Original Paper

The Phonetic Manifestation of French /s#ʃ/ and /ʃ#s/ Sequences in Different Vowel Contexts: On the Occurrence and the Domain of Sibilant Assimilation

Oliver Niebuhr[a]   Christine Meunier[b]

[a]Seminar für Allgemeine und Vergleichende Sprachwissenschaft, Christian-Albrechts-Universität zu Kiel, Kiel, Germany; [b]Laboratoire Parole et Langage, UMR 6057 CNRS, Université de Provence, Aix-en-Provence, France

Abstract

While assimilation was initially regarded as a categorical replacement of phonemes or phonological features, subsequent detailed phonetic analyses showed that assimilation actually generates a wide spectrum of intermediate forms in terms of speech timing and spectrum. However, the focus of these analyses predominantly remained on the assimilated speech sound. In the present study we go one step ahead in two ways. First, we look at acoustic phonetic detail that differs in the French vowels /i, a, u/ preceding single /s/ and /ʃ/ sibilants as well as /s#ʃ/ and /ʃ#s/ sibilant sequences. Second, our vowel measurements include not only F1 and F2 frequencies, but also traditional prosodic parameters like duration, intensity and voice quality. The vowels and sibilants were recorded as the central part of CVC#CVC pseudo-names in a contextualized read-speech paradigm. In the single-sibilant conditions we found that the vowels preceding /ʃ/ were longer, breathier, less intense, and had more cardinal F2 values than before /s/. For the /s#ʃ/ and /ʃ#s/ conditions we found regressive and progressive /s/-to-[ʃ] assimilation that was complete in terms of spectral centre-of-gravity measurements, although French is said to have only voice assimilation. Moreover, the vowels preceding the /s#ʃ/ sequences still bear an imprint of /s/ despite the assimilation towards [ʃʃ]. We discuss the implications of these findings for the time window and the completeness of assimilation as well as for the basic units in speech communication.

Copyright © 2011 S. Karger AG, Basel

1. Introduction

1.1. Research Background

Speakers of languages in the 'alphabetic world' have traditionally held the view that speech consists of a string of segments, that each segment is concatenated with

its neighbours by short transition phases, and that a stationary section between these on-glides and off-glides conveys the characteristic sound features of the segment. This linear segmental concept dominated descriptive phonetic research on speech production well into the 20th century [Sweet, 1906; Jespersen, 1926]. However, with instrumental phonetic analysis, Menzerath and de Lacerda [1933] were able to show that (a) the articulators start to move well before an acoustic sound is produced and that (b) the movements of the individual articulators towards and away from the expected segmental configurations strongly overlap without pausing for a clear stationary section.

Since the seminal work of Menzerath and de Lacerda [1933] the segmental concept has been shifted across the speech chain from the level of production to the levels of speech acoustics and perception [Morton et al. 1976; Liberman and Mattingly, 1985; Stevens, 1989]. However, due to methodological advances and the progression from read words and sentences to spontaneous dialogue data, the segmental concept has been continuously undermined at all levels of the speech chain. In fact, there can be no doubt today that the segment is an oversimplified concept. When we speak, our articulators are in constant motion, and these non-linear changes are reflected in the acoustic signal and in the perceptual behaviour of the listener. For example, rhoticity and secondary articulations like velarization, palatalization and labialization can be present in the acoustic signal several syllables before the speech segment they are ascribed to occurs [Bell-Berti and Harris, 1982; Kelly and Local, 1986; Tunley, 1999; Heid and Hawkins, 2000], and listeners make use of these long-term resonances in identifying words (particularly in noise) and in restoring masked speech sections [Benguerel and Adelman, 1976; Hawkins, 1995; West, 1999, 2000]. Moreover, in producing reduced speech (which is typical of spontaneous speech and of function words in particular), speakers can reorganize the sequencing of vocal-tract opening-closing movements in the articulation of words by superimposing articulatory prosodies on the surrounding opening-closing gestures, thus retaining the phonetic essence of reduced words in a non-segmental fashion. Listeners, in turn, can use the articulatory prosodies to perceive the word [Kohler, 1990; Kohler and Niebuhr, 2011; Niebuhr and Kohler, 2011]. Finally, there is evidence that when listeners restore a reduced word, this restoration process is not based on segments, but on suprasegmental units [Niebuhr, 2011].

Yet, despite the obvious evidence against the speech segment and its abstract phonological counterpart, the phoneme, the segmental concept still shapes our approaches to speech. In certain respects and under particular conditions, speech production and perception can of course be at least approximately segmental; and as a heuristic tool, for instance in transcriptions with the International Phonetic Alphabet [IPA, 1999], a segmental approach can be informative. But when it comes to investigating the interaction between sounds, as in cases that are subsumed under the heading of assimilation, a strictly segmental approach applies too narrow a time window to phonetic events and thus hinders insight into how speaking and understanding work in speech communication. When in a postulated segmental string XYZ, Y is classified as becoming assimilated to Z, the question of a potential imprint of Y on X is left unanswered.

This is the question that underlies our present study, which deals with alveolar-to-postalveolar assimilation in French sibilant sequences (as this assimilation should primarily concern the shape of the tongue rather than the place where it creates the constriction, we will refrain from using the term 'place assimilation'). We will show that alveolar and postalveolar sibilant sounds create systematically different imprints in the preceding vowels, and that these multiparametric phonetic details remain, even if

the alveolar sibilant noise itself is assimilated and becomes postalveolar. The nature of the sibilants' imprints in the preceding vowels as well as the presence of assimilation-independent vowel differences fit in well with a number of previous studies that are summarized in the following.

1.2. Previous Studies

1.2.1. Assimilation and the Issues of Completeness and Scope

Traditionally, assimilation is conceptualized as a process in which a sound takes over a phonological feature of a preceding or following sound. In other words, assimilation is claimed to change one sound into another in a categorical way [cf. Chomsky and Halle, 1968]. However, in more detailed phonetic analyses, like in the studies by Nolan [1992] and Holst and Nolan [1995], it was found that regressive assimilations of alveolar sounds across word boundaries in sequences of English plosives or sibilants are gradual rather than categorical processes that generate a broad array of intermediate forms in terms of both speech timing and spectrum. For example, even if a /s#ʃ/ sequence as in *Swiss shop* is realized with regressive sibilant assimilation, a residual of the /s/ can remain in a subtle initial transition in the phonetic quality of the friction. Thus, the assimilation is phonetically partial in the temporal and spectral domains. Alternatively, if the friction of the assimilated sequence does not change over time, it can have a phonetic quality in between [s] and [ʃ]. Such assimilations have been classified as complete in the temporal domain, but partial in the spectral domain. Since the influential work of Nolan [1992] and Holst and Nolan [1995], incomplete temporal and/or spectral assimilation patterns across word boundaries have been found for various assimilatory processes in many different languages, including German and French [Zsiga, 1995; Gow, 2001, 2002, 2003; Ellis and Hardcastle, 2002; Dilley and Pitt, 2007; Snoeren et al., 2008; Kuzla, 2009; Pouplier et al., 2011].

Despite the fact that assimilation underwent a conceptual change from a categorical to a gradual process, many of the above studies [including Holst and Nolan, 1995] continue to claim that complete assimilations across word boundaries do still occur at the extreme end of the assimilation continuum. These claims rest on articulatory or acoustic measurements that were made *within* the C#C dyad of assimilating and assimilated *segment*.

However, as was outlined in 1.1, consonants, including [s] and [ʃ], do not start with the onset of a non-vocoid sound quality like sibilant friction. The constriction, the corresponding tongue shape as well as the interruption of voicing are prepared in the preceding vowel. This kind of overlap, by which sound Y creates an imprint in sound X, has been studied extensively for many different sound sequences and languages and under various headings like coarticulation, articulatory control or segment-induced modifications. It has also been the subject of theoretical discussion [Barry and Hawkins, 1992; Browman and Goldstein, 1992; Hardcastle and Hewlett, 1999; Kohler and Niebuhr, 2011]. For example, in the latter paper, Kohler and Niebuhr [2011] argue on the basis of the theory of phrase-level phonetics that the operational units in the controlled interactions of speech sounds are whole opening-closing gestures, inspired by those suggested by Menzerath and de Lacerda [1933] and Öhman [1966]. From this theoretical standpoint, describing and classifying assimilations within a segmental grid is insufficient in two ways. First, the assimilated sound is not

an isolated segment, but an integral part of a vowel-consonant gesture VC#. Thus, phonetic aspects of the assimilated sound are also contained in the vowel portion of the VC# gesture. Second, assimilation is not encapsulated in a single unit, but consists of phonatory *and* articulatory components that shape the entire VC# gesture. Their involvement in assimilation can vary a great deal. For example, the vowel portion of the assimilated sound may be less strongly affected by assimilation than the consonant portion, or the phonatory component may be less strongly affected than the articulatory component. So far, assimilation research has often disregarded or underestimated these possibilities.

Initial evidence in favour of phonetic residuals of the assimilated sound in the preceding vowel portion comes from Nolan [1992]. He showed that listeners were able to distinguish those words ending in actual bilabial or velar sounds from those words whose final alveolar sounds were analysed on an articulatory basis as completely assimilated towards velar or bilabial sounds (e.g., *lead* vs. *leg* preceding *covered*). Nolan [1992] relates this ability of the listeners to his impressionistic observation 'that the vowel allophone before the lexical velar is slightly closer than before the lexical alveolar' [Nolan, 1992, p. 272]. This impression was later confirmed through acoustic measurement by Local [2003]. Similar evidence was provided by Gow [2001, 2002], who showed in terms of acoustic formant transitions that phonetic aspects of word-final alveolar plosives and nasals in English remain in their preceding vowel portions when the alveolar sound has been assimilated to the subsequent labial sound (e.g., assimilated *cat box* differs from *cap box*). Moreover, Gow [2001, 2002] argued that listeners can use these vowel details not only to identify the assimilated word, but also to predict the initial sound of the following word [a similar anticipatory perception was demonstrated in Kohler and Niebuhr, 2011].

1.2.2. French Sibilant Assimilation and Sibilant-Related Vowel Details

In recent phonetic studies Niebuhr et al. [2008, 2011] analyzed the acoustic realizations of sibilant sequences across word boundaries in French. In contrast to the general claim that French only knows voice assimilation [Ramus, 2001; Fagyal et al., 2006; Snoeren et al., 2008], the acoustic data of Niebuhr et al. [2008, 2011] clearly show regressive alveolar-to-postalveolar assimilations in sequences like /s#ʃ/. Parallel to the findings of Holst and Nolan [1995] for English sibilants (cf. 1.2.1), the French sibilant assimilations were overall gradual in both timing and spectrum, and they also included apparently complete assimilations. That is, the /s#ʃ/ sequences had substantially longer overall durations than single /ʃ/ sibilants, which argues against /s/ elision in the sequences. Yet, the spectral distribution of the noise energy, represented by spectral centre-of-gravity (CoG) measurements, was statistically the same in single /ʃ/ and assimilated /s#ʃ/ sequences.

Moreover, unlike English, French showed alveolar-to-postalveolar assimilations even in the progressive direction, i.e. in sibilant sequences like /ʃ#s/. The progressive assimilation was weaker than the regressive assimilation. However, the progressive process also yielded sequences that may be called completely assimilated on the basis of the temporal and spectral acoustic measurements within the sibilant noise section. Figure 1a presents spectral analyses of the two French utterances *C'est une trou<u>sse</u> <u>chargée</u>* ('This is a full bag') and *Tu te couches <u>si</u> j'veux* ('You go to bed when I want') that were elicited by Niebuhr et al. [2008, 2011] (sibilant sequences underlined). In view of the comparable noise patterns of sibilant sequences in the two spectrograms,
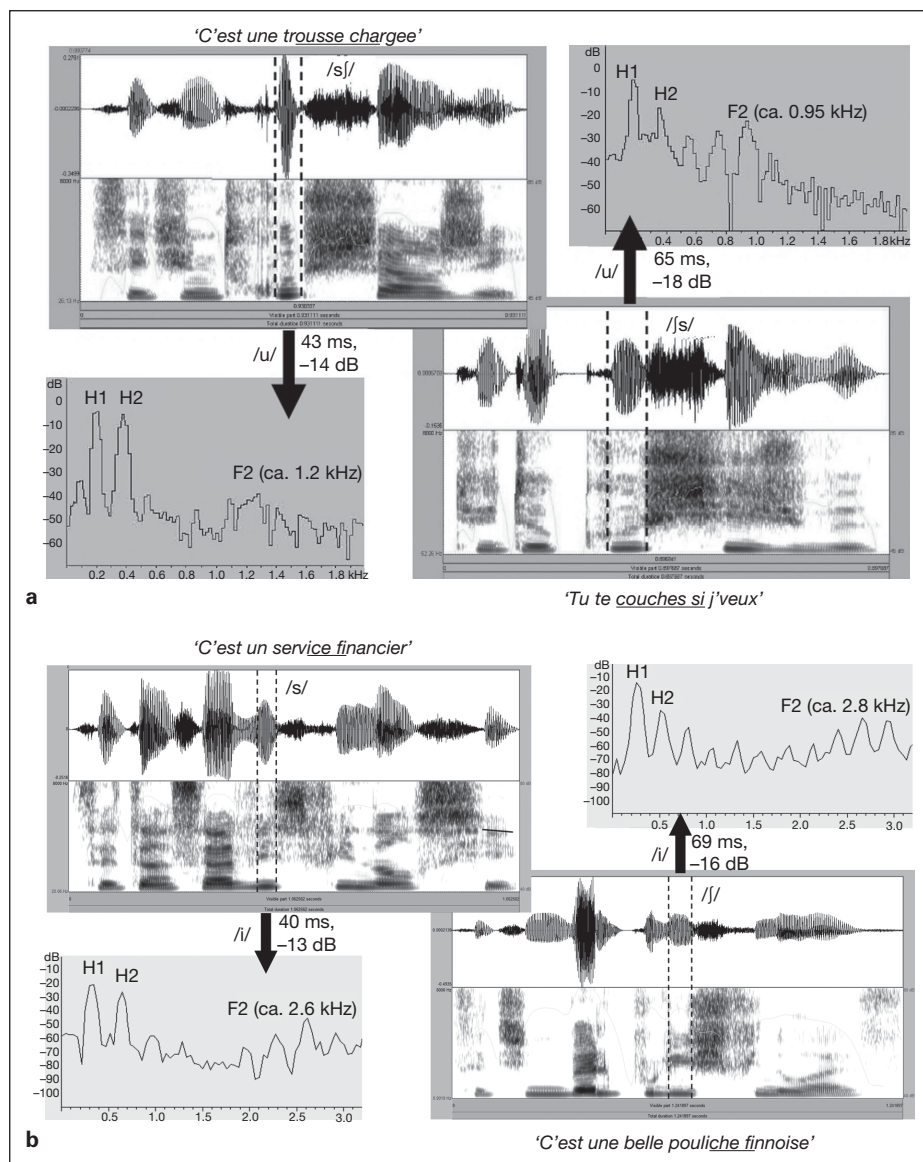
Fig. 1. Acoustic analyses of two pairs of French utterances. The pairs in **a** contain the sibilant sequences /s#ʃ/ (left) and /ʃ#s/ (right) after /u/ in *C'est une trousse chargée* (left) and *Tu te couches si j'veux*. The pairs in **b** contain the single sibilants /s/ and /ʃ/ after /i/ in *C'est un service financier* and *C'est une belle pouliche finnoise*. For all utterances oscillograms, spectrograms, and spectral slices (narrow-band DFT) are given, the slices come from the vowel centre and show the frequencies of the first two harmonics (H1, H2) and the second formant (F2). All utterances come from female speakers.

they were realized with equally strong alveolar-to-postalveolar assimilations in the regressive or progressive direction, respectively.

In the course of their assimilation analyses, Niebuhr et al. [2008, 2011] made an additional, informal observation. The vowels that preceded alveolar-postalveolar sequences seemed to differ from those preceding postalveolar-alveolar sequences. The observed differences were even multiparametric, i.e. they included more than just formant patterns that are typically at the centre of attention when relationships between sounds are addressed in the traditional segmental approach. Specifically, in the context of postalveolar-alveolar sibilant sequences, the preceding vowel portions were longer, breathier, less intense, and had a quality that was closer to the respective cardinal vowel (particularly in terms of the second formant, F2). These differences can also be seen in the two /u/ that precede the sibilant sequences in figure 1a. In order to illustrate the differences in vowel quality and breathiness, spectral slices taken at the centre of the vowels show the F2 frequencies and the amplitude levels of the first and second harmonic (H1, H2). The F2 value is lower, i.e. more [u]-like, for the vowel preceding /ʃ#s/, and compared with the vowel preceding /s#ʃ/, H1 of the vowel preceding /ʃ#s/ is clearly higher than H2, which is an indicator of a breathier voice quality [Klatt and Klatt, 1990].

Analogously, Niebuhr et al. [2011] observed that vowel sounds were longer, breathier, less intense and had a more cardinal quality before single postalveolar than before single alveolar sibilants. This analogous observation is exemplified in figure 1b by means of the vowel /i/ that precedes /s/ or /ʃ/ in the two utterances *C'est un serv<u>ice</u> financier* ('This is a financial service') and *C'est une belle poul<u>iche</u> finnoise* ('This is a nice Finnish filly'). Sound examples, including the utterance pairs of figure 1a, b, are provided in Niebuhr et al. [2009].

In their small comparative study of the two English question utterances *Who sharpened the meat cleaver?* and *Who's sharpened the meat cleaver?* Hawkins and Smith [2001] also observed differences in a vowel sound before alveolar and postalveolar sibilants. These differences concern again /u/ and fit in well with the observations of Niebuhr et al. [2008, 2011] (fig. 1a). That is, when directly followed by a postalveolar sibilant, as in the case of *Who sharpened*, the /u/ had a considerably lower F2 frequency than in the case of *Who's sharpened*, in which the /u/ is preceded first by an alveolar and then by a postalveolar sibilant. That is, the /u/ quality in *Who sharpened* was closer to the cardinal vowel quality than the /u/ in *Who's sharpened*. Moreover, the spectral analyses provided by Hawkins and Smith [2001] suggest that the /u/ preceding the postalveolar sibilant was longer, less intense, and breathier.

But the most striking fact was that the multiparametric vowel difference was there, even though the sibilant noises in *Who sharpened and* W*ho's sharpened* showed comparable sound qualities. That is, the alveolar sibilant in the sequence of *Who's sharpened* was assimilated to the following postalveolar sibilant. Likewise, the French data of Niebuhr et al. [2011] include sibilant sequences in which the vowel characteristics are not those observed before single /ʃ/, but those observed before single /s/, even though the corresponding /s#ʃ/ sequences contained only small or no spectral residuals of the sequence-initial /s/. For example, as is illustrated in figure 1a, the two /s#ʃ/ and /ʃ#s/ sequences are virtually identical in terms of spectral CoG and duration (as well as overall intensity). They could both be characterized as [ʃʃ]. Yet, the vowels preceding the two [ʃʃ] sequences differ, and these differences correspond to those found in vowels preceding single /s/ and /ʃ/ in figure 1b.

In summary, the previous study of Niebuhr et al. [2011] demonstrated that sibilant sequences can show alveolar-to-postalveolar assimilation not only in English, but also in French. In the wider context of this assimilation, informal observations in French and English suggest in line with the theoretical and empirical pictures outlined in 1.2.1 that there are multiparametric vowel differences that occur systematically and that are due to phonetic aspects of the sibilant that immediately follows the vowel. On this basis, the following four hypotheses are put forward in this study and will be tested in the first step for French.

Hypothesis 1: Replicating previous findings, the French sibilant sequences will be produced with clear regressive and progressive alveolar-to-postalveolar assimilation.

Hypothesis 2: Similar differences will be found between the vowels preceding the sibilant sequences /s#ʃ/ and /ʃ#s/, and the vowels preceding the single sibilants /s/ and /ʃ/.

Hypothesis 3: The vowel differences go beyond formant patterns and additionally include duration, voice quality (i.e. breathiness), and intensity. Vowels preceding /ʃ/ and /ʃ#s/ are longer, breathier, less intense and have formants closer to the related cardinal vowels.

Hypothesis 4: Those phonetic aspects of /s/ that are part of the preceding vowel portion remain independent of the /s/-to-[ʃ] assimilation. Thus, the vowel differences can also be found if the /s#ʃ/ and /ʃ#s/ sequences are judged to be completely assimilated to [ʃʃ] in acoustic terms.

2. Method

2.1. Rationale of the Speech Corpus

Ma et al. [2006] showed that the qualities of French vowels are subjected to strong anticipatory coarticulation [cf. also Nguyen and Fagyal, 2006]. Furthermore, it is generally known that the prosodic parameters of voice quality (i.e. degree of breathiness), duration, and intensity are involved directly or indirectly in a number of linguistic and paralinguistic measures of the speech code. For example, duration and intensity contribute cross-linguistically to the creation of perceptual prominence and hence to the signalling of stress and accent [Gay, 1978; Kohler, 2008]. Stress and accent levels in turn interact with the degree of speech reduction which is, among others, influenced by word frequency [Ernestus et al., 2006], number of syllables in the word [Lindblom, 1968; Torreira and Ernestus, 2009], presence of syntactic or phrasal boundaries [Kuzla, 2009], as well as by the semantic-pragmatic context [Local, 2003]. Regarding the latter, it must further be taken into account that duration and intensity as well as voice quality vary along with the pitch-accent category and the type of accentuation (i.e. emphatic vs. non-emphatic accents) [Niebuhr, 2010; Niebuhr and Pfitzinger, 2010]. Finally, it is a well established fact that voice quality, duration, and intensity are also part of the intrinsic phonetic properties of speech segments, particularly of vowels [Ladefoged, 1967; Lehiste, 1970; Clumeck, 1976; Klatt and Klatt, 1990]. So, since we were not sure how strong and consistent the hypothesized sibilant-related vowel differences would be, we decided to reduce the statistical noise induced by factors like those listed above as much as possible in order to test our hypotheses.

The existing French speech corpus of Niebuhr et al. [2011], whose examples motivated the present study, consisted of reasonably representative speech and met the experimental requirements for describing basic characteristics of sibilant realization and assimilation in French. However, the corpus does not offer the control of disturbing factors that seemed necessary for analysing reliably fine phonetic detail of vowel duration, intensity, and (voice) quality. For example, the target words in

the elicited sentences were high-frequency function words as well as low-frequency nouns or adjectives, the words contained different numbers of syllables, the target-word pairs had different degrees of syntagmatic cohesion (see the two sentences in fig. 1a; the syntactic boundary before *si* is stronger than before *sous*), and they were placed in different prosodic and coarticulatory environments, i.e. in different sentence frames. So, instead of using the existing corpus of Niebuhr et al. [2011], we set up an additional more tightly controlled speech corpus.

The new corpus was based on a set of pseudo-names. Pseudo-names have the advantage that they can be accepted by speakers as word-like units, while they simultaneously allow for a controlled segmental make-up and variation [Magnuson et al., 2003]. Moreover, they are all equally unfamiliar and hence equally infrequent for speakers. The pseudo-names that were created for the present study consisted of monosyllabic first and last names (e.g., *Vass Chame*), each with a consonant-vowel-consonant (CVC) structure. The resulting consonant clusters that bridged first name and late name (C#C) were used to form four sibilant conditions across word boundaries. As in the study by Niebuhr et al. [2011], these four conditions include the *two sequence conditions* /s#ʃ/ and /ʃ#s/ that differ in the order of the sibilants, i.e. alveolar-postalveolar or postalveolar-alveolar. Considering hypothesis 2, the two sibilant-sequence conditions were complemented by *two single-sibilant conditions* in which the voiceless alveolar and postalveolar sibilants were followed by the labial nasal /m/. The resulting /s#m/ and /ʃ#m/ sequences served as references. First, they allow us to test whether the vowel differences subsumed under hypothesis 3 in fact represent the vocalic imprint of the immediately following sibilant, or whether the occurrence or the phonetic composition of this imprint is in some way a characteristic of sibilant sequences. Second, comparing the measurements in the single-sibilant conditions with the measurements in the sibilant-sequence conditions allowed us to determine with reference to hypotheses 3 and 4 if and to what extent the degree of assimilation affects the hypothesized vowel differences.

The /m/ was used in the /s#m/ and /ʃ#m/ sequences to reduce influences of the context consonant on the analysed sibilants. Since /m/ is a labial, coarticulatory interactions with the apical or laminal tongue-tip gestures of the sibilants that could interfere with the acoustic measurements should be minimal. In order to control for coarticulatory interferences with the acoustic measurements in the vowels as well, labials were also used to delimit the first and second names. While the first names started with the voiced obstruents /b/ or /v/ (initial Cs in CVC#), the last names ended in either /m/ or /f/ (final Cs in #CVC). Finally, we refrained from attaching an 'e' to the first name. In the phonological representation of the name, the 'e' would correspond to a final schwa vowel. However, Niebuhr et al. [2011] noted that even though the schwa is typically elided in production, its orthographic/phonological presence could weaken the degree of sibilant assimilation. In the light of our hypotheses, including hypothesis 1, we aimed at triggering the strongest possible /s/-to-[ʃ] assimilations in the sibilant sequences.

Finally, three different symmetrical vowel pairs, /a_a/, /i_i/ and /u_u/, were inserted in the consonant frames of first and last names. By cross-combining the three symmetrical vowel conditions with each of the four sibilant conditions, a total of 12 CVC#CVC pseudo-names were created. They are listed in table 1 in the orthographic representations that were also presented to the speakers. The 12 pseudo-names were integrated into the constant sentence frame *J'ai vu ___ hier* ('I saw ___ yesterday').

2.2. Recording Procedure

The set of 12 sentences was arranged in a randomized order. The whole recording comprised four sets, i.e. four copies of each sentence or 48 sentences altogether. Each set was given a different randomization. The sentences were presented to the speakers on a computer screen. Between two sets the word *Pause* appeared on the screen and allowed the speakers to gather themselves again before continuing.

At the beginning of the recording session, the list of pseudo-names was given to the speakers, and they had some time to familiarize themselves with the names. Then, the speakers received the following background information. They were asked to pretend that they are secret agents who have been caught by the enemy. In the inevitable interrogation that follows now, the interrogator wants to know the name of the collaborator who was supposed to meet the speaker the other day. But since the speaker is a good secret agent, s/he invents names in the answers. In order to deceive the enemy it was

Table 1. The 12 pseudo-names that resulted from the three vowel and four sibilant conditions, and that were read by the speakers in the constant sentence frame *J'ai vu ___ hier*

	/sʃ/	/ʃs/	/sm/	/ʃm/
/a_a/	Vass Chame	Bach Same	Vass Mafe	Bach Mame
/i_i/	Biss Chime	Vich Sime	Biss Mife	Vich Mibe
/u_u/	Bouss Choume	Vouch Soume	Bouss Moufe	Vouch Moume

The vowels and sibilants in all names were framed by labial consonants. The names are given in the orthographic representation that was also presented to the speakers.

important that the speakers made the pseudo-names sound as if they were real names. The background information integrated the pseudo-names and the laboratory recording situation into a plausible communicative context. We expected that this context would help speakers to shift their focus away from the mere act of speech production, i.e. the actual aim of the recording, to the content of the sentences. Avoiding self-monitoring during speech production should generate more representative speech and facilitate sibilant assimilation. Such a contextualized elicitation of target words and sentences was not applied in the reading task of Niebuhr et al. [2011].

After the communicative context was set, the speakers got the more technical instruction that they were to read in a fluent natural-sounding fashion the 4 × 12 = 48 sentences that appeared on the screen one after the other. Moreover, they were told to produce every sentence twice: first with a careful pronunciation, as in a talk in front of an audience, and then in a fast, casual way, as in an informal conversation with a good friend. This procedure was applied successfully by Niebuhr et al. [2011]. The implicit aim of the first, careful production is for subjects to practise the sentence and the integrated pseudo-names. This practice and the speaking-style contrast facilitate assimilation in the sibilant sequence of the subsequent fast, casually produced sentence, in addition to the omission of 'e' at the end of the first name and the strong syntagmatic (i.e. syntactic and semantic) cohesions of first names and last names.

In view of the role of the slow, careful sentences as implicit practice trials, only the fast, casual productions were acoustically analysed. Prior to the actual recording session, the speakers produced 5 dummy sentences that were selected randomly for each speaker from the set of 12 sentences in order to familiarize the speakers with the recording procedure.

2.3. Speakers

Six female native speakers of French were recorded in separate sessions. Each session took around 20 min. The speakers were between 25 and 52 years old. They grew up and spent most of their life in Southern France. The speakers were recruited from the staff of the Laboratoire Parole et Langage (LPL) of the Université de Provence, Aix-en-Provence. Thus they had a background in Linguistics and/or Phonetics, but they were naïve with regard to the aim of the recording. Moreover, they were screened on an impressionistic basis with the aim to find strong assimilators, following empirical evidence that speakers of a language can differ considerably in their degree of assimilation [Zsiga, 1995; Ellis and Hardcastle, 2002; Niebuhr et al., 2011]. The recordings were made digitally (44.1 kHz, 16 bit) in a soundproof booth at the LPL.

2.4. Acoustic Analyses

The fast, casual sentences were labelled in Praat [Boersma, 2001] with regard to (a) beginning and end of the sentence, (b) type and boundaries of the vowel in the first target syllable (i.e. the first name), (c) and type and boundaries of the sibilant or the sibilant sequence.

The labelling was primarily guided by visual inspections of oscillogram and spectrogram sections, supported by auditory analyses of the first author. The fricative boundaries, including those of the sibilants and sibilant sequences, were defined in terms of the audible onset and offset of oral friction noise in the higher part of the spectrogram. Particularly for the boundary between the vowel and the following sibilant in VC#, this meant that the offset of voicing did not necessarily coincide with the onset of friction. In most cases, the voicing continued for a few periods beyond the labelled sibilant onset.

In other and comparatively rare cases, the offset of voice occurred before the onset of sibilant friction, in this way creating short [h]-like sections that were labelled as part of the vowels. In the case of the /s#m/ and /ʃ#m/ sequences, there was frequently a short section of nasal friction before the onset of voiced [m]. These sections were labelled as part of the nasal and not as part of the preceding sibilant. The remaining boundaries that concerned vowel onsets after nasals and plosives were set at spectral discontinuities or at sudden intensity increases (which coincided roughly in the vast majority of cases).

Overall durations of the sibilants and sibilant sequences were determined automatically using the Praat labels. A Praat script was used to calculate the mean CoG and the CoG range (i.e. maximum – minimum CoG value) for each sibilant or sibilant sequence. Our CoG (which is sometimes also referred to as centroid) is based on individual frequencies, not on frequency bands. It hence represents the average value of all frequencies in the spectrum that were weighted by their amplitudes before they were summed up and divided by the overall number of frequencies (the frequency resolution was determined by the FFT points).

A number of studies showed cross-linguistically that CoG values are good representatives of the acoustic and perceptual differences between alveolar and postalveolar sibilants [Heinz and Strevens, 1961; Nartey, 1984; Hoole et al., 1993; Gordon et al., 2002; Li et al., 2007; Toda, 2007; Niebuhr, 2009; Niebuhr et al., 2011]. For the sake of completeness it should be noted that we also considered alternative spectral measures like the 'ampRatio' that was shown to distinguish significantly between the /s/ and /ʃ/ of female American English speakers [Li et al., 2007]. Spot-check comparisons suggested that the ampRatio measure is not consistently superior to the CoG measure in representing sibilant qualities and changes in French. Therefore, we adhered to the well-established CoG measure.

Postalveolar sibilants yield substantially lower CoG values than alveolar sibilants. For example, the CoG values found for the French postalveolar sibilants in the study of Niebuhr et al. [2011] were between 5 and 7.5 kHz, whereas those of the alveolar sibilants were not lower than 6.5 kHz and went up to 10 kHz. The CoG measurements of the present study were based on the same method as in Niebuhr et al. [2011]. A Hamming window with a length of 30 ms (512 FFT points) was shifted through the sibilant or the sibilant sequence in steps of 7 ms. The window shift started with the left side of the window 20 ms after the onset label of the sibilant section and ended when the distance between the right side of the window and the offset label of the sibilant section fell below 20 ms.

At each step of the window shift a spectral DFT analysis of the 30-ms slice was done, and the CoG was calculated within a frequency range from 1.5 to 15 kHz. The lower threshold of the band pass was to minimize the effects of coupled back cavity resonances as well as of voice energy on the measurements (see our note on the decay of voicing after the sibilant onset above). Regarding the lower threshold, our CoG measure was similar to the 'CentHigh' measure of Li et al. [2007]. The upper threshold value was adopted from Niebuhr et al. [2011] and has been determined on a trial-and-error basis with a separate spontaneous-speech sample of French sibilants. When the window shift arrived at the end of the sibilant or sibilant sequence, the mean CoG and the CoG range were determined based on the previously calculated CoG values.

The mean CoG and CoG range measurements were used to estimate the degree of alveolar-to-postalveolar assimilation in the sibilant sequences. Lower mean CoGs and CoG ranges point to a stronger assimilation. However, since the sibilant assimilation can also result in sequences which are spectrally constant (i.e. small CoG range), but with a (blended) friction quality somewhere in between [s] and [ʃ] (creating intermediate mean CoGs), the two CoG measures are in principle independent indicators of the degree of assimilation in the time and frequency domains. We further took into account that the CoG range only represented spectral variability, but not the direction of spectral changes. For this reason we additionally included the individual initial and final CoG values of each sibilant and sibilant sequence in our set of spectral measures. The differences between the initial and

final CoG values need not correspond to the (moduli of the) CoG ranges. However, a certain correspondence between these measurements can be assumed and is intended.

The analyses of the preceding vowels included automatic, label-based duration measurements. The labelled vowel boundaries were also used to determine the mid-vowel short-term energy (i.e. intensity). By taking the mid-point value, influences of the surrounding segments were minimized. Moreover, the centre value was typically located close to the intensity maximum of the vowel. The intensity measurements were also done automatically in Praat.

In addition to the duration and intensity measurements, the frequencies of the first two formants F1 and F2 as well as the amplitude differences between the first two harmonics (H1–H2) were measured for each vowel at three points in time: 20 ms after vowel onset, in the middle of the vowel, and 20 ms before vowel offset. It should be noted that by measuring the vowel-offset values 20 ms before the boundary labels we excluded most of the few [h]-like sections mentioned above. The measurements were done manually by means of WaveSurfer (www.speech.kth.se/wavesurfer). The formant values for F1 and F2 were determined with spectral LPC slices.

The harmonic amplitude differences H1–H2 were quantified with WaveSurfer on the basis of narrow-band DFT spectra at the same points in time as the LPC slices of the formant measurements. The H1–H2 differences were used as an acoustic correlate of voice quality, following a number of previous empirical studies [Fischer-Jørgensen, 1967; Henton and Bladon, 1985; Klatt and Klatt, 1990]. The differences were to estimate the degree of breathiness across the vowel. Higher H1–H2 values point to a breathier vowel.

3. Results

3.1. Sibilant Patterns

Since each of the 6 speakers produced the 48 sentences, a total number of 288 vowel + sibilant tokens were analysed acoustically, 144 in the sibilant-sequence and another 144 in the single-sibilant condition. We will first present the results of the sibilants and sibilant sequences.

As for duration, we found a similar overall pattern as in the study by Niebuhr et al. [2011]. The sibilant sequences were on average about twice as long as the single sibilants (152 vs. 84 ms; SD 46 vs. 29 ms). In a t test for paired samples, this duration difference proved to be highly significant (t = 57.1; p < 0.001; two-tailed). Moreover, a t test that compared the paired subsamples of the single alveolar and single postalveolar sibilants showed that there is a statistical trend for the postalveolar sibilants to have greater durations than the alveolar sibilants (86 vs. 79 ms on average, SD 19 vs. 24 ms; t = 1.89; p = 0.07; two-tailed). Likewise, in an analogous t test the sibilant sequences that ended postalveolar (/s#ʃ/) tended to be longer than the sequences that ended alveolar (/ʃ#s/) (161 vs. 147 ms on average, SD 37 vs. 30 ms; t = 1.81; p = 0.08; two-tailed).

The results of the spectral sibilant measurements are displayed in figure 2. Figure 2a shows the single sibilants. The mean CoGs of the postalveolar sibilants cluster around 6–7.5 kHz, whereas the means of the alveolar sibilants are mainly located between 8 and 9.5 kHz. Thus, the two clusters are well separated along the x axis within ranges that are consistent with Niebuhr et al. [2011]. Each cluster shows further subdivisions along the x axis, which is due to the fact that the mean CoG values of the single postalveolar and alveolar sibilants varied with the vowel quality. The vowel-related variation amounts up to 1.5 kHz. The descriptive analysis of the distribution of the mean CoG values along the x axis was confirmed in a two-way repeated-measures ANOVA run with sibilant type (/s/ vs. /ʃ/, two levels) and vowel quality (/a/ vs. /i/ vs. /u/, three levels) as fixed
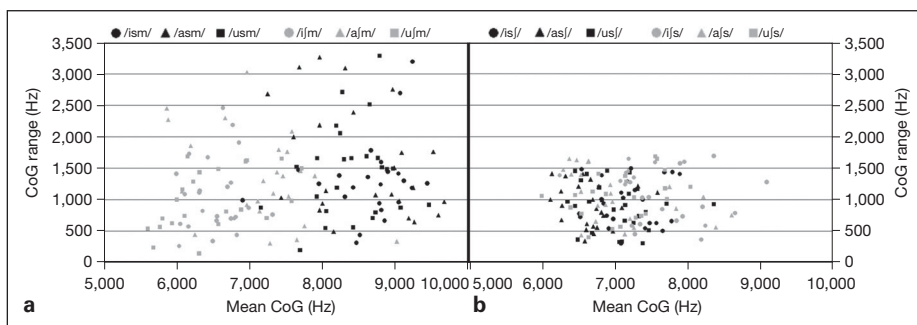
Fig. 2. Scatterplots of the spectral measurements for the single sibilants (**a**) and the sibilant sequences (**b**). For each of the 144 tokens in each panel, the CoG range (y axis) is plotted against the mean CoG (x axis). The tokens representing or starting with postalveolar sibilants (/ʃm/ and /ʃs/) are shown in grey. The alveolar ones (/sm/ and /sʃ/) are given in black. The grey and black scatterplots are further subdivided by the quality of the preceding vowel, i.e. /i/ = circles, /a/ = triangles, and /u/ = squares.

within-subject factors and mean CoG as dependent variable. Both sibilant type and vowel quality contributed highly significantly to explaining the variance of the mean CoGs [$F(1, 71) = 19.3$; $p < 0.001$; $F(2, 142) = 25.6$; $p < 0.001$]. Visual inspections of spectrograms for a subset of vowel-sibilant tokens suggest that the first reason for the vowel-quality effect is that the lower spectral energy boundary of the noise is positively correlated with the F2 of the vowel context. So, the lower spectral energy boundary was substantially higher (between 1 and 2 kHz) for /i/ than for /a/ and /u/; and this then raised the mean CoGs for /i/. The second reason is that, compared with /i/ and /a/, the main energy peaks and hence also the mean CoGs of the sibilant noise were shifted to lower frequencies in /u/ contexts. Both of these effects are in line with findings on vowel-fricative coarticulation [Nartey, 1984; Hoole et al., 1993].

The CoG ranges of the single alveolar sibilants in figure 2 look higher than those of the single postalveolar sibilants. This observation was confirmed by a second repeated-measures ANOVA with the same fixed factors as in the first ANOVA, but with CoG range instead of mean CoG as dependent variable. While the effect of sibilant type on the CoG range was significant [$F(1, 71) = 5.4$; $p = 0.02$], there was no effect of vowel quality. In both ANOVAs the interactions between the two fixed factors sibilant type and vowel quality were not significant.

Unlike the CoG results for the single sibilants, the CoGs for the sibilant sequences are not in all respects similar to those of Niebuhr et al. [2011]. For example, it is obvious from figure 2b that the two clusters of the /s#ʃ/ and /ʃ#s/ sequences overlap completely with regard to both mean CoGs and CoG ranges. The area of the scatterplot in which the CoG values overlap corresponds to the area of the single postalveolar sibilants (i.e. /ʃ#m/) in figure 2a. Accordingly, two ANOVAs, analogous to those of the single sibilants above, did not even yield a weak significant effect of the underlying sibilant type (/s#ʃ/ vs. /ʃ#s/) on the mean CoGs or the CoG ranges of the sibilant sequences. However, there was a significant effect of vowel quality on mean CoGs, which parallels the effect received for the single sibilants [$F(2, 142) = 2.9$; $p = 0.04$; i.e. mean CoGs for /i/ > /a/ > /u/]. Interactions were not significant.
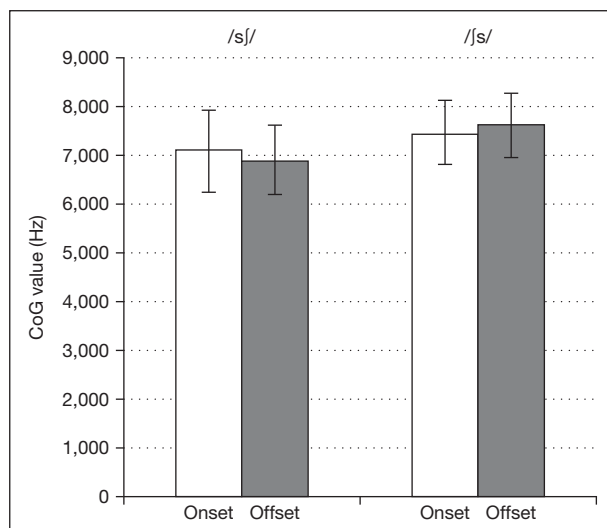
Fig. 3. Means and standard deviations of the individual CoG values measured 20 ms after the onsets (white bars) and 20 ms before the offsets (grey bars) of the /sʃ/ (left) and /ʃs/ (right) sibilant sequences. Each bar represents 72 measurements.

In addition to the mean CoGs and CoG ranges which both represent holistic properties of the sibilant sections, we looked at the individual CoG values that were measured 20 ms after the onsets and 20 ms before the offsets of the sibilant sequences. Comparing the individual CoGs of the onsets and offsets of each sequence can in principle provide more detailed information about the direction of frequency changes in noise energy across the sequences, in particular whether the noise energy is shifted to lower frequencies in /s#ʃ/ and to higher frequencies in /ʃ#s/ sequences. So, even though the CoG ranges of the two types of sibilant sequences are similarly distributed along the y axis in figure 2, this similar distribution could still emerge from different temporal variation. The results are displayed in figure 3. In terms of the mean values alone, figure 3 shows indeed that the CoGs decreased from the onsets to the offsets of /s#ʃ/ and increased from the onsets to the offsets of /ʃ#s/. However, as is clearly indicated by the great standard deviations, these mean differences did not occur consistently in the two subsamples of /s#ʃ/ and /ʃ#s/ sequences. Accordingly, two t tests for paired samples yielded no significant differences between onset and offset CoG values of the /s#ʃ/ sequences (t = 0.7; p = 0.471; two-tailed) and of the /ʃ#s/ sequences (t = 0.5; p = 0.686; two-tailed).

In summary, in terms of the acoustic sibilant noises the assimilation behaviour of the female French speakers cannot be considered to be gradual. Rather, in the narrow segmental time window of the assimilated and the assimilating sound, acoustic analysis and inferential statistics point to complete alveolar-to-postalveolar assimilation. The sibilant sequences were produced as [ʃʃ].

The overall complete assimilation found in this study is the major difference to Niebuhr et al. [2011]. Niebuhr et al. [2011] found that the mean CoGs were higher
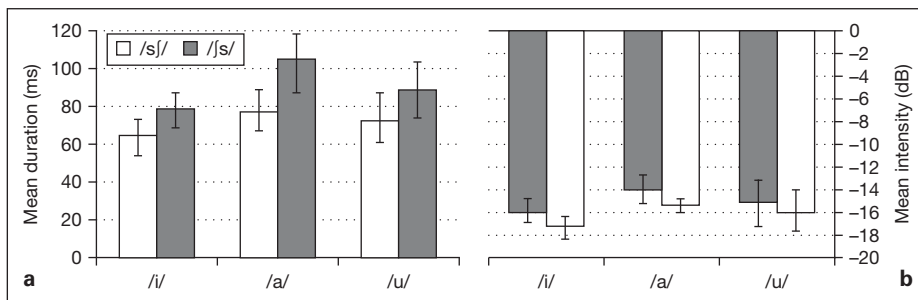
Fig. 4. Means and standard deviations for the durations (in ms, **a**) and intensities (in dB, **b**) of the three vowels /i/, /a/, and /u/ preceding the sibilant sequences /sʃ/ or /ʃs/ (white or grey bars). Note that the dB values use the highest possible 16-bit value as reference and are hence negative (i.e. smaller numbers indicate higher intensity values). Each bar represents 24 measurements.

and hence the degree of /s/-to-[ʃ] assimilation was lower for the /ʃ#s/ than for the /s#ʃ/ sequences. Moreover, there were fewer cases in which the sibilant noises were judged to be completely assimilated.

The consistently high assimilation level in the sibilant sequences shifts the focus in the signalling of /s#ʃ/ and /ʃ#s/ to the preceding vowels, but before we turn to the vowel results in detail, it should be noted that we found no significant correlations between the mean CoGs or CoG ranges and any of the vowel measurements. So there is no reason to assume that the vowel properties, presented in the following section, change with the degree of the assimilation in the sibilant sequences.

3.2. Vowel Patterns

Overall, the vowel patterns found in the two sibilant-sequence and the two single-sibilant conditions are similar. Most importantly, all three vowels /i/, /a/, and /u/ show consistent systematic differences depending on whether the immediately following sibilant was phonologically alveolar or postalveolar. Therefore the presentation of the vowel results will focus on the main conditions /s#ʃ/ and /ʃ#s/.

Figures 4–6 provide the mean values and standard deviations of duration, intensity, harmonic amplitude difference (H1–H2) and formant frequency (F1, F2) that were yielded by the measurements in the three vowels /i/, /a/, and /u/, which preceded the alveolar-postalveolar and the postalveolar-alveolar sibilant sequences. Parallel means and standard deviations were obtained for the vowels that preceded the single sibilants in the reference conditions /s#m/ and /ʃ#m/. These means and standard deviations are summarized in table 2, together with inferential statistics analogous to those of the sibilant sequences that are explained in the following.

Individual one-way ANOVAs were calculated for each of the 11 measurements of each vowel (for the complete list of measurements see table 2 of the single-sibilant results). The type of the following sibilant (/s/ vs. /ʃ/) served as two-level fixed factor. Each ANOVA based on 48 tokens (6 speakers × 4 repetitions × 2 types of following
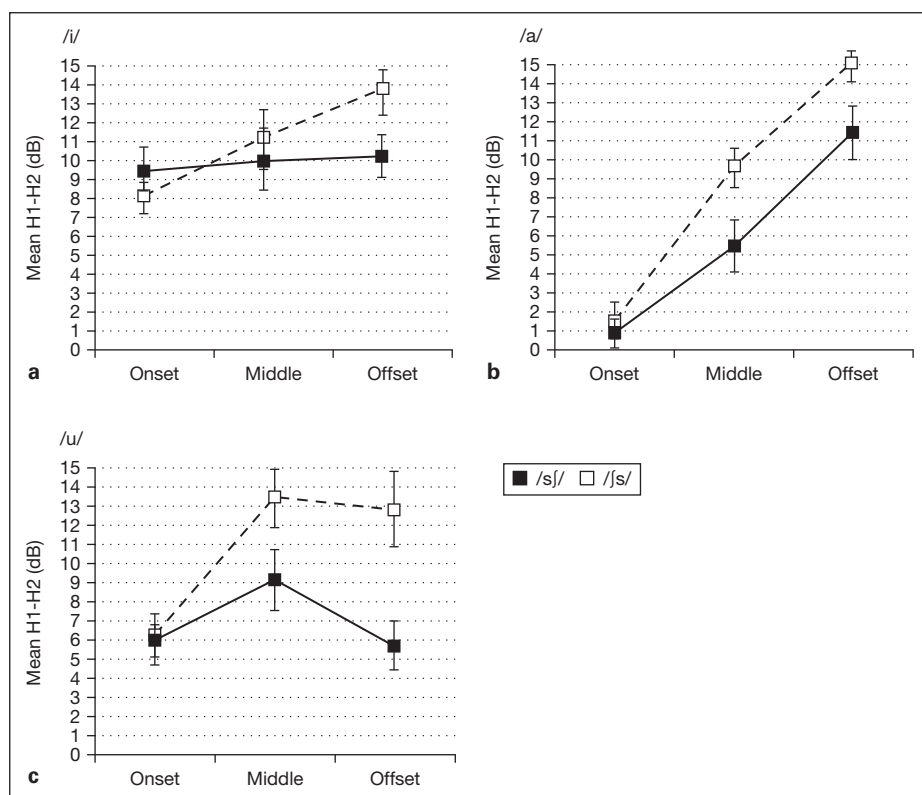
Fig. 5. Means and standard deviations of the amplitude differences (in dB) between the first two harmonics H1 and H2 in the sibilant-sequence conditions /sʃ/ and /ʃs/ (black and grey values), determined 20 ms after the onset, in the middle, and 20 ms before offset of the three vowels /i/ (**a**), /a/ (**b**), and /u/ (**c**). The curves result from a linear interpolation between the mean values. Each data point represents 24 measurements.

sibilants). Due to the high number of statistical tests (i.e. 33), only results with $p \leq 0.01$ were regarded as significant.

As is displayed in figure 4a, the two sibilant-sequence conditions had a clear effect on the durations of the preceding vowels. The sequences starting with postalveolar sibilants (/ʃ#s/) were preceded by significantly longer vowels than the sequences starting with alveolar sibilants (/s#ʃ/). These duration differences were on average larger for the intrinsically longer /a/ [about 27 ms, $F(1, 46) = 18.7$, $p < 0.001$] than for /i/ [about 15 ms, $F(1, 46) = 9.2$, $p = 0.003$] and /u/ [about 20 ms, $F(1, 46) = 25.1$, $p < 0.001$; see Lehiste, 1970, for the notion of intrinsic vowel duration]. Moreover, figure 4a shows that the longer vowels preceding /ʃ#s/ yielded 1–2 dB lower intensity values (i.e. greater negative values) than the shorter vowels preceding /s#ʃ/. However, the intensity differences were only significant for /i/ [$F(1, 46) = 7.1$, $p = 0.009$] and for /a/ [$F(1, 46) = 7.4$, $p = 0.007$]. In addition, it must be noted that the intensity values for /i/ and /a/ were negatively correlated with the H1–H2 values that measured the degree of
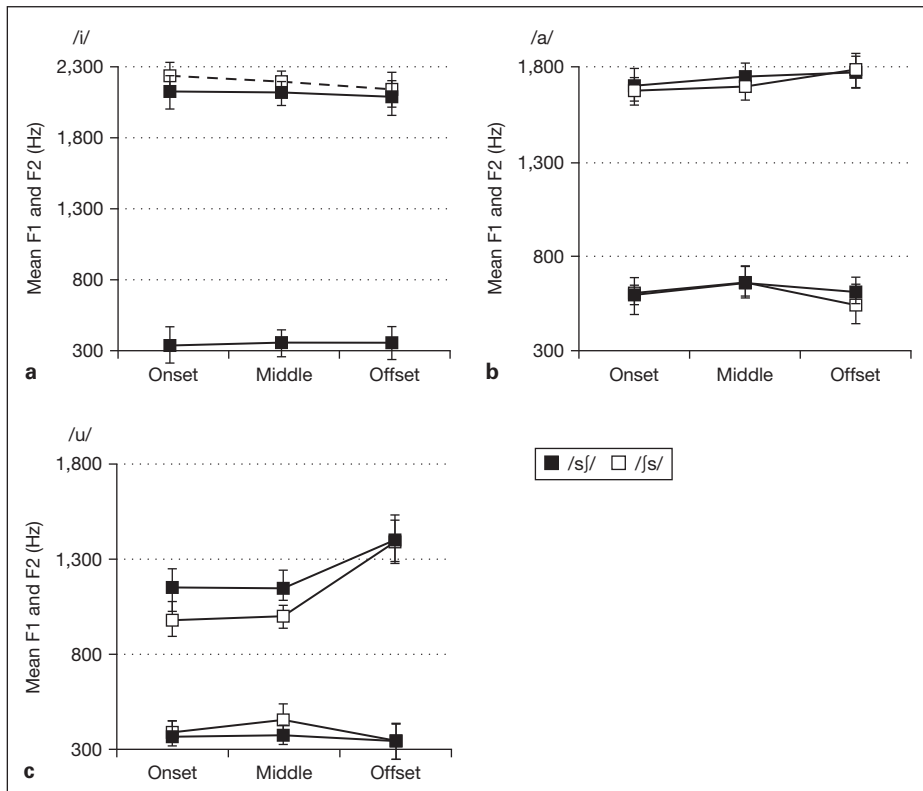
Fig. 6. Means and standard deviations for the formant frequencies (in Hz) of F1 and F2 in the sibilant-sequence conditions /sʃ/ and /ʃs/ (black and grey values). Frequencies were measured 20 ms after the onset, in the middle, and 20 ms before offset of the three vowels /i/ (**a**), /a/ (**b**), and /u/ (**c**). The curves result from a linear interpolation between the mean values. Each data point represents 24 measurements.

breathiness ($-45.1 \geq r \geq -64.3$, d.f. = 94, p < 0.001; Pearson product-moment correlation). That is, breathier vowels were also less intense. On the other hand, for none of the three vowels intensities were correlated significantly with duration. This indicates that the intensity differences are not a measuring artefact of vowel duration.

As for the harmonic-amplitude differences (H1–H2), figure 5 shows that higher values were found for vowels preceding the postalveolar than for vowels preceding the (assimilated) alveolar sibilant. More precisely, the discrepancies between the H1–H2 differences in the two sibilant-sequence conditions increase over time in all three vowels. While the grey and black curves in figure 5 have similar starting points after the vowel onsets, the curves diverge considerably over the middle to the offsets of the vowels. These divergences are also reflected in the corresponding ANOVAs, which yielded significant differences between the H1–H2 values of the /ʃ#s/ and /s#ʃ/ conditions only for the middle and offset values in /a/ [middle: $F(1, 46) = 9.4$, p = 0.002; offset: $F(1, 46) = 7.2$, p = 0.008] and /u/ (middle: $F(1, 46) = 22.2$, p < 0.001; offset: $F(1,$

Table 2. Means and standard deviations (in parentheses) of the 11 vowel–related acoustic measurements calculated for the three vowels /i/, /a/, and /u/ in the two single-sibilant reference conditions /s#m/ and /ʃ#m/

		Dur	Int	H1–2 on	H1–2 mid	H1–2 off	F1 on	F1 mid	F1 off	F2 on	F2 mid	F2 off
/i/	/s#m/	44	–17.1	5.1	8.9	9.4	310	340	300	1,830	2,000	1,910
		(9)	(3.4)	(0.8)	(2.7)	(1.5)	(40)	(30)	(20)	(120)	(150)	(100)
	/ʃ#m/	59*	–15.1*	5.7*	9.1	10.1*	320	340	290	2,180*	2,250*	2,100*
		(13)	(2.6)	(1.7)	(3.3)	(1.9)	(30)	(40)	(30)	(170)	(190)	(140)
/a/	/s#m/	76	–13.7	6.9	7.8	12.3	630	650	510	1,500	1,470	1,730
		(12)	(3.9)	(1.9)	(2.4)	(2.9)	(70)	(90)	(60)	(110)	(130)	(120)
	/ʃ#m/	79	–13.6	8.1*	11.1*	15.4*	600	590	480	1,650*	1,680*	1,900*
		(15)	(3.0)	(2.2)	(4.1)	(4.0)	(80)	(90)	(60)	(90)	(160)	(130)
/u/	/s#m/	58	–17.0	7.2	8.5	11.2	400	410	340	1,250	1,180	1,390
		(11)	(4.1)	(2.0)	(2.3)	(2.7)	(70)	(90)	(40)	(100)	(110)	(130)
	/ʃ#m/	67*	–16.8	8.3*	9.0	13.1*	370	350	320	1,000*	980*	1,560*
		(12)	(4.4)	(2.5)	(2.3)	(4.1)	(50)	(70)	(40)	(100)	(80)	(120)

From left to right the measurements refer to duration (in ms), intensity (in dB), H1–H2 amplitude differences (in dB), F1 and F2 frequencies (in Hz). The indices 'on', 'mid', 'off' refer to measurements taken 20 ms after vowel onset, in the middle of the vowel, and 20 ms before vowel offset. Asterisks behind /ʃ#m/ mean values indicate that these values differ significantly from the corresponding /s#m/ values in the table cell above. The significances result from one–way ANOVAs analogous to those run for the sibilant–sequence comparisons (i.e. two–level fixed factor type of following sibilant /s/ vs. /ʃ/, significance threshold $p \leq 0.01$).

46) = 63.5, p < 0.001]. In the case of /i/, a significant effect of the two sibilant-sequence conditions is restricted to the H1–H2 measurements before the vowel offset [F(1, 46) = 9.3, p = 0.003]. Taking into account that larger H1–H2 differences point to a higher degree of breathiness, the results mean that vowel breathiness increases (more quickly) towards postalveolar sibilants in /ʃ#s/ sequences than towards (assimilated) alveolar sibilants in /s#ʃ/ sequences.

Apart from the fact that the formant patterns of /a, i, u/ correspond well to those found in previous studies on French vowels [Meunier et al., 2003], the results of the formant frequencies also show a systematic effect of the two sibilant-sequence conditions. The effect may be summarized by saying that the /ʃ#s/ condition led to more cardinal vowel qualities than the /s#ʃ/ condition. This applies primarily to the F2 level in the vowels /i/ and /u/. As is illustrated in figure 6, F2 is higher for /i/ and lower for /u/ if the following sibilant sequence started with the postalveolar sibilant (/ʃ#s/). However, it can also be seen in figure 6 that the F2 differences for /i/ and /u/ in the /s#ʃ/ and /ʃ#s/ conditions almost disappear towards the vowel offsets. In line with these descriptive analyses, the ANOVAs yielded significant F2 differences at the onsets and in the middle, but not at the offsets of the vowels /i/ [onset: F(1, 46) = 13.4, p < 0.001; middle: F(1, 46) = 7.3, p = 0.008] and /u/ [onset: F(1, 46) = 33.6, p < 0.001; middle: F(1, 46) = 8.6, p = 0.005]. The findings for /u/ are compatible with the observations made by Hawkins and Smith [2001] (see 1.2.2). Comparing the formant patterns of British English /u/ before a single postalveolar and an alveolar-postalveolar sequence

in the sentences *Who̲ s̲harpened the meat cleaver* versus *Who's sharpened the meat cleaver*, they noted that F2 was 'considerably higher in frequency in who's than in who' [Hawkins and Smith, 2001, p. 9]. In the case of /a/, F2 differences were restricted to the midpoint of the vowel. Here, F2 was higher in the context of /ʃ#s/ than in the context of /s#ʃ/ [F(1, 46) = 7.0, p = 0.010]. This F2 difference also disappeared at the vowel offset.

Like in the two sibilant-sequence conditions, the two single-sibilant conditions did not cause a significant difference on F1. But, the results of the single-sibilant and sibilant-sequence conditions differed with respect to the second formant. In the single-sibilant condition, the F2 differences found before the sibilant onsets are greater, and they affect /a/ in the same way as /i/ and /u/ (in terms of a higher F2 before /ʃ/). Moreover, for all three vowels the F2 differences between the /s/ and /ʃ/ conditions are present across the whole vowel instead of decreasing or even disappearing (in statistical terms) towards the vowel offset (table 2).

3.3. The Vowel Contribution to Sibilant-Sequence Discrimination

In 3.1 we showed that the spectral patterns of the /s#ʃ/ and /ʃ#s/ sequences did not differ statistically in terms of mean CoGs and CoG ranges (fig. 2), nor were there systematic CoG differences between the onset and offset CoG values in either type of sibilant sequence (fig. 3). Finally, the two types of sibilant sequences did not differ in terms of overall duration. We concluded on this basis that the regressive and progressive alveolar-to-postalveolar assimilations were comparably strong in the two sibilant-sequence conditions. More specifically, from a strictly segmental standpoint the sibilant assimilation can even count as complete, i.e. [ʃʃ]. Contrariwise, in 3.2 we showed clearly that the vowels preceding /s#ʃ/ and /ʃ#s/ did still differ significantly, in several respects and in similar ways as before single alveolar and postalveolar sibilants.

These findings are in principle sufficient to discuss our hypotheses. Therefore the following supplementary analyses are primarily meant to estimate the implications of our findings for conceptual and communicative aspects of assimilation. We compared how well the underlying type of sibilant sequence can be predicted (a) based on the durational and spectral patterns of the sibilant sequences themselves and (b) with sole reference to the phonetic patterns of the preceding vowels. To this end, two sets of discriminant analyses were run separately for each of the three vowels /i/, /a/, and /u/. The first set addressed the sibilant-sequence patterns and included the five predictor variables mean CoG, CoG range, CoG onset value, CoG offset value, and sequence duration. The predictors of the second, vowel-related set comprised all 11 duration, intensity, voice quality, and formant measurements.

Using the CoG and duration measurements of the sibilant sequences, it was *not* possible to predict reliably whether the sibilant sequence was /s#ʃ/ or /ʃ#s/. The discriminant functions of all three vowel conditions were clearly not significant. Between 40 and 50% of the /s#ʃ/ tokens were confused with /ʃ#s/. In the other direction, 50–60% of the /ʃ#s/ tokens were classified as /s#ʃ/. The small difference in misassignment rate could indicate that the assimilation towards [ʃʃ] was slightly stronger in /s#ʃ/ than in /ʃ#s/ sequences, which would be in line with the findings of Niebuhr et al. [2011].

Unlike the sibilant patterns themselves, the multiparametric vowel differences were able to *predict successfully* whether the following sibilant sequence was /s#ʃ/

or /ʃ#s/. The discriminant functions were highly significant for /i/ (χ^2 = 34.8; Wilks' lambda = 0.424; p < 0.001), /a/ (χ^2 = 80.3; Wilks' lambda = 0.138; p < 0.001), and /u/ (χ^2 = 61.8; Wilks' lambda = 0.217; p < 0.001). In terms of the standardized canonical discriminant-function coefficients (that specify the contributions of the individual predictors to the separation and correct classification of the sibilant-sequence tokens as either /s#ʃ/ or /ʃ#s/), the most successful predictors – in order of importance – were: vowel duration, voice quality, F2 level, vowel intensity, and F1 (the latter did not make a substantial contribution at all). While duration and intensity were in all three vowels similarly important or unimportant predictors of the following sibilant sequence, the contribution of voice quality and F2 varied across the vowels. For /a/, F2 was the least important predictor (except for F1) whereas voice quality was roughly as important as duration. In the case of /i/, F2 contributed more than voice quality to the prediction of the upcoming sibilant sequence; and for /u/, F2 and voice quality were roughly equally important. However, in the case of all three vowels, the importance of voice quality increased, whereas the contribution of F2 decreased over the measurements taken at vowel onset, middle, and offset.

In concrete terms the multiparametric differences in /a/ and /u/ were able to correctly classify 87–96% of the subsequent sibilant as /s#ʃ/. Subsequent /ʃ#s/ sequences were even correctly predicted in 91–100% of the cases. The classification statistics that based on the multiparametric vowel differences of /i/ is worse, but with 70–80% correct sibilant-sequence classifications still remarkably efficient. As in the cases of /a/ and /u/, /i/-based classifications were slightly more successful for /ʃ#s/ than for /s#ʃ/. The lower classification rate of /i/ could be mainly due to the less clear and consistent voice quality differences between the two sibilant-sequence conditions (fig. 5).

4. Interpretation of the Findings

4.1. Hypotheses

4.1.1. Hypothesis 1: Alveolar-to-Postalveolar Sibilant Assimilation

For six of the 12 pseudo-names of the present study the first name ended in a single alveolar or postalveolar sibilant, and the last name started with a labial nasal. In these /s#m/ and /ʃ#m/ contexts the two sibilants were produced by the 6 female French speakers with clearly different spectral characteristics that are reflected in almost discrete ranges of mean CoG values. In terms of this holistic spectral measure the postalveolar sibilants yielded lower values than the alveolar sibilants, which is consistent with the fact that the friction noise of the narrow-grooved [s] is concentrated in a high-frequency band beyond 6 kHz, whereas the wide-grooved [ʃ] creates noise energy in a greater frequency band that can reach as low as 2 kHz (see the acoustic studies listed in 2.4).

In the other set of six pseudo-names the offsets and onsets of the first and last names together created sequences of alveolar and postalveolar sibilants across word boundaries in either sibilant order, i.e. /s#ʃ/ and /ʃ#s/. With respect to their overall spectral characteristics, these sequences were produced with similar mean CoGs and CoG ranges as the single postalveolar sibilants. In connection with the CoG ranges, there was also no clear shift towards a lower CoG from the onset to the offset of /s#ʃ/ or towards a higher CoG from the onset to the offset of /ʃ#s/. So, the noise energy in

the sequences was spectrally similarly distributed *and* similarly constant as the noise energy of the single postalveolar sibilants, and the remaining spectral energy variation in the sequences was not systematically related to the underlying order of the sibilants. Thus the only crucial difference between single postalveolar sibilants and the sibilant sequences was that the latter were about twice as long as the former, which rules out the interpretation in terms of /s/ elision in the sibilant sequences.

Against the backdrop of these results, hypothesis 1 may be regarded as confirmed. Alveolar-to-postalveolar assimilation is in no way an exotic (or rare) process in French sibilant sequences; and unlike in English and German, the assimilation occurs not only regressively, but also progressively. Hence, the present study replicated the previous findings Niebuhr et al. [2008, 2011] that have been made on the basis of sibilant sequences across pairs of real words. Within the labelled boundaries of the assimilated and assimilating segments, the present acoustic measurements support the assumption of complete assimilation, i.e. both /s#ʃ/ and /ʃ#s/ sequences have been realized as [ʃʃ]. In this respect the present findings differ from the more heterogeneous findings of Niebuhr et al. [2008, 2011], in which the majority of assimilations (especially those in the progressive direction) were only partial in spectral and/or temporal terms.

The consistently stronger assimilation of the present study may be explained with regard to four points. Firstly, the assimilation in the present study was produced within syntactically and prosodically highly cohesive pairs of first and last names. In contrast, the syntactic boundaries in the study by Niebuhr et al. [2008, 2011] varied from similarly cohesive sequences of nouns and adjectives to less cohesive sequences in which the word boundary in between the two sibilants coincided with the boundary of a syntactic and/or a prosodic phrase. It has been shown before that syntactic cohesion is a major factor in the degree of assimilation [Kohler, 1990; Kuzla, 2009]. Secondly, in most of the target-word pairs of Niebuhr et al. [2008, 2011] the phonological representation of the initial target word ended in a /ə/; and this word-final schwa was in fact produced in the majority of the sentences with the slow speaking rate, but practically always elided in the analysed target sentences with the fast speaking rate [as it was expected with reference to Gadet, 1992]. Due to the pseudo-names of the present study, the phonological/orthographic representations of the first names did not show a final schwa. The two sibilants were always directly adjacent, which is also reflected in the fact that schwas could be found neither in the slow and careful nor in the fast and casual productions of our sentences. Niebuhr et al. [2008, 2011] assumed that sibilant assimilation could be less strong in those cases in which it presupposes schwa elision. Thirdly, the target-word pairs of Niebuhr et al. [2008, 2011] were read in lists of isolated sentences, whereas in the present study they were additionally embedded in communicative secret-agent setting. It has been argued before that assimilation is essentially a communicative phenomenon, and is thus best elicited within a communicative framework [Kohler, 2007; Niebuhr and Kohler, 2011]. Fourthly, it is likely that the stronger and more consistent alveolar-to-postalveolar sibilant assimilation is due to the previous screening of the speakers, which aimed at recruiting those speakers that were known to the experimenters as strong assimilators, based on impressions received during everyday communication. In general, points three and four show that in the recording and analysis of speech corpora the suitability and representativity of speakers and the creation of a communicative framework should be taken into account more conscientiously.

4.1.2. Hypotheses 2–4: Vowel Differences

Over and above the alveolar-to-postalveolar sibilant assimilation the acoustic analyses of the recorded pseudo-names revealed that vowels preceding alveolar sibilants differed in their fine phonetic detail from those vowels preceding postalveolar sibilants. This held for all three vowels /i/, /a/, and /u/. The vowel differences in the two single-sibilant contexts were roughly the same as in the two sibilant-sequence contexts (with F2 at vowel offset being the only noticeable exception). This correspondence and the consistency of the differences across the three vowel qualities suggest that the fine phonetic detail results from a phonetic imprint that the following either single or sequence-initial sibilant created in the preceding vowel. That is, the imprints reflect the alveolar sibilant before /s#m/ and /s#ʃ/, and they reflect the postalveolar sibilant before /ʃ#m/ and /ʃ#s/. This conclusion is in accord with hypothesis 2.

In the two postalveolar contexts /ʃ#m/ and /ʃ#s/, the vowels were found to be longer, breathier, less intense, and they had a more cardinal vowel quality in terms of a lower F2 for /u/ and a higher F2 for /i/ and /a/ (more clearly so in the /ʃ#m/ context). So, in line with the initial informal observations in the study by Niebuhr et al. [2011], the phonetic spectrum of the imprints that alveolar and postalveolar sibilants create in the preceding vowels goes well beyond formant differences, which were so far the predominant phonetic measure in studies on the interactions between vowel and (sibilant) consonant sounds (see, for instance, Mann and Repp [1980] or Soli [1981]), including those studies that investigated vowel differences in the context of adjacent consonant assimilation [Hawkins and Smith, 2001; Gow, 2001, 2002; Local, 2003]. Thus the results of the present study are consistent with hypothesis 3.

In terms of all spectral measurements, the sibilant sequences were realized with noise qualities whose energy distributions and time courses may be judged to be indistinguishable from those of the single postalveolar sibilant in /ʃ#m/. That is, in the speakers' productions both /s#ʃ/ and /ʃ#s/ equally became [ʃʃ]. Yet, the vowels preceding /s#ʃ/ and /ʃ#s/ contexts still differed in their phonetic detail in the same way as before the single-sibilants of the /s#m/ and /ʃ#m/ contexts. Limiting this general statement, it must be noted that not all phonetic parameters actually differed significantly in and across all vowels. For example, the F2 differences of vowels preceding alveolar and postalveolar sibilants diminished or disappeared towards the vowel-sibilant boundary, and F1 did not show any sibilant-related differences. However, there was no vowel detail in the postalveolar contexts /ʃ#s/ and /ʃ#m/ that went in the direction of the alveolar contexts /s#ʃ/ and /s#m/, and vice versa. On this basis, it can be claimed that the findings support hypothesis 4, according to which an imprint of /s/, as against /ʃ/, remains in the preceding vowel, irrespective of the degree of regressive /s/-to-[ʃ] assimilation in /s#ʃ/ sequences. This conclusion also fits in with the lack of statistical correlations between CoG and vowel measurements (see 3.1).

Finally, as this study started with a critical discussion of segments, but also acknowledged that segmentation is a useful and necessary heuristic tool, we feel that we should comment on how the use of this tool may influence the reliability of our conclusions above. Most importantly, additional analyses and tests with subsamples of our sentences showed that both the vowel and the sibilant patterns do not crucially depend on the manually set boundary labels and their acoustic and auditive definitions. It is true that different definitions would have had consequences for our results. Particularly if the vowel-sibilant boundary of the first name (i.e. VC#) had been defined in terms of voice offset or the onset of acoustic friction in general, this would have affected

all sibilant and vowel measurements. However, the effects would have been primarily quantitative, not qualitative. For example, setting the vowel-consonant boundary in VC# at the onset of acoustic friction would have weakened the voice-quality differences, but enhanced the duration differences of the vowels. The CoG ranges and the differences between the initial and final CoG values would have become larger etc. However, the statistical differences and sibilant classifications would remain more or less unaffected. On this basis, we judge our findings to be robust against segmentation artefacts.

4.2. Theoretical and Communicative Implications

As was pointed out above, the present study is not the first one that deals with vowel differences in the context of adjacent consonant assimilation. However, as far as we are aware, our study is the first in which the relationship between assimilation and consonant-related vowel detail is investigated with a larger set of acoustic phonetic parameters. Due to this multiparametric approach, we were able to show that the imprint that a consonant creates in a preceding vowel and that can persist in the assimilation of the corresponding sound feature of this consonant goes beyond formant patterns. The imprint can also include duration, intensity and voice quality.

Moreover, our study is the first to include two reference contexts (here /s#m/ and /ʃ#m/), by which it was possible to relate the vowel differences clearly to the immediately adjacent consonant. The reference contexts also allowed comparing the vowel differences with and without simultaneous assimilation. In this way it was possible to estimate how the imprints of the consonants on the preceding vowels changed with simultaneous assimilation. For instance, the most noticeable change concerned the F2 differences. They were constantly present for the vowels in the non-assimilation contexts of the single sibilants, but disappeared across the vowels in the assimilation contexts of the sibilant sequences. In contrast, our analyses indicated that the duration and voice-quality differences were greater between the vowels of the assimilation contexts (/s#ʃ/ and /ʃ#s/) than between the vowels of the non-assimilation contexts (/s#m/ and /ʃ#m/).

Correspondingly, the discriminant analyses identified duration and voice quality of the vowels as the major predictors for the type of the following sibilant or sibilant sequence. F2 values also contributed to this prediction, but only those values that come from the onset or the middle of the vowel. The F1 values were practically irrelevant. This fits in with the fact that the formant differences between alveolar and postalveolar sibilants are stronger for F2 than for F1 [Soli, 1981]. Independent of this detail, the main conclusion from the discriminant analyses was that the assimilated /s#ʃ/ and /ʃ#s/ sequences could be separated by means of the vowel characteristics, but *not* by means of the noise characteristics in the sibilant sequences themselves.

With respect to speech communication, particularly the results of the discriminant analyses imply that the view of assimilation as 'one of the biggest problems for human and machine speech recognizers' [Mitterer and Blomert, 2003, p. 956] underestimates the fine phonetic detail around the sequence of assimilated and assimilating segments. Many assimilations that look complete at first sight turn out to be partial when fine phonetic detail is taken into account. Future phonetic analyses must be more careful in this respect. Provided that the fine phonetic detail is detected and used by listeners as

a means of disambiguation (and we have initial experimental evidence that this is the case, see the 'outlook' section), then 'one of the biggest problems' will hardly ever be a problem in speech communication, which also means that developing sophisticated cognitive compensation strategies for these problems is unnecessary, at least unless the listener is confronted with manipulated stimuli under experimental conditions. Of course, we do not want to claim that assimilation must always be incomplete. It is very likely that certain semantic, prosodic, and syntactic conditions as well as certain idiomatic collocations can trigger complete assimilation (which can then become conventional and end in a morphological and/or phonological language change).

Against the backdrop of the communicative implications, the related theoretical implications of our findings are also quite obvious. Most importantly, our results demonstrate the limitations of the linear-segmental approach to speech that relies on the phoneme theory. A description of French sibilant assimilation within the segmental boundaries of the sibilant sequences would have been insufficient, and, in view of our data, it would have probably led to the inadequate conclusion that the speakers produced overall complete assimilations, i.e. they neutralized the sequence-initial /s/ versus /ʃ/ contrast. This, in turn, would have entailed specifying perceptual mechanisms that can detect the assimilation and restore the original sound. Even if the assimilation perspective had been expanded to the F1–F2 formant patterns before the vowel offset, the conclusions would have been similar, as the formant patterns were the least robust residuals of the sequence-initial /s/–/ʃ/ difference in the preceding vowel.

However, when traditional prosodic parameters like duration, voice quality and intensity are also taken into account in the wider VC# time window that includes the preceding vowel, it becomes obvious that the alveolar-to-postalveolar sibilant assimilation in French was clearly not complete. On this basis, our findings also argue against the theoretical division of the speech signal into segmental and prosodic layers. Like the phoneme theory, drawing a theoretical line between segmental and prosodic layers can be a useful heuristic tool (cf. the F0 autosegments in the case of AM phonology [Ladd, 1996]). But actual communicative analyses and models of the speech signal must integrate traditional segmental and traditional prosodic perspectives. This is true for the segmental layer and its core aspects like assimilation, and it is also true for the prosodic layer and its core aspect intonation [Niebuhr, 2008, 2009].

In this paper, speech sounds have been analyzed (a) in a broader time window and (b) in an approach that integrates traditional segmental and prosodic properties, in line with the theory of phrase-level phonetics outlined in Kohler and Niebuhr [2011]. The operational speech units in this theory are opening-closing vocal-tract gestures. Even though these gestures contain vocalic and consonantal sounds, the gesture as a whole is equally shaped by both vowel and consonant articulations. In the same way as articulatory-phonetic aspects of the vowel type are reflected in the following sibilant and caused vowel-specific mean CoGs (fig. 2, 3.1), certain consonantal aspects that concern, for example, locations and shapes of the constrictions for the sibilant noises of [s] and [ʃ] as well as the interruption of voicing are prepared in the vowel portion of the gesture, and our results suggest that this part of the overall sibilant consonant gesture is not or not strongly affected by the alveolar-to-postalveolar assimilation. The assimilation rather affects that part of the consonant articulation in the gesture that creates the sibilant noise. On this basis, the following 'Outlook' section will outline ideas and perspectives for future articulatory and perceptual research.

5. Outlook

Our acoustic data raise a number of questions as to the origin of the sibilant-specific imprints on the preceding vowels and the persistence of these imprints in alveolar-to-postalveolar sibilant assimilation. Answers to these questions can only be given by subsequent articulatory studies. The articulatory studies could be guided by the following considerations that were derived from our acoustic data and from previous findings on fricative articulation.

Our key consideration is based on the fact that the productions of alveolar and postalveolar sibilants are bipartite. They must consist of a *supraglottal* (i.e. articulatory) component and of a *glottal* (i.e. phonatory) component. In French the supraglottal component includes the initiation of friction as well as 'distinct lip rounding' [Armstrong and Jones, 1967, p. 122] for postalveolar sibilants in adjacency to /u/ [Price, 1991]. The glottal component concerns the interruption of voicing. If the sibilant production is bipartite, then sibilants' imprints on the preceding vowel must also be bipartite. Is it possible on this basis that our acoustic findings result (a) from sibilant-specific coordinations of the articulatory and phonatory components and (b) from the restriction of the assimilation to the articulatory component?

The [s] friction is created between the narrow-grooved apical part of the tongue and the (back part of the) alveolar ridge of the mouth. In contrast, [ʃ] friction results from the wider-grooved laminal part of the tongue and hence involves a stronger raising of the front of the tongue body than [s] [Dart, 1998]. The acoustic consequence of this tongue-shape difference is that [ʃ] yields higher F2 values than [s] in the adjacent vowel, except when the preceding vowel is /u/. For /u/, the rounded lips in the production of [ʃ] lower F2 [Mann and Repp, 1980; Soli, 1981].

The F2 differences that we found between the vowels of the single-sibilant contexts /s#m/ and /ʃ#m/ fit in with the acoustic consequences of the articulatory patterns above. So these patterns may be the origin of the more cardinal vowel qualities before /ʃ/. Our F2 measurements suggest additionally that the sibilant-specific tongue-shape and lip patterns were present right after the vowel onsets, preparing the respective friction noise, as it would be predicted by a theoretical framework in which opening-closing gestures are the basic units that are equally shaped by vowel and consonant articulations. On the other hand, the successive disappearance of the F2 differences across the vowels in the /s#ʃ/ and /ʃ#s/ contexts as well as the lack of spectral differences between the noise patterns of two types of sibilant sequences together indicate that the articulatory component of /s/ in the preceding vowel is successively changed towards postalveolar and resembles the postalveolar component with the onset of sibilant friction.

The phonatory component must underlie an abduction of the vowel folds. It has been shown before for different languages that voiceless fricatives show even greater glottal abductions than plosives, which 'assists in the build-up of oral pressure necessary for driving the noise source' [Löfqvist and McGarr, 1987, p. 399; see also Löfqvist and Yoshioka, 1984; Hoole, 1999; Yeou et al., 2008]. In contrast, a glottal constriction would considerably reduce the transglottal airflow and the related build-up of oral pressure, which makes it a highly inefficient and unlikely strategy for the interruption of voicing in voiceless fricatives [Ladefoged, 1967]. Moreover, it is also known from the studies above that different fricatives can have differently timed glottal abduction gestures in otherwise identical prosodic contexts. The consequence of a successive preparatory

vocal-fold abduction during the vowel that precedes a voiceless fricative is an increasing vowel breathiness, which is reflected acoustically in increasing H1–H2 differences. Note that a greater H1–H2 difference is the acoustic consequence of a longer opening phase in the vibrations of the vocal folds; a turbulent supraglottal airflow, which can sound similar to glottal breathiness, does not increase H1–H2 differences [Klatt and Klatt, 1990]. The increase in breathiness in turn will decrease the intensity of the vowel. That intensity changes as a by-product of voice quality differences was found in previous studies [Gordon and Ladefoged, 2001]. In our data, this by-product assumption is supported by the significant correlations of H1–H2 and intensity values (see 3.2).

Against this backdrop, the acoustic differences in H1–H2, intensity and duration that we found in the vowels preceding /s/ and /ʃ/ could have the following phonatory origins that should be investigated in articulatory studies. The interruption of voicing for /s/ and /ʃ/ is based on vocal-fold abduction that starts for both types of sibilants in the preceding vowel. However, the abduction sets in earlier and proceeds more slowly for /ʃ/ than for /s/, which makes the vowels preceding /ʃ/ not only breathier and less intense, but also longer. Note that the laminal stricture of /ʃ/ should facilitate the initiation of friction compared with /s/. Therefore, /s/ may require an active and hence abrupt abduction of the vocal folds that can rapidly build up a high oral pressure behind the constriction, whereas a successive relaxation of the medial compression of the vocal folds may be sufficient for /ʃ/ to create voiceless sibilant friction. Finally, the persistence of the duration, intensity and voice-quality differences in the context of alveolar-to-postalveolar sibilant assimilation suggests that the phonatory components of /s/ and /ʃ/ in the preceding vowel portions are not or not strongly affected by sibilant assimilation.

The idea of a decoupling of articulatory and phonatory components and the restriction of assimilation to just one of the two components is not new. It has been suggested before, implicitly as well as explicitly, but in the opposite direction. That is, the phonatory component of a consonant is subjected to assimilation, whereas the articulatory component remains constant. For example, Kohler and Künzel [1979] report with regard to *voice* assimilation in French that the assimilation towards voicing or voicelessness need not affect the duration and formant patterns that are related to the voice-voiceless distinction in the adjacent vowel. Similar observations were made for English by Smith [1997] and Myers [2010] as well as for German by Kuzla et al. [2007].

In addition to determining the origins of our acoustic findings in articulatory studies, perception studies must pursue two different questions. The first question concerns the cross-validation of our findings by perceptual judgements. Most importantly, are the sibilant sequences that were judged on an acoustic basis to be completely assimilated also perceptually identical to actual geminates of postalveolar sibilants? The second question concerns the communicative relevance of our findings. That is, do French listeners use the revealed fine phonetic detail in the vowels to anticipate or – if the /s/-to-[ʃ] assimilation is perceptually complete – to identify the following sibilant? We are currently running such experiments, and preliminary results indicate that listeners are in fact unable to detect the /s/ in the completely assimilated sibilant sequences of the present study. But when given the appropriate vowel contexts listeners perceive an initial /s/ in the assimilated sequences of the present study and even in actual /ʃ#ʃ/ sequences, i.e. they mistake /ʃ#ʃ/ for /s#ʃ/. The cues that trigger the perception of a sequence-initial /s/ are vowel-specific and agree with the predictor hierarchies that emerged from the discriminant analyses.

Acknowledgements

References

Armstrong, L.E.; Jones, D.: The phonetics of French: a practical handbook (Bell & Sons, London 1967).

Barry, W.; Hawkins, S.: Comments on: On types of coarticulation (Hewlett/Shockey); in Docherty, Ladd, Papers in Laboratoy Phonology. II: Gesture, segment, prosody, pp. 138–145 (Cambridge University Press, Cambridge 1992).

Bell-Berti, F.; Harris, K.: Temporal patterns of coarticulation: lip rounding. J. acoust. Soc. Am. *71*: 449–454 (1982).

Benguerel, A.P.; Adelman, S.: Perception of coarticulated lip rounding. Phonetica *33*: 113–126 (1976).

Boersma, P.: Praat, a system for doing phonetics by computer. Glot int. *5*: 341–345 (2001).

Browman, C.P.; Goldstein, L.: Articulatory phonology: an overview. Phonetica *49*: 155–180 (1992).

Chomsky, N.; Halle, M.: The sound pattern of English (Harper & Row, New York 1968).

Clumeck, H.: Patterns of soft palate movements in six languages. J. Phonet. *4*: 337–351 (1976).

Dart, S.: Comparing French and English coronal consonant articulation. J. Phonet. *26*: 71–94 (1998).

Dilley, L.C.; Pitt, M.A.: A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. J. acoust. Soc. Am. *122*: 2340–2353 (2007).

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

Ernestus, M.; Lahey, M.; Verhees, F.; Baayen, R.H.: Lexical frequency and voice assimilation. J. acoust. Soc. Am. *120*: 1040–1051 (2006).

Fagyal, Z.; Kibbee, D.; Jenkins, F.: French: a linguistic introduction (CUP, Cambridge 2006).

Fischer-Jørgensen, E.: Phonetic analysis of breathy (murmured) vowels in Gujarati. Indian Ling. *28*: 71–139 (1967).

Gadet, F.: Le français populaire (PUF, Paris 1992).

Gay, T.: Effect of speaking rate on formant movements. J. acoust. Soc. Am. *63*: 223–230 (1978).

Gordon, M.; Barthmaimer, P.; Sands, K.: A cross-linguistic study of voiceless fricatives. JIPA *32*: 141–174 (2002).

Gordon, M.; Ladefoged, P.: Phonation types: a cross-linguistic overview. J. Phonet. *29*: 383–406 (2001).

Gow, D.: Assimilation and anticipation in continuous spoken word recognition. J. Memory Lang. *45*: 133–139 (2001).

Gow, D.: Does English coronal place assimilation create lexical ambiguity? J. exp. Psychol hum. Perception Performance *28*: 163–179 (2002).

Gow, D.: How representations help define computational problems: commentary on Grossberg, Gaskell and Greenberg. J. Phonet. *31*: 487–493 (2003).

Hardcastle, W.J.; Hewlett, N.: Coarticulation, theory, data and techniques (Cambridge University Press, Cambridge 1999).

Hawkins, S.: Arguments for a non-segmental view of speech perception. Proc. 13th ICPhS, Stockholm, 1995, pp. 18–25.

Hawkins, S.; Smith, R.: Polysp: a polysystemic, phonetically-rich approach to speech understanding. J. Ital. Ling. Riv. Ling. *13*: 99–188 (2001).

Heid, S.; Hawkins, S.: An acoustical study of long-domain /r/ and /l/ coarticulation. Proc. 5th ISSP, Kloster Seeon 2000, pp. 77–80.

Heinz, J.M.; Strevens, K.N.: On the properties of voiceless fricative consonants. J. acoust. Soc. Am. *33*: 589–596 (1961).

Henton, C.G.; Bladon, R.A.W.: Breathiness in normal female speech: inefficiency versus desirability. Lang. Commun. *5*: 221–227 (1985).

Holst, T.; Nolan, F.: The influence of syntactic structure on [s] to [ʃ] assimilation; in Connell, Arvaniti, Papers in Laboratory Phonology. IV: Phonology and phonetic evidence, pp. 315–333 (CUP, Cambridge 1995).

Hoole, P.: Laryngeal coarticulation. Section 1: Coarticulatory investigations of the devoicing gesture; in Hardcastle, Hewlett, Coarticulation: theory, data and techniques, pp. 105–121 (CUP, Cambridge 1999).

Hoole, P.; Nguyen-Trong, N.; Hardcastle, W.J.: A comparative investigation of coarticulation in fricatives: electropalatographic, electromagnetic, and acoustic data. Lang. Speech *36*: 235–260 (1993).

IPA: Handbook of the International Phonetic Association (Cambridge University Press, Cambridge 1999).

Jespersen, O.: Lehrbuch der Phonetik; 4. Aufl. (Teubner, Leipzig 1926).

Kelly, J.; Local, J.: Long domain resonance patterns in English. Proc. IEE Conf. on Speech Input/output: Techniques and Applications, London 1986, pp. 304–209.

Klatt, D.H.; Klatt, L.C.: Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. acoust. Soc. Am. *87*: 820–858 (1990).

Kohler, K.J.: Segmental reduction in connected speech in German: phonological facts and phonetic explanations; in Hardcastle, Marchal, Speech production and speech modelling, pp. 69–92 (Kluwer Academic Publishers, Dordrecht 1990).

Kohler, K.J.: Beyond laboratory phonology: the phonetics of speech communication; in Solé, Beddor, Speeter, Ohala, Experimental approaches to phonology, pp. 41–53 (OUP, Oxford 2007).

Kohler, K.J.: The perception of prominence patterns. Phonetica *65*: 257–269 (2008).

Kohler, K.J.; Künzel, H.J.: Produktion und Perzeption der Plosive und Frikative im heutigen Standard-Französisch: wort- und satzphonetische Untersuchungen. AIPUK *11*: 135–190 (1979).

Kohler, K.J.; Niebuhr, O.: On the role of articulatory prosodies in German message decoding. Phonetica *68*: 57–87 (2011).

Kuzla, C.: Prosodic structure in speech perception and production; PhD diss. Max Planck Institute of Psycholinguistics, Nijmegen (2009).

Kuzla, C.; Cho, T.; Ernestus, M.: Prosodic strengthening of German fricatives in duration and assimilatory devoicing. J. Phonet. *35*: 301–320 (2007).

Ladd, D.R.: Intonational phonology (Cambridge University Press, Cambridge 1996).

Ladefoged, P.: Three areas of experimental phonetics (OUP, London 1967).

Lehiste, I.: Suprasegmentals (MIT Press, Cambridge 1970).

Li, F.; Edwards, J.; Beckman, M.: Spectral measurements for sibilant fricatives of English, Japanese, and Mandarin Chinese. Proc. XVIth Int. Congr. Phonet. Sci., Saarbrücken 2007, pp. 917–920.

Liberman, A.M.; Mattingly, I.G.: The motor theory of speech perception revised. Cognition *21*: 1–36 (1985).

Lindblom, B.: Temporal organization of syllable production. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 9, pp. 1–5 (1968).

Local, J.: Variable domains and variable relevance: interpreting phonetic exponents. Proc. TIPS, Aix-en-Provence 2003, pp. 101–106.

Löfqvist, A.; Yoshioka, H.: Intrasegmental timing: laryngeal-oral coordination in voiceless consonant production. Speech Commun. *3*: 279–289 (1984).

Löfqvist, A.; McGarr, N.: Laryngeal dynamics in voiceless consonant production; in Baer, Sasaki, Harris, Laryngeal function in phonation and respiration, pp. 391–402 (College Hill, Boston 1987).

Ma, L.; Perrier, P.; Dang, J.:Anticipatory coarticulation in vowel-consonant-vowel sequences: a crosslinguistic study of French and Mandarin speakers. Proc. 7th Int. Semin. on Speech Production, Ubatuba 2006, pp. 151–158.

Magnuson, J.S.; Tanenhaus, M.K.; Aslin, R.N.; Dahan, D.: The time course of spoken word learning and recognition: studies with artificial lexicons. J. exp. Psychol. gen. *132*: 202–227 (2003).

Mann, V.A.; Repp, B.H.: Influence of vocalic context on perception of the [ʃ]-[s] distinction. Perception Psychophysics *28*: 213–228 (1980).

Menzerath, P.; Lacerda de, A.: Koartikulation, Steuerung und Lautabgrenzung (Dümmlers, Berlin 1933).

Meunier, Ch.; Frenck-Mestre, C.; Lelekov-Boissard, T.; Le Besnerais, M.: Production and perception of vowels: does the density of the system play a role? Proc. 15th ICPhS, Barcelona 2003, pp. 723–726.

Mitterer, H.; Blomert, L.: Coping with phonological assimilation in speech perception: evidence for early compensation. Perception Psychophysics *65:* 956–969.

Morton, J.; Marcus, S.; Frankish, C.: Perceptual centers. Psychol. Rev. *83*: 405–408 (1976).

Myers, S.: Regressive voicing assimilation production and perception studies. JIPA *40*: 163–179.

Nartey, J.: Coarticulation effects on fricative consonants across languages. J. acoust. Soc. Am. *75*: 66 (1984).

Nguyen, N.; Fagyal, Z.: Acoustic aspects of vowel harmony in French. J. Phonet. *36*: 1–27 (2006).

Niebuhr, O.: Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German. J. acoust. Soc. Am. *142*: 1252–1263 (2008).

Niebuhr, O.: Intonation segments and segmental intonations. Proc. 10th Interspeech Conf., Brighton 2009, pp. 2435–2438.

Niebuhr, O.: On the phonetics of intensifying emphasis in German. Phonetica *67*: 170–198 (2010).

Niebuhr, O.: On the domain of auditory restoration in speech; in Kokinov, Karmiloff-Smith, Nersessian, European perspectives on cognitive science, pp. 1–6 (New Bulgarian University Press, Sofia 2011).

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

Niebuhr, O.; Kohler, K.J.: Perception of phonetic detail in the identification of highly reduced words. J. Phonet. *39*: 319–329 (2011).

Niebuhr, O.; Lancia, L.; Meunier, Ch.: On place assimilation in French sibilant sequences. Proc. 8th ISSP, Strasbourg 2008, pp. 221–224.

Niebuhr, O.; Meunier, Ch.; Lancia, L.: The role of the vowel context in the differentiation of French /sʃ/ and /ʃs/ sequences. PaPI 2009, Las Palmas, 2009. http://www.ipds.uni-kiel.de/on/downloads/Pres_PaPI_NiebuhrEtAl_2009_final.ppt

Niebuhr, O.; Pfitzinger, H.R.: On pitch-accent identification – the role of syllable duration and intensity. Proc. 5th Speech Prosody Conf., Chicago 2010.

Nolan, F.: The descriptive role of segments: evidence from assimilation; in Ladd, Docherty, Papers in Laboratory Phonology. II, pp. 261–280 (CUP, Cambridge 1992).

Öhman, S.E.G.: Coarticulation in VCV utterances: spectrographic measurements. J. acoust. Soc. Am. *39*: 151–168 (1966).

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

Price, G.: An introduction to French pronunciation (Blackwell, Oxford 1991).

Ramus, F.: Outstanding questions about phonological processing in dyslexia. Dyslexia *7*: 197–216 (2001).

Smith, C.: The devoicing of /z/ in American English: effects of local and prosodic context. J. Phonet. *25*: 471–500 (1997).

Snoeren, N.D.; Segui, J.; Hallé, P.A.: Perceptual processing of partially and fully assimilated words in French. J. exp. Psychol. hum. Perception Performance *34*: 193–204 (2008).

Soli, S.: Second formants in fricatives: acoustic consequences of fricative-vowel coarticulation. J. acoust. Soc. Am. *70*: 976–984 (1981).

Stevens, K.N.: On the quantal nature of speech. J. Phonet. *17*: 3–45 (1989).

Sweet, H.: A primer of phonetics (Oxford University Press, Oxford 1906).

Toda, M.: Speaker normalization of fricative noise: considerations on language-specific contrast. Proc. 16th ICPhS, Saarbrücken2007, pp. 825–828.

Torreira, F.; Ernestus, M.: Probabilistic effects on French [t] duration. Proc. Interspeech Conf. 2009, Brighton 2009, pp. 448–451.

Tunley, A.: Coarticulatory influences of liquids on vowels in English; PhD diss. University of Cambridge (1999).

West, P.: The extent of coarticulation of English liquids: an acoustic and articulatory study. Proc. 14th ICPhS, San Fransisco1999, pp. 1901–1904.

West, P.: Perception of distributed coarticulatory properties of English /l/ and /ɹ/. J. Phonet. *27*: 405–426 (2000).

Yeou, M.; Maeda, S.; Honda, K.: Laryngeal activity in the production of consonants clusters and geminates in Moroccan Arabic. Proc. 8th Int. Semin. on Speech Production, Strasbourg 2008, pp. 249–252.

Zsiga, E.: An acoustic and electropalatographic study of lexical and postlexical palatalization in American English; in Connell, Arvaniti, Papers in Laboratory Phonology. IV: Phonology and phonetic evidence, pp. 282–302 (CUP, Cambridge 1995).