that any changes in spectral quality found in the main analysis are due to assimilations, rather than elisions.

Finally, we report on the number of analysable sentences with no pauses or schwas that interrupted the target word pairs in the fast and slow conditions. In the fast condition 1092 (or 94.8%) of the target word pairs showed single sibilants and sibilant sequences in the French and English sentences, all French and English speakers managed to produce the speaking-style difference in consistent and clear ways. Moreover, given the similarity in speaking rates between the fast, casual conditions and the actual spontaneous speech, the production of the sibilant conditions may be regarded as representative of everyday communication (the French speaker CDL being the only exception).

The second preliminary analysis compared the durations of the reference sibilants and sibilant sequences in the French and English fast, casual sentences. The t-tests yielded highly significant differences between single sibilants and the sibilant sequences of each speaker (French: \(-29.55 \leq t \leq 33.67; 276 \leq df \leq 282; p < 0.0001\); English: \(17.41 \leq t \leq 21.84; 156 \leq df \leq 164; p < 0.0001\)). For most of the speakers, the sibilant sequences were between 140 and 180 ms long, whereas the single sibilants had overall durations of 60–90 ms, making them roughly half as long as the sibilant sequences. Nevertheless, there were overall just 11 cases (out of 1700) in which a speaker produced a single sibilant that was longer than her shortest sibilant sequence. Separate assessments revealed that in all of these cases the overlap was due to exceptionally long single sibilants rather than to exceptionally short sibilant sequences and these long single sibilants could be accounted for by hesitational, pre-pausal, or accent-clash-related lengthening (cf. Delattre, 1966; Klatt, 1976). Thus the overall durations of the single sibilants and the sibilant sequences fell into discrete distributions, suggesting that any changes in spectral quality found in the main analysis are due to assimilations, rather than elisions.

Finally, we report on the number of analysable sentences with no pauses or schwas that interrupted the target word pairs in the fast and slow conditions. In the fast condition 1092 (or 94.8%) of the French sentences and 663 (or 98.7%) of the English sentences were analysable. In the slow, careful sentences, the English speakers inserted pauses and/or prosodic phrase boundaries between 215 (or 22.9%) of the overall 672 target word pairs. However, in French even 795 (or 66.2%) of the target word pairs were interrupted. These interruptions were all due to schwa insertions, partly produced in connection with a pause after the end of the first target word. This result lends clear support to the claim that word-final /a/ elision is not at all a default process for our Southern French speakers but that it is also quite common (in our study 94.8% of the time) for speakers of this region under fast, casual speaking-style conditions (cf. Gadet, 1992; Pustka, 2008). To what extent the /a/ elisions entailed place assimilations in the sibilant sequences of the Southern French speakers and what
these assimilation patterns looked like was explored by the following analyses.

3.2. Analyses of the French sibilants

3.2.1. Duration ratios within the sequences

The results of the mixed-model, linear-regression analysis given in Table 1 show a strong main effect of the order of place of articulation on the duration ratios \( \text{Sibilant1/(Sibilant1 + Sibilant2)} \times 100 \) calculated for the sibilant sequences by means of the ‘B’ label segmentation. This reflects a tendency for the postalveolar section to be longer than the alveolar section. That is, in the postalveolar–alveolar sequences the initial section contributes more to the overall duration of the sibilant sequence than the second section. In the alveolar–postalveolar sequences it is the other way round.

This effect is illustrated in Fig. 3. It shows scatterplots of the individual sibilant durations within the sequences for ABO, MTE, CDL, and IVI. Except for IVI, it is clear that the black symbols of the alveolar–postalveolar and the grey symbols of the postalveolar–alveolar sequences are mainly located on different sides of the dotted diagonal that represents equal sibilant durations within the sequence. From the perspective of absolute durations it is interesting to note that particularly ABO and MTE produced the postalveolar sections in the sequences apart from a few exceptions clearly longer (i.e. > 90 ms) than the single reference sibilants. In contrast, the alveolar sections of the sequences were mostly below 50 ms and hence shorter than the single references. The speaker-specific scaling of the axes reflect the statement in Section 3.1 that the sibilant durations of CDL are remarkably shorter than those of the other three speakers.

Significant effects of the control variables are restricted to the durations of the adjacent vowels. Word frequencies, surrounding vowel qualities, and phonological voice patterns did not affect the internal temporal structure of the sibilant sequences (cf. Table 1).

3.2.2. Centre-of-gravity values within the sequences

As regards the experimental variables (group-(a)), the mean CoGs showed a significant effect of the voice pattern (cf. Table 2). Phonologically voiceless sequences yielded higher mean CoGs...
than voiced sequences and sequences with heterogeneous voice features. For example, it can be seen in the panels of Fig. 4 that for all four speakers at least one of the /z/ or /r/ sequences yielded a lower mean CoG than the voiceless /s/ or /P/ counterpart. Moreover, for ABO, CDL, and IVI either /z/ or /r/ had the lowest mean CoG of all sequences. In addition to the voice-pattern effect, Table 2 shows clear significant effects of the order of place of articulation. Across all speakers, the mean CoGs were lower and the CoG ranges were smaller for the alveolar–postalveolar than for the postalveolar–alveolar sequences. At the same time, alveolar–postalveolar as well as postalveolar–alveolar sequences both differed significantly in their CoG means and ranges from the single alveolar and postalveolar references (cf. Table 3).

However, it becomes obvious from Fig. 4 that not all four speakers contribute equally to these overall differences. The effects of the order of place of articulation are primarily due to ABO and MTE. The fact that the sequences differ from either of the reference sibilants relies primarily on MTE and IVI.

In order to assess the production patterns in more detail, Fig. 5 displays the mean CoG (horizontal axis) and the CoG range (vertical axis) values for the individual tokens of the sibilant sequences and reference sibilants in speaker-specific scatterplots. Fig. 5 shows even clearer than Fig. 4 that distribution of the sibilant sequences is more variable than the distributions of the reference sibilants in terms of both means and ranges. At the same time, the alveolar–postalveolar and postalveolar–alveolar sequences overlap for all four speakers. That is, each speaker produced some tokens with (1) large ranges and intermediate mean CoGs, which points to transitioning CoGs across the sequences; and other tokens were produced with (2) CoG means and ranges similar to those of the single postalveolar sibilants. However, (1) can be found more frequently for IVI and CDL, whereas (2) is mainly true for ABO and MTE, more specifically for their alveolar–postalveolar sequences. Thus the distributions of the alveolar–postalveolar and postalveolar–alveolar sibilant-sequence tokens overlap less strongly for ABO and MTE than for IVI and CDL. The scatterplot of CDL also shows that the overlap between the sequences and the postalveolar references is actually stronger than Fig. 4 suggested. The more /s/-like mean CoGs of the sequences in Fig. 4 are based on a very large range of values, and many of these values fall into the range of /j/. Besides the speaker-specific patterns, whose temporal

<table>
<thead>
<tr>
<th>Comp. var. levels</th>
<th>Mean CoG</th>
<th>CoG range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group-(a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order place of articulation</td>
<td>117.63</td>
<td>54.31</td>
</tr>
<tr>
<td>Voice (vlvl–vdvd)</td>
<td>241.20</td>
<td>73.90</td>
</tr>
<tr>
<td>Voice (vlvl–vlvd)</td>
<td>189.34</td>
<td>87.44</td>
</tr>
<tr>
<td>Voice (vlvl–vdvl)</td>
<td>241.82</td>
<td>94.82</td>
</tr>
<tr>
<td>Group-(b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preceding vowel (/u/–/i//a/)</td>
<td>−61.93</td>
<td>59.98</td>
</tr>
<tr>
<td>Following vowel (/u/–/i//a/)</td>
<td>92.79</td>
<td>61.87</td>
</tr>
<tr>
<td>Vowel duration (initial)</td>
<td>1357.40</td>
<td>1015.83</td>
</tr>
<tr>
<td>Vowel duration (final)</td>
<td>1676.50</td>
<td>1074.83</td>
</tr>
<tr>
<td>Frequency (1st word)</td>
<td>0.30</td>
<td>7.47</td>
</tr>
<tr>
<td>Frequency (2nd word)</td>
<td>−1.74</td>
<td>6.33</td>
</tr>
</tbody>
</table>

Table 2
Results of the mixed-model, linear-regression analyses for the sibilant sequences, based on the predictor variables mean CoG (left) and on the CoG range (right) for French. From left to right, the names of the (levels of the) fixed-effects variables as well as the corresponding regression coefficients, standard errors, and T-values are given. As for the voice features, ‘vd’ and ‘vl’ refer to ‘voiced’ or ‘voiceless’. Asterisks indicate significant outcomes. Unreported values are not significant.

organizations are examined more closely in Section 3.2.4, Fig. 5 displays one crucial aspect very clearly. While all speakers produced many or most sibilant sequences with CoG means and ranges similar to those of the single postalveolar sibilants, there are only a few or no sibilant sequences similar to the alveolar referents. This fact is also reflected in the T-values of the significant mean-CoG conditions in Table 3, which are considerably higher in connection with the alveolar than with the postalveolar reference. On this basis, it is not an exaggeration to say that, despite the significant CoG differences to both references, the sibilant sequences were overall more similar to single postalveolar than to single alveolar sibilants.

As regards the reference sibilants, Fig. 5 shows that they form two clear groups of tokens, differing in mean CoG. The mean CoGs were significantly lower for postalveolar than for alveolar reference sibilants (cf. Table 3). Additionally, there is a smaller effect of sibilant type on the CoG ranges. Across all speakers, they were on average slightly higher for the alveolar than for the postalveolar sibilants (cf. Table 3). This is more clearly illustrated in Fig. 4 than in Fig. 5.
None of the group-(b) control variables had an influence on the mean CoG values. This holds for the single reference sibilants as well as for the sibilant sequences. However, both single sibilants and sibilant sequences have in common that the CoG ranges were significantly affected by the following vowel quality. The ranges were larger before back, rounded /u/ than before /a/ and /i/ (cf. Tables 2 and 3).

3.2.3. Summary

Our analyses of sequences of alveolar and postalveolar sibilants produced by Southern French speakers revealed a clear tendency to change the alveolar in favour of the postalveolar element. The duration of the alveolar section in the sequences was reduced relative to the postalveolar, and/or the alveolar was changed in its spectral characteristics towards the postalveolar. These alveolar-to-postalveolar shifts in the time and frequency domains occurred for both alveolar–postalveolar and postalveolar–alveolar sibilant sequences, but they were stronger in the former sequences. In terms of CoG, the spectral alveolar-to-postalveolar shifts created sequences that fell within the range of the single postalveolar references. However, within this overall pattern notable speaker-specific differences emerged. Some of the differences were presented in Section 3.2.2. In the following this presentation is supplemented with a focus on the temporal organization of the sequences in the time and frequency domains.

3.2.4. Inter-speaker differences

The significant shift in the internal temporal structure of the sibilant sequences in favour of the postalveolar portion can mainly be ascribed to the three speakers ABO, MTE, and CDL.

![Fig. 6. Examples of CoG time courses (in Hz) that characterized the postalveolar reference sibilants as well as the alveolar–postalveolar and postalveolar sequences of the four French speakers. As CoG values were taken at intervals of 7 ms, the numbers of values within the time courses vary with the overall durations of the sibilants or sibilant sequences.](image-url)
However, these similar relative shifts underlie individual strategies that are reflected in Fig. 3. MTE shifted the relative sibilant portions towards postalveolar mainly by shortening the alveolar section in the postalveolar–alveolar sequences, while leaving the mean duration of the initial sibilant in both the postalveolar–alveolar and the alveolar–postalveolar ordering almost constant. Correspondingly, the black and grey symbols of MTE’s scatterplot in Fig. 3 are well separable when they are projected onto the y-axis of the second sibilant, whereas they overlap largely when they are projected onto the x-axis of the first sibilant. As a consequence of MTE’s strategy, her scatterplot also indicates that the overall durations of alveolar–postalveolar sequences were longer compared with postalveolar–alveolar sequences. CDL used a strategy that was diametrically opposed to the strategy of MTE. Instead of shortening the alveolar portion in the postalveolar–alveolar sequences, CDL lengthened the postalveolar portion in these sequences. At the same time, the duration of the second sibilant in the postalveolar–alveolar and alveolar–postalveolar sequences only contributed slightly to shift the relative sibilant portions towards postalveolar. In consequence, CDL’s postalveolar–alveolar sequences showed longer overall durations than her alveolar–postalveolar sequences. Finally, it is clearly illustrated in Fig. 3 that ABO was the only speaker who created the longer postalveolar portions in the sibilant sequences by means of complementary changes of the durations of both sibilants. In her scatterplot the black and grey symbols overlap in similar strong ways when projected onto the x- and y-axes. Thus the overall durations of all eight sibilant sequences remained approximately constant for ABO.

Inter-speaker differences also show up in the spectral domain. In order to illustrate some of these differences, Fig. 6 presents separately for each of the four French speakers characteristic examples of CoG time courses across a postalveolar reference sibilant as well as across an alveolar–postalveolar and a postalveolar–alveolar sequence. In order to facilitate reliable comparisons between the CoG time courses, all examples come from symmetrical /a_–a/ contexts. The examples reflect the mean CoG and CoG range patterns that are shown in Figs. 4 and 5. That is, the CoG time courses of the postalveolar reference sibilants in Fig. 6 fluctuate within a constant narrow range, though at a speaker-specific level. Such narrowly limited CoG fluctuations also characterize the alveolar–postalveolar sequence of MTE as well as the alveolar–postalveolar and postalveolar–alveolar sequences of ABO. However, in all other cases, the CoG time courses changed across the sibilant sequences, which led to higher mean CoGs and larger CoG ranges in Figs. 4 and 5. The most important additional information that is contained in Fig. 6 concerns the different dynamics of these changes across the CoG time courses. For CDL, the CoG values fall or rise continuously across the alveolar–postalveolar or postalveolar–alveolar sequences, respectively. In contrast, IVI produced her alveolar or postalveolar sibilant sections typically with two high or low plateau-like CoG sections that were separated by an abrupt transition. A similar discontinuous CoG pattern characterizes the postalveolar–alveolar sequences of MTE. But while the two plateau-like CoG sections of IVI had roughly equal durations (which were longer in the postalveolar alveolar than in the alveolar–postalveolar sequences), the plateau-like postalveolar section of MTE was typically clearly longer than her plateau-like alveolar section.

3.3. Analyses of the English sibilants

3.3.1. Duration ratios within the sequences

The mixed-model, linear-regression analysis yielded a strong significant effect of the order of the place of articulation. Alveolar–postalveolar sequences showed shorter duration ratios than postalveolar–alveolar sequences (cf. Table 4). That is, the internal temporal structures of the sibilant sequences are shifted in favour of the postalveolar portion. This effect, which also manifests itself in the different clustering of absolute durations along the x-axis or y-axis in Fig. 7, is the same as in French, though quantitatively less pronounced. Moreover, it becomes obvious from Fig. 7 that in all scatterplots the black and grey symbols cluster in the upper left section. This means that the second sibilant section in the sequence was produced consistently longer than the first, irrespective of the place-of-articulation effect. While average reference sibilant durations were between 60 and 90 ms, the second sibilant in a sequence was frequently produced with durations of more than 100 ms (in some cases up to 150 ms). As regards the group-(b) control variables, the only significant effect on the duration ratios was due to the preceding vowel contrast /u/ vs. /a/ and /i/ vs /a/. Compared with the latter vowel qualities, /u/ or /i/ enhanced the order of place of articulation effect by increasing the duration of the postalveolar section at the cost of the alveolar section in both sibilant orders. In general, the English temporal patterns showed fewer speaker-specific differences than the French patterns. However, in terms of the number of measurements per type of sibilant sequence the English data set was smaller than the French set. This holds particularly for the alveolar–postalveolar sequences. A main reason for the cross-linguistically different sample sizes is that far more of the English sequences were realized with voiceless /v/ and ‘vl’ refer to ‘voiced’ or ‘voiceless’. Asterisks indicate significant outcomes.

<table>
<thead>
<tr>
<th>Comp. var. levels</th>
<th>Coefficient</th>
<th>Std. error</th>
<th>T-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group-(a) Order of place of articulation</td>
<td>0.11</td>
<td>0.02</td>
<td>4.47*</td>
</tr>
<tr>
<td>Voice (vlvl–vdvd)</td>
<td>−0.01</td>
<td>0.05</td>
<td>−0.26</td>
</tr>
<tr>
<td>Voice (vlvl–vlvd)</td>
<td>−0.04</td>
<td>0.03</td>
<td>−1.15</td>
</tr>
<tr>
<td>Voice (vdvl–vdvl)</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Group-(b) Preceding vowel (/u/–/i/</td>
<td>/a/)</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Preceding vowel (/i/–/a/)</td>
<td>0.02</td>
<td>0.03</td>
<td>0.73</td>
</tr>
<tr>
<td>Following vowel (/u/–/i/</td>
<td>/a/)</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Following vowel (/i/–/a/)</td>
<td>−0.05</td>
<td>0.02</td>
<td>−1.90</td>
</tr>
<tr>
<td>Vowel duration (initial)</td>
<td>0.01</td>
<td>0.17</td>
<td>0.07</td>
</tr>
<tr>
<td>Vowel duration (final)</td>
<td>−0.69</td>
<td>0.37</td>
<td>−1.89</td>
</tr>
<tr>
<td>Frequency (1st word)</td>
<td>0.01</td>
<td>0.01</td>
<td>1.85</td>
</tr>
<tr>
<td>Frequency (2nd word)</td>
<td>−0.01</td>
<td>0.01</td>
<td>−1.24</td>
</tr>
</tbody>
</table>

3.3.2. Centre-of-gravity within the sequences

According to Table 5 order of place of articulation had an effect on CoG means and ranges. As was found for French, the alveolar–postalveolar sequences showed lower mean CoGs and smaller CoG ranges than the postalveolar–alveolar sequences. Voiceless sequences also had significantly higher mean CoGs and smaller CoG ranges than voiceless–voiced sequences (cf. particularly DGO and FDU in Fig. 8).

Furthermore as is displayed in Figs. 8 and 9 and statistically supported in Table 6, the CoG means and ranges of the postalveolar–alveolar sequences differed from both the alveolar
and the postalveolar references. The majority of the means and ranges fell almost exactly between the values of the two references. The intermediate means and ranges suggest that the sequences were realized with transitions between two clear postalveolar and alveolar endpoints. In contrast, the alveolar–postalveolar means and ranges differed only from the alveolar, but not from the postalveolar CoG references. This descriptive and statistical finding represents a major difference from the French results. It implies that the alveolar–postalveolar sequences were mostly realized with a roughly stable postalveolar quality. With regard to spectral stability, it is worth noting that, as in French, the single reference sibilants show a small, but significant difference in CoG ranges. They are larger for the alveolar than for the postalveolar references. However, as can most clearly be seen in Fig. 9, the range difference is primarily based on CMO and FDU. The postalveolar references are also produced with significantly lower mean CoGs than the alveolar references. Unlike the range difference, the difference in mean CoGs is similarly large for all four speakers.

The spectral patterns of the English sibilant sequences were additionally influenced by the group-(b) control variables (cf. Table 5). For example, there was an effect of the vowel context.

Table 5
Results of the mixed-model, linear-regression analyses for the sibilant sequences, based on the predictor variables mean CoG (left) and on the CoG range (right) for English. From left to right, the names of the (levels of the) fixed-effects variables as well as the corresponding regression coefficients, standard errors, and T-values are given. As for the voice features, ‘vd’ and ‘vl’ refer to ‘voiced’ or ‘voiceless’. Asterisks indicate significant outcomes. Unreported values are not significant.

<table>
<thead>
<tr>
<th>Comp. var. levels</th>
<th>Mean CoG</th>
<th>CoG range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std. error</td>
</tr>
<tr>
<td>Group-(a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order place of articulation</td>
<td>841.85</td>
<td>43.37</td>
</tr>
<tr>
<td>Voice (vlvl–vlvd)</td>
<td>254.36</td>
<td>113.16</td>
</tr>
<tr>
<td>Voice (vlvl–vdvl)</td>
<td>−76.28</td>
<td>56.84</td>
</tr>
<tr>
<td>Group-(b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preceding vowel (/u–/i//a/)</td>
<td>−237.92</td>
<td>40.77</td>
</tr>
<tr>
<td>Following vowel (/u–/i//a/)</td>
<td>−72.41</td>
<td>37.30</td>
</tr>
<tr>
<td>Frequency (1st word)</td>
<td>13.797</td>
<td>14.829</td>
</tr>
<tr>
<td>Frequency (2nd word)</td>
<td>29.725</td>
<td>12.087</td>
</tr>
</tbody>
</table>

Fig. 7. Durations (in seconds) of the first and second sibilant portions in the sub-samples of English alveolar–postalveolar (black symbols) and postalveolar–alveolar sequences (grey symbols) that were given a ‘B’ label. The sample sizes across the 4 different sibilant sequences and speakers are given at the top.
However, unlike French, the English effect concerned not the CoG ranges, but mean CoGs, and it was caused by the preceding and not the following vowel quality. The CoG means were significantly lower for preceding back, rounded /u,ʊ/ compared to preceding /i,ɨ/ and /a,ʌ/. Furthermore, while there were no effects of word frequency in French, a higher frequency of occurrence of the second word in the English target-word pairs led to higher mean CoGs and larger CoG ranges in the sibilant sequences. We will return to the group-(b) effects in more detail in Section 4.3 of the discussion.

3.3.3. Inter-speaker differences

The French speakers clearly differed in their ‘B’-label related temporal organizations of the sibilant sequences. The English speakers, however, all behaved very similarly in this respect. Yet, there is also an inter-speaker difference that concerns the dynamics of the spectral change across the postalveolar–alveolar sibilant sequences.

Fig. 8. Mean CoGs (dots) and CoG ranges (vertical bars) of the three single reference sibilants (grey, left) and the four different sibilant sequences (right), averaged across all corresponding tokens of each of the four English subjects.

Fig. 9. Distribution of individual tokens for the reference sibilants (postalveolar: grey circles, alveolar: grey triangles) and sibilant sequences (alveolar–postalveolar: crosses, postalveolar–alveolar: open circles) for each of the four English subjects.
Table 6
Excerpt of the results of the mixed-model, linear-regression analyses for the comparison between sibilant sequences and single sibilants for English, starting from the alveolar (top) or the postalveolar (bottom) reference, with the predictor variables mean CoG (left) and CoG range (right). From left to right, the names of the (levels of the) fixed-effects variables as well as the corresponding regression coefficients, standard errors, and T-values are given. As for the voice features, 'vd' and 'vl' refer to 'voiced' or ‘voiceless’; ‘av’ and ‘pav’ refer to alveolar and postalveolar. Asterisks indicate significant outcomes. Unreported values are not significant.

<table>
<thead>
<tr>
<th>Comp. var. levels</th>
<th>Mean CoG</th>
<th>CoG range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std. error</td>
</tr>
<tr>
<td><strong>Alveolar</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place (av–av/pav)</td>
<td>2107.47</td>
<td>117.34</td>
</tr>
<tr>
<td>Place (av–pav/av)</td>
<td>1285.12</td>
<td>113.47</td>
</tr>
<tr>
<td>Place (av–pav)</td>
<td>2044.56</td>
<td>123.41</td>
</tr>
<tr>
<td>Frequency (2nd word)</td>
<td>−12.07</td>
<td>12.84</td>
</tr>
<tr>
<td><strong>Postalveolar</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place (pav–av/pav)</td>
<td>−62.91</td>
<td>116.32</td>
</tr>
<tr>
<td>Place (pav–pav/av)</td>
<td>759.45</td>
<td>112.98</td>
</tr>
<tr>
<td>Place (pav–av)</td>
<td>2044.56</td>
<td>123.41</td>
</tr>
<tr>
<td>Frequency (2nd word)</td>
<td>−12.07</td>
<td>12.84</td>
</tr>
</tbody>
</table>

Fig. 10. Examples of CoG time courses (in Hz) that characterized the postalveolar reference sibilants as well as the alveolar–postalveolar and postalveolar sequences of the four English speakers. As CoG values were taken at intervals of 7 ms, the numbers of values within the time courses vary with the overall durations of the sibilants or sibilant sequences.
sequences of each of the four English speakers. All examples were
taken from symmetrical /a__a/ vowel contexts. In line with the
CoG means and ranges of Figs. 8 and 9, it can be seen in Fig. 10 that
all four English speakers produced both postalveolar references and
alveolar–postalveolar sequences CoG values that oscillated in the
narrow range between 7 and 8 kHz. Compared with these spectrally
almost constant patterns, the CoG time courses of the four speakers
showed clear changes across the postalveolar–alveolar sequences.
However, while the CoG transition from the lower postalveolar to
the higher alveolar level typically took place in a roughly continuous
fashion for CMO and FDU, the CoG time courses of FAY and DGO
were characterized by quite abrupt transitions in between two more
or less flat running low and high CoG sections.

3.4. Voice assimilation patterns

While the four French speakers ABO, MTE, CDL, and IVI showed
speaker-specific place assimilation patterns, their voice assimilation
patterns were all very similar. Therefore, Fig. 11 presents the
measurements for each sibilant sequence pooled across the four
French speakers. As can be seen in Fig. 11, the two voiced–voiceless
sequences /zζ/ and /sζ/ were largely phonetically voiceless, whereas
the two voiceless–voiced sequences /ζz/ and /ζζ/ were produced
predominantly voiced. A t-test showed that this difference between
the two pairs of sequences was highly significant ($t = -12.66; \ df = 180; \ p < 0.0001$). Moreover, as is implied by the error bars in
Fig. 11, we found a small number of regressive assimilations that led
to complete voicing or voicelessness. However, in the vast majority
of cases, the regressive assimilations left a small portion of the
original voice feature.

While the pairs of voiceless–voiced and voiced–voiceless sequences differed clearly and significantly in their relative voice-
less intervals, no such differences can be observed in Fig. 11
between the two alveolar–postalveolar and postalveolar–alveolar pairs. Correspondingly, a t-test that compared the voiceless intervals
for the two pairs with the opposite orders of place of articulation was far from being significant.

As for the English sibilant sequences, Fig. 11 shows that
different voicing patterns were found for CMO compared to
FAY, DGO, and FDU. The latter three speakers produced both
voiceless–voiced and voiced–voiceless sequences on average
85–90% voiceless. CMO shows a different behaviour in the
two kinds of sequences. While her voice–voiceless sequences
were also around 85% voiceless, the voiceless–voiced sequences
remained about half-way voiced. According to a separate t-test for
CMO, this difference in the relative voiceless interval was very
significant ($t = -3.72; \ df = 11; \ p < 0.01$).

4. Discussion

4.1. Basic interpretation of the French results

In the French sequences the postalveolar sibilant sections are
consistently longer than alveolar sections. Likewise, the mean CoGs
of the sibilant sequences are on the whole more similar to the
postalveolar than to the alveolar references, and there is a strong
overlap between the CoG ranges of the sequences and single
sibilants. At the same time, sibilant sequences are consistently
longer than single sibilants. This overall pattern can be interpreted
as assimilation towards postalveolar in either direction, i.e. progres-
sive for postalveolar–alveolar sequences and regressive for alveo-
lar–postalveolar sequences. However, both the duration and mean CoG
measures show a stronger influence of postalveolar when it is the
second sibilant in the sequence suggesting that regressive place
assimilation is stronger (presumably more often complete) and
more consistent across speakers than progressive assimilation.

For the first time, the present study provides systematic
empirical evidence for the presence of assimilation of place of
articulation in French. In line with hypothesis (F1), the assimila-
tion of place of articulation was found in sequences of alveolar
and postalveolar sibilants. Furthermore, as hypothesized in (F2)
the assimilation showed up for Southern French speakers after
schwa deletion and beyond the frequent and possibly fixed expressions ‘je sais’ and ‘je suis’. Assimilation of place of articulation in French sibilant sequences may hence be regarded as
a productive, widespread phenomenon that takes place across
different speakers (from different regions), and combinations of
words. Finally, as was claimed in hypothesis (F3), the assimila-
tion went in the direction alveolar-to-postalveolar and occurred pro-
gressively as well as regressively. However, there were also
substantial speaker-specific patterns. Speakers differed in the
extent they assimilated regressively and progressively (ABO vs.
MTE), in the assimilatory strategy – which could be a strengthen-
ing of the postalveolar and/or the weakening of the alveolar
element (MTE vs. CDL) – and in the extent they showed assimila-
tion in the time and frequency domains (ABO vs. CDL vs. IVI).

In terms of individual differences in the degree of assimilation,
the speakers can be put into three groups. According to the CoG
means and ranges, ABO produced almost all sibilant sequences with
the same spectral energy distribution and spectral stability as the
postalveolar references. In the few spectrally transitioning sequences
that occurred for ABO, the postalveolar sections were lengthened at
the cost of the alveolar sections. Thus ABO can be classified as
a strong progressive and regressive assimilator. The opposite is true for
IVI. In the vast majority of cases, she produced spectrally well
separated alveolar and postalveolar sections without durational
precedence of either section. Therefore, IVI may be regarded as a
weak assimilator. The other two speakers fall in between ABO and IVI.
In the alveolar–postalveolar sequences, MTE exhibited as strong
regressive assimilations as ABO. But in the postalveolar–alveolar
sequences, MTE’s progressive assimilations were largely restricted
to the temporal domain, i.e. to a lengthening of the postalveolar
section. So, MTE represents a strong regressive but moderate pro-
gressive assimilator. The fourth speaker, CDL, may be classified as
a moderate regressive and progressive assimilator. Although most of
her sibilant sequences were realized with clear spectral transitions,
these transitions were continuous. Compared with abrupt spectral
changes (cf. IVI, MTE), continuous transitions suggest that CDL’s
alveolar and postalveolar gestures were at least partly merged.
Moreover, her continuous spectral transitions went along with
durational changes in favour of postalveolar. On the whole, we
conclude that the alveolar-to-postalveolar assimilation in French
sibilant sequences is a gradual rather than a categorical process, as
was hypothesized in (F3).
It should also be noted that CDL differed from the other French speakers in that many mean CoGs of her sibilant sequences (though not as much as suggested by Fig. 4) show an upward shift towards the alveolar reference sibilant. However, we think that CDL’s upward shift is an artefact of the extremely high speaking rate that she employed (as noted earlier). Previous studies have shown that changes in speaking rate or style have an effect on the spectral patterns of sibilant noises (cf. Maniwa studies have shown that changes in speaking rate or style have an effect on the spectral patterns of sibilant noises (cf. Maniwa et al., 2009). This hypothesis is further supported by the fact that CDL behaved like ABO and MTE in her slow, careful sentences and in her spontaneous speech data, both of which had more similar speaking rates to the other speakers.

4.2. Basic interpretation of the English results

The English sibilant sequences show two duration patterns: The second sibilant section is longer than the first, and the postalveolar section is longer than the alveolar section. Both types of sequences are longer than the single reference sibilants. The CoG means and ranges of the alveolar–postalveolar sequences are statistically indistinguishable from the postalveolar reference sibilants. These findings are clear manifestations of a strong regressive place assimilation process in line with previous studies (cf. Section 1.2) and our hypothesis (E1). There is also clear evidence that this regressive assimilation process is not symmetrical, but restricted to alveolar–postalveolar sequences, as hypothesized in (E2). Furthermore, the mean CoG distributions of the /s/ and /z/ sequences and the necessity to set ‘B’ labels in a fair portion of these sequences both imply that the regressive assimilation is overall a gradual process.

However, there are findings which do not fit in with the general pattern of postalveolar dominance. First, the alveolar sections were sometimes longer than the post alveolar sections in /s/ and /z/ sequences. In fact, in some of these sequences the postalveolar portion almost completely disappeared. The /s/ in “Spanish sea” (produced by DGO) that is shown in Fig. 12(a) is an example of such a sequence. An analogous displacement of the postalveolar portion also occurred in the spectral domain. Most cases were found for DGO and CMO. One of them is contained in the sentence fragment “Rosebush season” of Fig. 12(b). The /s/ sequence in this fragment, which also comes from DGO, is spectrally entirely comparable with the surrounding alveolar sibilants (and with the alveolar references in the same vowel contexts, cf. the individual tokens in Figs. 9 and 10). The two /s/ sequences that are displayed in Fig. 12 also sound impressionistically like alveolar sequences, i.e. [ss].

It is possible that examples like these are due to slips of the tongue. However, the speakers were instructed to repeat sentences with mispronunciations. So, the fact that the speakers did not repeat these sentences – along with the lack of hesitations or any other kinds of disfluencies – suggests that they judged their own productions as acceptable. Consequently, these findings raise doubts about the general validity of the postalveolar dominance, whereas they are still consistent with regressive place assimilation. Follow-up studies should deal with this issue in more detail.

4.3. Interpretations of the voice assimilations patterns

As regards French, the significant differences between the voiceless intervals of /z/ and /s/ on the one hand and /s/ and /s/ on the other fit in well with previous findings of regressive voice assimilation (cf. Carton, 1974; Fagyal et al., 2006; Malmberg, 1969; Snoeren et al., 2006). Unlike the place assimilation, the voice assimilation was consistent across the four speakers. It included completely voiceless or voiced sequences, but mostly a small portion of the original voice feature was kept. This is consistent with Snoeren et al. who concluded from their acoustic and perceptual findings that voice assimilation in French is a gradual process. But while Snoeren et al. found that the assimilation towards voicing was stronger than the assimilation towards voicelessness, our data suggest a reversed asymmetry. Regressive voiced-to-voiceless assimilations only left about 15% voicing in the sequences, whereas 35–40% voicelessness remained after voiceless-to-voiced assimilations.

One possible reason why our findings deviate from those of Snoeren et al. is that we examined fricatives, whereas the study of Snoeren et al. was based on stops. Fricatives and stops impede the airflow through the vocal tract and hence raise the supraglottal pressure. Thus, both kinds of obstructions are prone to become voiceless in production. However, while maintaining voicing in stops is a matter of articulatory compensation for the intrinsic breakdown of vocal fold vibration, maintaining voicing in fricatives additionally interferes with the creation of friction (cf. Ohala, 1983). Avoiding this additional interference could have biased the regressive voice assimilation behaviour in our sibilant sequences in favour of voicelessness. Supporting this preliminary explanation, we found weak but significantly positive correlations

Fig. 12. Examples of the two target-word pairs “Spanish sea” (a, left) and “Rosebush season” (b, right), produced by DGO. The /s/ sequences in both pairs are overall spectrally very similar to the adjacent single alveolar sibilants (/s/,/z/). Only in (a), there is a discontinuity in the sequence that may be regarded as reflecting a short initial postalveolar section.

between the relative voiceless intervals and the overall durations of the two sibilant sequences \(/\text{zf}/\) \((r=0.46;\, \text{df}=47;\, p<0.01)\) and \(/\text{s}/\) \((r=0.55;\, \text{df}=47;\, p<0.001)\). That is, the longer the sibilant sequences, the stronger the voiced-to-voiceless assimilation.

Crucially, a comparison of the two alveolar–postalveolar sequences with the two postalveolar–alveolar sequences yielded no significant difference in the voiceless interval. Thus even though we found clear evidence that French sibilants can interact simultaneously in terms of their voice and place features, there is no evidence that the voice-related process is affected by the place-related process. This outcome is in accord with hypothesis (F2). Assimilation of place of articulation in French sibilant sequences is not triggered by voice assimilation, but occurs independently of the latter. Moreover, the two processes can even go into opposite directions. Progressive place assimilation can co-occur with regressive voice assimilation.

As for the English sibilant sequences, the different voicing patterns in the \(/\text{zf}/\) and \(/\text{sf}/\) sequences of CMO may be interpreted in accord with hypothesis (E3). The \(/\text{zf}/\) sequences became voiceless due to a process of regressive voiced-to-voiceless assimilation. The \(/\text{sf}/\) sequences were not subject of an analogous voiceless-to-voice assimilation and showed thus similarly long portions of voiceless and voiced frication. Hence, we found supporting evidence that voice assimilation can occur in English (fricatives) across word boundaries and beyond function words. From this point of view English is more similar to French than previously thought. However, unlike the regressive process in French, the English process is asymmetrical insofar as it only proceeds from voiced to voiceless. Moreover, compared with the consistent voice assimilation behaviour of the French speakers, English voice assimilation seems to be a more speaker-specific or dialect-specific process (cf. Wells (1982)) for voice assimilations in dialects of Yorkshire, where our speakers came from).

When comparing the voice assimilation processes of English and French, it must also not be forgotten that the phonological voice contrast in English relies on a large bundle of phonetic parameters (cf. Lisker, 1986). Therefore the phonological contrast is also represented by the features ‘lenis’ versus ‘fortis’. In contrast, the phonological contrast in French is more closely related to the actual presence or absence of vocal-fold vibration (cf. Snoeren et al., 2006). That the phonological contrast is based on a richer phonetic coding in English than in French is interesting from two points of view. First, in the case of the English speakers, the difference in the phonation behaviour of CMO on the one hand and DGO, DGO, and FDU on the other could proceed from two points of view. First, in the case of the English dialects of Yorkshire, where our speakers came from).

4.4. General discussion

4.4.1. Comparing French and English

The present study demonstrated that assimilation of place of articulation exists in French as well as English. Furthermore, the assimilation processes in the two languages show fundamental differences in both direction and target quality. The French sibilant assimilation is clearly determined by the target quality. That is, the functional and spectral changes are all in favour of the postalveolar sibilant. Direction does not appear to play a major role in French. However, the changes were stronger and less

![Diagram 13](image.png)

Fig. 13. Schematic illustrations of spectral patterns found for alveolar–postalveolar sequences in French and English, based on the types differentiated by Holst and Nolan (1995). Type A and B require a ‘B’ label, type C and D not.

speaker-specific in the regressive than the progressive direction. In contrast, in English, direction is the central characteristic. Assimilation is strictly regressive and dominated by the postalveolar target quality. But, in view of the examples of assimilations towards alveolar, the target quality is not as important and restricting as in French. So, from a simplified point of view, the assimilations of place of articulations within the sibilant sequences of French and English may be described as primarily quality-guided or primarily direction-guided, respectively. This comparative statement may be regarded as the quintessence of the discussion of the hypotheses.

Moreover, the assimilation processes of both French and English must be characterized as gradual. The ranges of regressive and progressive assimilations in French as well as the range of regressive assimilations in English cover all four types of patterns in Fig. 13 (with as well as across speakers) based on the type A–D distinctions described by Holst and Nolan (1995). These schematically represent the acoustic outcomes that are created when a postalveolar sibilant gesture dominates the adjacent alveolar gesture to different degrees. For example, type A results when the two gestures are simply co-existing, representing non-assimilation. At the opposite end, type D shows up when the postalveolar gesture fully dominates the alveolar one, representing complete assimilation. Types B and C are intermediate dominations.

Although all four assimilation types occurred in French and English, we can show by means of the ‘B’ labels that the types are differently distributed in the two languages. We only set ‘B’ labels in those sibilant sequences whose CoG ranges were larger than the maximum range that occurred within the single sibilants. Therefore, the ‘B’ labels represent an indicator of the spectral stability of a sibilant sequence. We classified each of the sibilant sequences into A, B, C or D in the following way. Sequences without a ‘B’ label had a constant spectral quality and were thus classified as instances of type D assimilations; ‘B’ labels placed in the middle of a sibilant sequence (between 45% and 55% of the total sibilant duration) indicated a type A assimilation. Sequences whose ‘B’ labels were strongly displaced towards the beginning or end of the sibilant sequence were classified as type B/C.

According to this tripartite classification, more than 60% of the regressive alveolar-to-postalveolar assimilations in English were of type D. In contrast, in French only 25% of the regressive assimilations were of type D. Around 15% of the alveolar–postalveolar sequences in French were of the non-assimilation type A, as opposed to less than 5% in English. A Chi-square test found the frequencies of the three assimilation patterns (A vs. B–C vs. D) in the alveolar–postalveolar sequences to be different for French and English (Chi square = 28.02, df = 2; \(p<0.0001\)). Taking into consideration the CoG results for French and English, it is reasonable to conclude that the regressive assimilation was stronger for the English than the French
speakers. However, this need not be true for all French speakers since speakers of Northern varieties may show even higher rates of word-final /a/ elision and thus might show regressive assimilation that is as strong as for our English speakers.

4.4.2. Explaining the cross-linguistic differences

How can we explain the cross-linguistic differences that concern both the degree and the direction of the assimilation? The Hypo-Hyper (H&H) theory of Lindblom (cf. Lindblom, 1990), postulates that the phonetics of an utterance is the result of a balance between two antagonistic forces: maximizing production efficiency (i.e. the utterance should be as articulatorily reduced and simplified as possible, which includes assimilation) and minimizing listener's error (which depends on the cognitive mechanisms and processes in speech perception). The assimilationary differences between French and English may be approached in terms of the H&H balance, viewed in light of language-specific, linguistic factors.

Firstly, closed syllables are much more common in English than in French. According to written-text analyses, 76% of the French syllables are CV or V, whereas 60% of English syllables are CVC or CVCC (Delattre & Olsen, 1969; Goldman, Content, & Frauenfelder, 1996). In addition, speech may be processed on the basis of syllables in French and phonemes in English (cf. Cutler, Mehler, Norris, & Segui, 1986), which potentially increases the difference in processing of adjacent consonants. Secondly, sibilants make up approximately 12.5% of the consonants in English and only 8.5% in French (Adda-Decker, 2006; Fry, 1947). Thus, sequences of sibilants are more likely to occur in English, and more frequent speech patterns are reduced more strongly and more consistently (e.g. Ernestus, Lahey, Verhees, & Baayen, 2006; Kohler, 1990). Thirdly, French and English are thought to differ in speech rhythm. English rhythm may be stress-timed (cf. Barry et al., 2007). This means that there is a stronger temporal compression between perceptually salient syllables which may provide a higher incentive for speakers to reduce their speech. Overall, these facts may explain why the regressive place assimilation in the English sibilant sequences was more often complete across speakers than in French.

To explain the directional differences between the French and English assimilations, we must also consider the perception side of the production-perception trade off. Perception is the probability-guided interpretation of incoming stimuli (cf. Goldstein, 1989), and probabilistic concepts have been used in recent studies to model the decoding of assimilated speech as well as to explain the differential use of acoustic cues (cf. Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Gaskell, 2003). Given this, we searched for reasons why French does and English does not show progressive assimilation in language-specific probabilities of (co-)occurrences of alveolar and postalveolar sibilants across word boundaries, based on corpora of phonologically transcribed, canonical forms. In the case of French, we decided to include all relevant tokens with word-final schwa in the census which may have slightly overestimated the potential for assimilation.

Sequences of two postalveolar sibilants /s/z/, /s/sl, /s/sl, /s/sl turned out to be highly infrequent in French. Estimates derived from texts of about 10 million words in the newspaper ‘Le Monde’ found that these sequences occur with a probability of less than 2% (cf. van Rullen et al., 2005). Similar values result from corpora of spoken French (cf. Bertrand et al., 2007; New et al., 2007). Thus, French listeners coming across a sequence of two postalveolar sibilants can be almost certain that this sequence results from place assimilation. Furthermore, three out of the four sibilant sequences which can show progressive assimilation, viz. /z/z/, /s/sl, and /s/sl, are relatively rare across word boundaries (cf. van Rullen et al., 2005; New et al., 2007). The only more frequent sequence is /s/sl, which may be assumed to consist to a large extent of “je sais” and “je suis” cases after schwa elision. So, when French listeners come across postalveolar sibilant sequences, they may expect that these are instances of regressive assimilations. The relatively few cases of progressive assimilations (particularly “je sais” and “je suis”) could be represented cognitively as separate exemplars, or they could be identified with little extra effort on the basis of the semantic and/or the syntactic context.

Analysis of the British National Corpus (BNC, Leech et al., 2001) reveals that the same is not true for English. In English, words ending in postalveolar sibilants (i.e. /f/) are at least three times more frequent than in French (approx. 5%), and words that start with alveolar sibilants (particularly /s/) are also quite common. Hence, progressive assimilations in English sibilant sequences would affect far more combinations of words than in French and in combination with regressive assimilation would create many lexical ambiguities (cf. also the BNC-related online data base on ‘Phrases In English’, PIE, http://pie.usna.edu). In summary, unlike English, French can afford to allow more comprehensive reductions and hence more efficient speech productions in terms of regressive and progressive assimilations in sibilant sequences, since this does not entail much extra cognitive effort and/or a considerably higher risk of ambiguities or misunderstandings at the perceptual side of the speech chain.

4.4.3. Effects of word frequency, vowel context, and voice

In English, higher frequency of the second word meant higher mean CoGs and larger CoG ranges. We think that this was due to the fact that in the majority of these cases, both target words were realized with clear pitch accents, which are known to lower the degree of speech reduction (cf. de Jong, 1995) and that most were postalveolar–alveolar sibilant sequences which showed less assimilation in general. We also found larger CoG ranges for alveolar and for postalveolar sibilants in both French and English, which indicates that /s/ is in general more susceptible to context effects than /f/ (cf. Dart, 1998; Gordon et al., 2002; Jongman et al., 2000; Narrey, 1982).

We also found vowel-related CoG and temporal effects in French and English. As these effects occurred for /u, i/ vs. /i, a/ and /a, a/, they are likely due to lip rounding, which is known to spread to adjacent segments (Bell-Berti & Harris, 1982). However, we think the effect of vowel duration on the temporal structure of the sibilant sequences in French has a different origin. The duration of the second sibilant increased with an increasing duration of the following vowel. We expect vowel lengthening due to accent to affect the syllable onsets but not the preceding coda (Hirst et al., 1998). On the whole, it is interesting to see that all of our vowel-context effects concerned the following vowel in French, but the preceding vowel in English. This supports the claim of Hoole et al. (1993), according to which anticipatory co-articulation is characteristic for French, whereas English is dominated by carry-over co-articulation.

The lower mean CoGs that were found for voiced single sibilants and sibilant sequences are likely to have a twofold origin: First, the vibration of the vocal folds slows down the supraglottal airflow, which in turn reduces the acoustic energy of the friction, particularly for higher frequencies. Secondly, the voice source introduces energy in the lower frequency region of the spectrum.

4.4.4. Implications for the modelling of assimilation

The place assimilation patterns of both French and English cover a continuum from non-assimilated to completely assimilated sibilant sequences. At a phonetic level, our temporally and
spectrum of intermediate patterns, particularly the continuously changing CoG time courses in Figs. 6 and 10, contradict the notion of acoustic quantal changes in the production of alveolar and postalveolar sibilants. Rather, our data favour a close correspondence between gradual articulatory and acoustic changes in sibilant production, as was previously argued by Nolan et al. (1996) and Tabain (2001).

At the level of the speech code itself, the variety of temporally and spectrally intermediate sibilant patterns of the present study is incompatible with the view of assimilation as a phonological process which takes place at a cognitive level prior to the actual speech production, and in which a feature of a certain segment is or is not replaced by the feature of another segment (cf. Section 1.1). Such a process entails categorically different phonetic patterns. Although the English speakers come closer to categorical patterns than the French speakers, they are still far from this simple behaviour. Several approaches have attempted to fill this gap between empiricism and theory.

For example, Holst and Nolan (1995, p. 330) proposed that in general “an apparent continuum of assimilation involves two separate phenomena”. While the non-assimilated and completely assimilated extremes are manifestations of a categorical phonological process, the spectrum of intermediate forms is caused by “mechanical effects” (Holst and Nolan, 1995, p. 330) that occur during articulatory implementation. This proposal is not supported by the results of the present study. In particular, two aspects pose a problem. First, the present study revealed that even those sibilant sequences that were not completely assimilated differed systematically within as well as across French and English. For example, we showed that the second element in spectrally discontinuous sequences tended to be longer in English than in French. This is not predicted by a mechanical process. Secondly, we found that spectrally constant type-D patterns can sometimes have friction qualities which are in between those of the two alveolar and postalveolar reference sibilants. This is not predicted by a categorical process.

The framework of articulatory phonology (cf. Browman and Goldstein, 1992) might account for a range of assimilation strengths in terms of different degrees of temporal overlap of two adjacent articulatory (i.e. alveolar and postalveolar) gestures. However, this too is problematic. For example, if a gesture is successively merged with an adjacent gesture, the overall temporal pattern of the resulting merger should be symmetrical. However, this is clearly inconsistent with the asymmetrical shifts in our ‘B’-label based durations. Another important point is that if the two gestures overlap more strongly, i.e. if the degree of assimilation is higher, the overall duration of the sibilant sequence should decrease. In fact, Browman (1995) re-analysed the data of Holst and Nolan (1995) on alveolar-to-postalveolar place assimilation in English sibilant sequences and found evidence for such a correlation. We correlated our CoG mean and range measures with the corresponding overall durations of the sibilant sequences. Although the coefficient was higher for English ($r=0.247$) than for French ($r=0.013$), both correlations are negligible. The degree of assimilation in a sibilant sequence was not reflected systematically in its overall duration. Furthermore, Nolan et al. (1996) noted that their data also contradicted the concept of complete gestural overlap insofar as all sibilant sequences, including type D, were longer than a single sibilant. Our own data is consistent with Nolan et al. in this respect and in fact our sequences were not just 16% but more than 30% longer than the reference sibilants. Thus, in order to model our French and English findings within the framework of articulatory phonology, it seems necessary to include changes in the temporal extensions of the involved gestures, in addition to changes in their temporal overlap.

In summary, it can be concluded that none of the above frameworks is able to account for the complete spectrum of observed time and frequency patterns in French and English in a straightforward way. Hence, further improvements are necessary. Instead of trying to understand assimilation as a contrast-neutralizing process in order to facilitate speech production, assimilation should be addressed from a perspective that also takes the listener and communicative functions into account.

5. Outlook

Based on sibilant sequences, our comparative study provided the first systematic acoustic evidence for the existence of assimilation of place of articulation in French. Place assimilations occurred independently of voice assimilations, but the two assimilations can coincide and even go into different directions. One obvious task of follow-up studies on French is to examine, if and in which ways assimilations of place of articulation also occur for other types of consonants. For example, with reference to single-case examples it is claimed by Kohler (2002, p. 19) that “if an apical nasal precedes [a labial plosive], it may be shifted to labial place of articulation, even if the plosive is no longer produced: une petite [ýmit]”. Supporting Kohler’s claim, we found already many instances of nasal place assimilations, including [ým(þ)ti(t)], in the French Corpus of Interaction Data. These instances are a good starting point for detailed phonetic analyses. Against the backdrop of our English findings, it should be further investigated to what extent, under which conditions, and for which types of consonants the primarily direction-guided regressive place assimilations can induce changes towards the alveolar place of articulation. In addition, the voice assimilation patterns in English and their possible interactions with place assimilations are worth being addressed more systematically and with a larger group of speakers.

In dealing with the degree of assimilation, future studies should integrate temporal and spectral data and extend the perspective beyond the pair of assimilating and assimilated consonants. For example, in labelling and analysing the present data we made the initial observation that the vowels preceding alveolar and postalveolar sibilants can have different durations, energies, and voice qualities, and that these differences can persist irrespective of the occurrence and degree of place assimilation.

Acknowledgements

We are particularly grateful to Gareth Gaskell, Noel Nguyen, Mirjam Ernestus, Ulrich Frauenfelder, Sarah Hawkins, Sophie Dufour, and many other people from the Marie Curie Research Training Network “Sound to Sense” for their fruitful and inspiring discussions in the course of the planning, performance, and interpretation of our study. Moreover, we would like to thank our two anonymous reviewers and the editor, Stefan Frisch, whose constructive and helpful comments contributed a lot to improving this paper. Finally, thanks are due to the Marie Curie Research training network MRTN-CT-2006-035561-S25 (“Sound to Sense”) that funded the present study.

Appendix A

The French sentence list, subdivided into the three subsets as shown in Table A1.
The primary subset
Toi tu triches si j’yeux
C’est la Miss chignon
Tu t’ouche sous l’idrap
C’est une rousse chouchou
Elle remâche sa viande
C’est une classe chargée
C’est un vestige Zimbabweenne
C’est une devise gitanes
C’est un rouge zoulou
J’ai vendu douze journaux
C’est un mariage zairois
C’est une phrase japonaise
Ils ont trouvé la biche zigouillée
Ils annoncent une crise chinoise
C’est un prestige si convoité
Si c’est à Nice Ty vais
J’ai rencontré de farouches zoulous
Il y a douze chouchous
Il bouge souvent
C’est une douce journée
Cette fumée cache Zagreb
C’est une danoise charmante
Il part en voyage samedi
Il ne chasse jamais
The secondary subset
Il se niche sous l’toit
C’est la Miss chouchou
Tu t’ouche si j’yeux
C’est la houesse chinoise
Tu t’ouche sous l’lit
C’est un passe choué
Une grosse mouthe ca pique
C’est une trouse chargée
J’étais sur la corniche samedi
C’est un fils charmant
Tu t’ouche si mal
C’est une tasse chinoise
C’est un vrai prodige zoulou
Nous passons une esqué journée
Ce peau-rouge zigouille ses ennemis
Il y avait douze gitanes
J’étais invité à un mariage zoulou
C’est une bourgeoise joifille
C’est un peau-rouge zairois
C’est une andalousie jalouse
J’ai vu les vestiges zairois
Je ne le précise jamais
On le considère comme un sape zigoto
C’est un vase gitan
The complementary subset
Rejoins-moi sur la place face au port
Tu t’ouche facilement
C’est une chinoise particulière
Ce message va à la mer
C’est un service financier
C’est une belle poudre finnoise
Il a une devise vitale
C’est un vestige biblique
Il a fait un couscous pour toi
Il achète des cartouches pour son arme
Il regarde sa peluche jouet
Il voit rouge pour tout
Elle tape sa sœur
C’est une sape charmante
C’est un pape zairois
Il ne la haffe jamais
Il se rebiffe si souvent
C’est une équipe chinoise
C’est une initiative Zimbabweenne
Si elle s’active je vais
Il bouffe sous les ponts
C’est un couche-chou-fleur
C’est un groupe zoulou
C’est une soupe journalière

References

Appendix B
The English sentence list, subdivided into the three subsets as shown in Table B1.

Table B1
The primary subset
They crossed the Spanish sea
He will steer this ship
He wanted to push Susan
He entered the house shoeless
The dog needs a brush sometimes
It’s great how the glass shines
They have shorn these sheep
For the cake you may use sugar
He broke Anna’s shower
She has a childish zeal
They want to ambush Zulus
In the football match, France tried to crush Zaire

The secondary subset
He buys fish soup
The girl has nice shoes
In May, it’s rosebush season
The cold wind made me shiver
hearts was the flush suit
He is a first-class shoemaker
The diplomat found Bush sarcastic
The gardener fixed the mouse shaver
She likes the British south
Her friends miss Charlotte
The spy tried to rush secretly
This is a brass shield

The complementary subset
The students went to the blue-grass party
The driver received a harsh fine
This is Barbara’s mouse
Do not try to kiss me
This is a fish filet
He can touch his feet
The park keeper found the goose food
The negotiations made Bush moody
The contestant was careful not to choose foolishly
I’ve had enough sardines
The bay was known for tough sharks
They planned to halve Zagreb
The team arrived at the crime scene
We have visited a steam ship
I bought a cheap ripper
We have to cut the bloom soon
During work, he has to wear fireproof shoes
The camera has a childproof zoom


