Vowel elision in casual French: the case of vowel /e/ in the word *c'était*

Francisco Torreira[a,b], Mirjam Ernestus[a,b]

[a] *CLS, Radboud Universiteit, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands*
[b] *Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands*

Abstract

This study investigates the reduction of vowel /e/ in the French word *c'était* /setɛ/ 'it was'. This reduction phenomenon appeared to be highly frequent, as more than half of the occurrences of this word in a corpus of casual French contained few or no acoustic traces of a vowel between [s] and [t]. All our durational analyses clearly supported a categorical absence of vowel /e/ in a subset of *c'était* tokens. This interpretation was also supported by our finding that the occurrence of complete elision and [e] duration in non-elision tokens were conditioned by different factors. However, spectral measures were consistent with the possibility that a highly reduced /e/ vowel is still present in elision tokens in spite of the durational evidence for categorical elision. We discuss how these findings can be reconciled, and conclude that acoustic analysis of uncontrolled materials can provide valuable information about the mechanisms underlying reduction phenomena in casual speech.

Keywords: vowel elision; reduction; French; casual speech.

1. Introduction

One of the main characteristics of spontaneous connected speech is the ubiquitous presence of reduced pronunciations. For a corpus of spontaneous American English, Johnson (2004) reports absences of segments in at least 20% of the word tokens, and deviations from canonical citation forms involving some sort of reduction for 60% of all word tokens (e.g. [pʰɛɾĩ] for *apparently*; [kʰz] for *because*, [pʰtʰkʰɚ] for *particular*). Although the pervasiveness of such reductions in everyday speech is now widely acknowledged (e.g. Ernestus, 2000; Kohler, 2000; Shockey, 2003; Johnson, 2004), little is

still known about how they are produced and about their exact phonetic characteristics. In fact, detailed descriptions of lenition phenomena in conversational speech are scarce, especially for non-Germanic languages such as French (excepting perhaps well-studied phenomena as schwa deletion (Bürki et al., 2007) and voicing assimilation in consonant clusters (Duez, 1995; Hallé and Adda-Decker, 2007), which for many researchers might not fall under the hood of casual speech processes).

This article contributes to filling this gap. We investigate the phonetic characteristics and conditioning factors of a reduction phenomenon affecting vowel /e/ in the word *c'était* /setɛ/ 'it was' in casual French. This reduction process, despite its high incidence in conversational French, has not received any attention in the phonetic and phonological literatures, and therefore offers a fresh ground for investigation.

Of particular interest to us is whether the elision of vowel /e/ results from the categorical absence of this segment in a reduced phonological form of the word *c'était* /stɛ/, or if it can be explained as a case of extreme gradient reduction. Browman and Goldstein (1990) made the hypothesis that casual speech reduction phenomena are mainly due to two gradient processes: reduction in gestural magnitude and increase in gestural overlap. Reduction in gestural magnitude occurs for instance when a consonantal constriction is not fully achieved, as when the word *sugar* is produced with a velar approximant rather than with a stop. This mechanism can be thought to underlie most undershoot phenomena. Gestural overlap involves the timing between articulatory movements, and may lead to different kinds of reduction including apparent segmental assimilations and deletions. For instance, the articulatory gestures in consonant clusters consisting of coronal and bilabial consonants (e.g. /tm/ in the utterance *perfect memory* or /dp/ in *hundred pounds*) are sometimes overlapped in such a way that the coronal constriction is acoustically masked by the bilabial closure, even if it is effectively achieved by the speaker.

The view that reduction phenomena in casual speech are mainly gradient has been supported by numerous studies. Davidson (2006) showed that pretonic schwa deletion in English (as in the words *potato* or *support*), which had been traditionally analyzed as the output of a phonological deletion rule (e.g. Zwicky, 1972), is the endpoint of a phonetic reduction continuum. Davidson's conclusions were based on the acoustic detail of segmental sequences with and without acoustically realized schwas, and of underlying consonant clusters lacking schwas (e.g. as in the word *sport*). This study showed, among

other things, that voiceless stops such as /p/ in the word *support* /səpɔrt/ maintain their characteristic aspiration when following [s] in cases of schwa elision, in contrast with underlying /sp/ clusters, which notably lack aspiration after the [p] release. Other connected speech phenomena now claimed to be gradient rather than categorical include nasal place assimilation in English (Nolan (1992); Byrd (1996), but see Ellis and Hardcastle (2002) for evidence of categorical assimilation in some speakers), palatalization of /s/ before /j/ in English (Zsiga, 1995), and progressive voice assimilation in German (Kuzla et al., 2007).

Although many reduction phenomena are now classified as gradient, it is uncontroversial that some reduced forms typical of casual connected speech such as English *gonna* (standard form: *going to*) and *wanna* (standard form: *want to*) do not result from online articulatory reduction. These cases can be viewed as fossilized versions of extreme gradient reductions in frequent words. Numerous lexicalized reduced forms can be found in other languages as well (e.g. *tuurlijk* for *natuurlijk* 'of course' in Dutch, *pa* for *para* 'for' in Spanish, *cê* for *você* 'you' in Brazilian Portuguese). Moreover, some sandhi and elision processes have been claimed to be categorical rather than gradient (e.g. schwa elision in French, Bürki et al. (2007, 2010); voicing assimilation in French, Hallé and Adda-Decker (2007); nasal place assimilation in Spanish and Italian, Honorof (1999); Farnetani and Busà (1994a,b); /s/ to /ʃ/ accomodation in English, Nolan et al. (1996); place assimilation in Korean, Kochetov and Pouplier (2008), among others). There is therefore evidence that connected-speech reduction phenomena may be both gradient and categorical.

According to traditional descriptions of French, in this language only schwa vowels can be elided categorically (Tranel, 1987; Coveney, 2001; Fagyal et al., 2006). Therefore, complete elision of [e] in the word *c'était* can be expected to occupy the endpoint of a gradient reduction continuum. However, *c'était* is a very recurrent expression in everyday speech, similar in this respect to English expressions with phonologically reduced variants such as *going to* and *want to*. For this reason, the possibility that the word *c'était* has a phonologically reduced variant without vowel /e/ should not be discarded a priori, even if the vowel subject to elision is not a schwa.

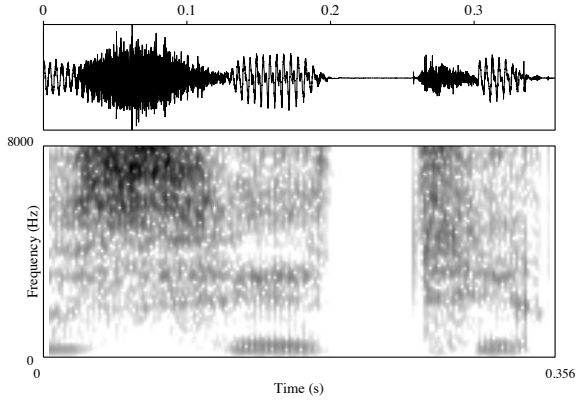
We investigate the nature of vowel elision in the word *c'était* using acoustic data from a corpus of casual speech. From a methodological perspective, we are interested in probing to what extent this sort of question, mostly addressed using laboratory speech materials and articulatory techniques (e.g.

Davidson (2006), Manuel (1992), Fougeron and Steriade (1997), Barnes and Kavitskaya (2002), Côté and Morrison (2007)), can be investigated on the basis of more naturalistic data. Although carefully controlled materials and articulatory instrumentation offer obvious advantages for the investigation of finely-grained phonetic phenomena, we suspect that many reduction phenomena characteristic of casual speech, such as the elision of vowel [e] in the word *c'était*, cannot be properly studied in a highly-controlled laboratory setting. Let us assume for a moment that vowel [e] in the French word *c'était* can be elided categorically *only in extremely relaxed and casual speech*. Under this hypothesis, which seems quite plausible to us, a controlled experiment using invasive instrumentation would probably inhibit speakers from adopting the speech register in which categorical elision is allowed to occur, and would lead to the wrong conclusion that [e] elision is gradient.

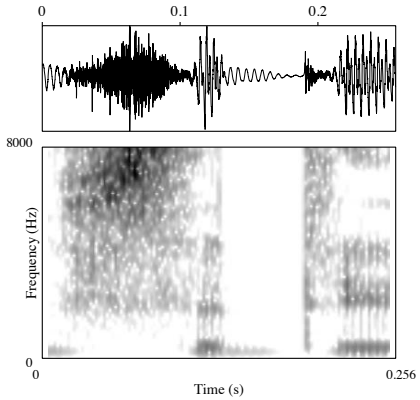
2. Realizations of *c'était* in connected speech

In this section we present several pronunciations of the word *c'était* illustrating the reduction phenomenon studied in this article. The examples in Figures 1-3 were extracted from the Nijmegen Corpus of Casual French (Torreira et al., 2010), and depict occurrences of the word *c'était* in non-prepausal position. The vocalic portion observable at the beginning of each example belongs to a word preceding *c'était*. This vocalic portion was included in the examples so that the realization of the word *c'était* can be fully appreciated from its beginning. Note also that all examples are drawn to scale in the time dimension.

Figure 1 shows two cases in which the word *c'était* received unreduced realizations containing clear [e] vocalic features between [s] and [t] (i.e. voicing, formant structure). Figure 2 shows realizations more reduced than those in Figure 1. Example 2a shows a pronunciation of *c'était* in which a few low-amplitude voicing periods attributable to /e/ can be observed only at the beginning of the stop closure. Resonances in the F2 and F3 regions appear to be reinforced before the closure and the onset of voicing (see arrows in the figure), suggesting that an opening of the vocal tract occurred between [s] and [t]. Example 2b does not display any periodicity attributable to vowel /e/, but it does exhibit an increase in energy in the F2 region before the oral closure (see arrow in the figure), indicating that, as in Example 2a, the speaker may have produced a vocal opening gesture between [s] and [t].



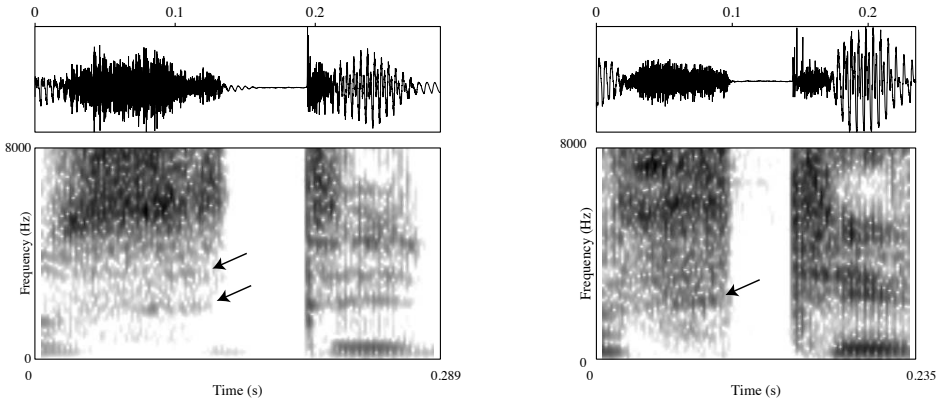
(a)



(b)

Figure 1: Examples of unreduced pronunciations of the word *c'était* /setɛ/.

Figure 3 illustrates cases in which the word *c'était* is further reduced. Example 3a does not display any trace of periodicity or formant structure attributable to /e/. Moreover, the duration of the voiceless frication portion is considerably shorter than that of Examples 2a and 2b above, making the presence of a reduced devoiced vowel between [s] and [t] rather unlikely. Example 3b exhibits additional reduction in the form of an incomplete [t] closure. In cases such as these, in which the main spectral and durational correlates of /e/ are absent from the signal, it is legitimate to wonder if the speaker attempted to pronounce a vowel at all. On the other hand, the fact that cases of intermediate reduction such as those in Figure 2 also occur



(a) (b)

Figure 2: Examples of reduced realizations of the word *c'était* /setɛ/. Arrows indicate traces of vowel /e/ between [s] and [t] in the form of F2 and F3 resonances.

in casual French suggests that those in Figure 3 might simply occupy the endpoint of an articulatory reduction continuum.

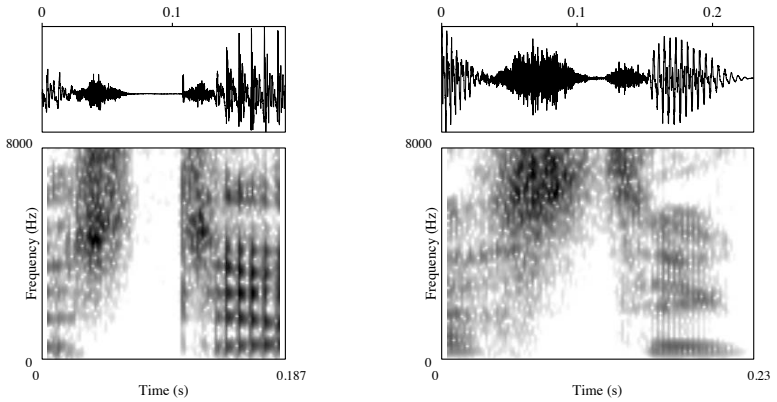
3. Quantitative study

We now investigate the reduction of vowel /e/ in the word *c'était* /setɛ/ using quantitative methods. The study consists of two parts. First, we examine the distributions of several acoustic measures taken in the region of the word *c'était* where /e/ is expected to manifest itself. We also compare some of these measures between tokens of reduced *c'était* and words beginning with underlying /st/ clusters. In the second part of the study, we investigate by means of multiple regression which among a series of prosodic, speech rate and segmental factors condition the occurrence of elision and the duration of [e] in pronunciations of the word *c'était*. Our focus is on data and analyses allowing us to determine whether /e/ elision in the word *c'était* is a gradient or a categorical phenomenon.

3.1. Method

3.1.1. The Nijmegen Corpus of Casual Speech

All materials were extracted from the Nijmegen Corpus of Casual French (NCCFr). This corpus contains 35 hours of conversations featuring 46 French speakers (24 female, 22 male). At the time of the recording, all speakers were



(a) (b)

Figure 3: Examples of extremely reduced pronunciations of the word *c'était* /setɛ/.

living in the Paris region. While most of them had been born in Île-de-France, some came from other regions in Central and Northern France. Except for two speakers, all had ages between 18 and 28. The recording equipment consisted of an Edirol R-09 solid-state stereo recorder, Samson QV head-mounted unidirectional microphones and a stereo microphone preamplifier. Microphones were placed at an approximate distance of 5 cm from the left corner of the speakers' lips. The sampling rate used was 48 KHz, while quantization was set to 32 bits. A detailed description of the preparation and recording of the NCCFr corpus can be found in Torreira et al. (2010).

3.1.2. Selection of materials

Excerpts containing the word *c'était* were randomly extracted from the NCCFr. We considered materials from all speakers excepting the two who were considerably older than the rest (i.e. 52 and 55 years old) on the grounds that they might behave differently. Excerpts in which *c'était* carried a pitch accent were disconsidered, since their low number did not allow for statistical modeling. For the same reason, we excluded tokens in which intervocalic /s/ was voiced, or in which speakers failed to achieve a full [t] closure after clearly pronouncing [e]. Excerpts in which *c'était* was immediately preceded by a voiceless obstruent were also discarded, since the beginning of [s], an important measure in our analyses, could not be determined reliably in these cases. Finally, excerpts in which *c'était* was part of an accentual phrase containing disfluencies, laughter or intrusive overlapping speech were also excluded. The

resulting dataset contained 450 tokens from 44 different speakers.

In order to compare [st] clusters in reduced *c'était* pronunciations (as in Examples 2b and 3a) with underlying /st/ clusters, we extracted all tokens of words beginning with underlying /st/ clusters followed by a non-high vowel. Word types with high vowels after /st/ were discarded because /t/ was often severely palatalized in this context. Examples of investigated word types include *stage* 'internship', *station* 'station', *stérile* 'sterile' and *stéréotype* 'stereotype'. After discarding tokens occurring in sentences with disfluencies and laughter, those preceded by a voiceless obstruent, and tokens spoken by the two oldest speakers, a total of 104 underlying /st/ clusters from 15 word types remained for analysis.

3.1.3. Measurements

The acoustic measurements taken include [e] and [s] durations, the duration of the [t] closure, the duration of voicing during the [t] closure and the spectral center of gravity in the second half of the interval marked as [s]. All measurements were done manually unless indicated otherwise.

It should be noted that the durational intervals defined below, particularly [e] and [s], do not correspond strictly to traditional segmental intervals, but rather to intervals delimited by specific acoustic landmarks attributable to the articulation of specific phonological segments. These landmarks were chosen on the basis of a clear association with the articulation of /s/, /e/ and /t/ in uncontrolled casual speech (i.e. onset of high-frequency noise as a landmark of /s/, voicing onset as a landmark of vowel /e/, start of stop closure as a landmark of /t/). For instance, in the case of [s] to [e] transitions, sudden and instantaneous increases in formant amplitudes indicative of a vocal tract opening, which can provide a reasonably consistent landmark in laboratory speech recordings, could not be used reliably in our materials. Sudden and simultaneous increases in formant amplitude as in the [se] transition of Figure 1b were rare in our dataset.

Duration of [s]: The beginning of [s] was marked at the onset of aperiodic energy in the high frequencies (> 4 KHz), while its end was defined as the onset of periodicity attributable to vowel /e/. In tokens lacking any periodicity between [s] and [t], the end of [s] was marked at the beginning of the [t] closure. Tokens lacking both a portion of [e] periodicity and a complete [t] closure did not receive an [s] duration value. According to this definition, the interval marked as [s] may contain a voiceless vocalic part.

Duration of [e]: The duration of [e] was defined as the interval extending

from the onset of periodicity after the aperiodic portion of [s] up to the start of the [t] closure. We refer to this measurement as [e] duration rather than [e] periodicity only for the sake of simplicity. Due to this definition, the initial part of this interval may still contain frication attributable to /s/. Also, cases with an [e] duration of 0 ms may actually contain a completely devoiced vowel, which in our measurement scheme is included in the [s] interval.

Duration of [s(e)]: The duration of [s(e)] includes the interval extending from the onset of [s] frication to the start of the [t] closure, regardless of whether periodicity is present within this interval. Tokens without a full [t] closure were not assigned any [s(e)] duration value, since the presence of a full [t] closure was needed to determine the end of the [s(e)] interval.

Duration of [t] closure: the start of the [t] closure was marked at the offset of energy in the F2-F3 region, while the end was marked at the beginning of [t] release. Tokens without a clear [t] closure (e.g. Figure 3b) did not receive a [t] closure duration value.

Intrusive voicing: A considerable number of [t] closures had a portion of periodicity attributable to the preceding /e/ segment. We refer to this portion of periodicity as intrusive voicing. Since the end of periodicity during the closure was often difficult to determine manually due to a very gradual decay, its duration was measured automatically by using the auto-correlation method for pitch detection available in Praat, with default settings except for time step (5 ms), silence threshold (0.03) and voicing threshold (0.04). The details of this algorithm can be found in Boersma (1993).

Spectral balance in the second half of [s]: for underlying /st/ clusters and tokens of *c'était* without voicing between [s] and [t], we calculated the spectral center of gravity in the last part of the [s] interval. Speech signals were low-pass filtered (10 Khz), and then an FFT spectrum was computed from a Hamming window placed over the second half of the [s] interval (between the center of [s] and the beginning of the [t] closure). This measure should identify gross spectral differences between the underlying and elision [st] clusters in the region where acoustic traces of a reduced/devoiced vowel might be present. If elision clusters are produced with a relaxation or opening of the [s] constriction before the closing gesture of [t], or a more posterior articulation of [s] (resulting from coarticulation with vowel /e/), a downward shift in spectral balance should be observed.

Phrase-medial	
… parce que c'était un aquarium.	'because it was an aquarium'
… et c'était assez.	'and it was enough'
… mais c'était pas vrai.	'but it was not true'
AP-initial	
moi aussi, c'était comme ça.	'me too it was like that'
plus exactement, c'était comment?	'more exactly, how was it?'
en plus, c'était la première fois.	'on top of that, it was the first time'
IP-initial	
Si. C'était marrant.	'Yes. It was funny.'
Non. C'était pendant le PACS.	'No. It was during the PACS.'
[pause] *C'était foutu pour eux.*	[pause] 'They were screwed.'

Table 1: Examples of *c'était* in each annotated phrasal position.

3.1.4. Contextual annotation

Each token was provided with an annotation of several prosodic characteristics and of the preceding segmental context. The main use of this annotation was to provide a pool of potential predictors of [e] reduction that could be used in the regression part of this study (Section 3.3).

Prosodic annotation: phrasal prosody is known to affect articulation and therefore the occurrence of reduction phenomena. Most importantly, phrase-initial segments tend to be articulated with increased articulatory strength (e.g. Fougeron, 2001; Keating et al., 2003; Cho and McQueen, 2005). For this reason, we marked whether *c'était* was preceded by a major prosodic boundary or if it was in phrase-medial position (note that *c'était* tokens in phrase-final position were not included in the dataset). Two different prosodic boundaries were distinguished: Accentual Phrase (AP) boundaries were marked after continuation rises, while Intonational Phrase (IP) boundaries were placed after final falls (typically in sentences of declarative modality, but also in some questions) or final rises in certain questions (typically yes-no questions). Table 1 illustrates utterances representative of each phrasal position type. The dataset contained a total of 71 tokens in phrase-medial position, 161 tokens in AP-initial position and 218 tokens in IP-initial position.

The beginning and end of the AP containing *c'était* were marked following the same criteria. The number of canonical syllables (referred to simply as syllables from now on) in each AP was also annotated, as well as the number of syllables separating the first syllable of *c'était* from the beginning and end of its AP. AP length ranged from 3 to 11 syllables, with an average of 4.97 syllables. Long APs occurred when speakers did not place intonational boundaries in the usual locations (i.e. at major syntactic boundaries), a phenomenon that typically occurs at fast speech rates (Fougeron and Jun, 1998).

Segmental context: we annotated whether the word *c'était* was immediately preceded by a consonant, a vowel or a silent pause.

3.1.5. Reliability check

The manual measurements and annotations mentioned above were all made by the first author. A check was performed in order to assess the reliability of these measurements and annotations. One hundred *c'était* tokens (21.9% of the dataset) were randomly selected, and independently analyzed by an assistant unaware of the purposes of our study. We computed the mean differences and correlations between the continuous measurements mentioned above ([s], [e] and [t] closure durations and the number of syllables in the AP). For the other annotations (the location of AP edges, the presence and type of major prosodic breaks before *c'était*, the presence of a complete [t] closure in *c'était*), we computed the percentages of agreement between the two annotators.

No major discrepancies were found between the two annotators. Duration measurements were generally highly correlated ($r = .96$ for [e] duration, $r = .92$ for [s] duration, $r = .88$ for closure duration, $r = .78$ for the number of syllables in the AP) and their mean absolute differences were always below 15% of the mean of the variable being checked (3.6 ms for [e] duration, 6.1 ms for [s] duration, 4.4 ms for closure duration, 0.67 for the number of syllables in the AP). The annotation of phrasal position was highly consistent between the two annotators (93%), as was the annotation of complete [t] closures (99%). The annotation of the beginning and end of the AP showed an agreement of 82%.

3.2. Results: Distribution of acoustic characteristics

3.2.1. Duration of [e]

We first examined the distribution of [e] duration in the dataset. As explained above, we defined this duration as the interval extending from the onset of voicing after the voiceless [s] segment to the beginning of the [t] closure. It should be noted that all cases without a complete [t] closure ($n = 95$, 21.1% of the total dataset) received an [e] duration value of 0 ms, since none of them exhibited any voicing between the beginning of [s] and the onset of voicing in the vowel following the reduced [t].

Figure 5 shows a kernel density plot[1] of [e] durations. Interestingly, 62% of the data points ($n = 279$) did not have any voicing between [s] and [t]. The rest of the data points ($n = 171$) exhibit positive values around a mean of 37 ms. Inspection of tokens with low [e] duration values (< 10 ms) revealed that many of them had a considerable stretch of voicing extending into the [t] closure (defined and measured as intrusive voicing; see section 3.1.3). Figure 2a illustrates this phenomenon. For the sake of simplicity, we henceforth refer to tokens with 0 ms of periodicity between [s] and [t] as *elision* cases, and to tokens with positive periodicity in this interval as *non-elision* cases.

3.2.2. Duration of [s(e)]

We then examined the distribution of [s(e)] duration, which we defined as the interval extending from the onset of [s] frication to the start of the [t] closure, regardless of whether it included any periodicity. If elision results from the absence of a complete segmental slot corresponding to vowel /e/, a bimodal distribution of [s(e)] duration should be found, and elision tokens should have shorter durations than non-elision tokens. On the other hand, if tokens with [e] duration of 0 ms actually contain a devoiced [e] vowel resulting from gradient reduction, the distribution of [s(e)] durations is expected to be unimodal.

The distribution of [s(e)] duration (solid line in Figure 5) exhibits an asymmetrical pattern, with a clear mode between 100 and 150 ms, and a

[1]Kernel density plots display the estimated probability density function (y-axis) of a continuous random variable (x-axis), and have a purpose similar to that of histograms. However, whereas histograms group observations into a discrete number of bins, kernel density plots provide a continuous estimate of the distribution of a variable. The kernel density plots shown in this article were computed using the *density* function in R with default parameter settings (R Development Core Team, 2008).

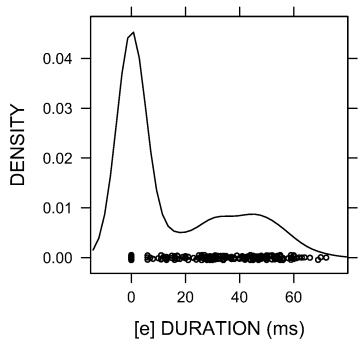


Figure 4: Kernel density plot of [e] duration. Circles represent individual data points.

salient bump between 50 and 100 ms. Interestingly, the main mode and the bump correspond closely to the modes of the distributions of the non-elision and elision groups (see dotted and dashed lines in Figure 5), suggesting that the asymmetrical pattern of [s(e)] duration is due to the pooling of two different durational populations (elision and non-elision groups). In our view, this distribution is consistent with the categorical absence of a temporal slot corresponding to /e/.

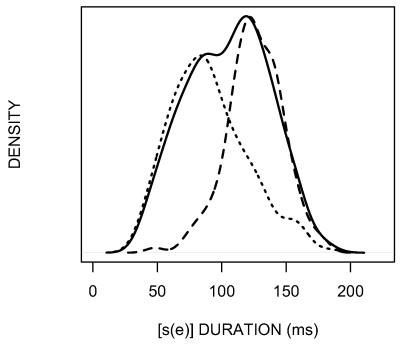


Figure 5: Kernel density plots of [s(e)] duration for pooled data (solid line), elision cases (dotted line) and non-elision cases (dashed line).

3.2.3. Duration of [s] and [t]

Previous work on segmental durations in English and Dutch has shown that consonants tend to be shorter when part of a consonant cluster (Klatt, 1974; Crystal and House, 1988; Waals, 1999). The same durational pattern may also hold for French. We therefore wondered if [s] and [t] closures in elision tokens were shorter than in non-elision tokens. If so, the difference might be taken as evidence that elision *c'était* tokens contain genuine /st/ clusters.

Mixed-effects linear models (Bates and Sarkar, 2006) were fitted with speaker as a random factor, elision as the predictor and [s] and [t] closure durations (in ms) as dependent variables. Tokens without a full [t] closure ($n = 95$), all of them elision cases, were excluded from this comparison, since they did not offer measurable [s] and [t] closure intervals. The resulting dataset contained 184 elision and 171 non-elision tokens. Results indicated that both [s] and [t] closure durations were slightly shorter in elision than in non-elision cases ([s]: $\beta = -4.97, t = -2.16, p < .05$; [t] closure: $\beta = -4.69, t = -3.16, p < .005$), supporting the possibility that in elision cases they are genuine /st/ clusters, rather than apparent /st/ clusters containing an extremely reduced /e/ vowel. Importantly, similar results were obtained when several covariates including speech rate (see Section 3.3 below for details) were included in the regression model.

3.2.4. Comparison of elision and underlying /st/ clusters

In order to further test the hypothesis that vowel /e/ was categorically absent in a subset of pronunciations of the word *c'était*, we compared [st] clusters resulting from [e] elision in *c'était* with underlying /st/ clusters in word-initial position. Three acoustic parameters were examined: [s] duration (in ms), [t] closure duration (in ms) and the spectral center of gravity (in Hz) computed in a Hamming window placed over the second half of [s] (where traces of [e] in *c'était* might be present). Mixed-effects regression with speaker as random factor was used in order to test for statistical differences between the two groups. In the case of spectral center of gravity, the gender of the speaker was used as a covariate.

Only tokens with a complete [t] closure were considered for analysis, since [s] and [t] closure durations were not available for tokens with an incomplete closure. The resulting dataset contained 184 *c'était* elision tokens and 73 tokens of words containing underlying /st/ clusters. Since underlying /st/ clusters were never preceded by a major prosodic break (because they always

occurred in content words preceded by articles or pronouns), we controlled for the preceding prosodic context in the following way. We first checked if the preceding prosodic context had a statistical effect on each of our dependent variables in the subset of *c'était* tokens. Only [s] duration was affected by the preceding prosodic context ($F(2, 181) = 5.8, p < .005$). For the analysis of [s] duration, therefore, only *c'était* tokens not preceded by a prosodic boundary were retained for comparison with underlying /st/ clusters. For [t] closure duration and the spectral center of gravity, we concluded that the preceding prosodic context was not a confounding factor and proceeded to an analysis of all available data.

Neither [s] nor closure duration differed between underlying clusters and clusters resulting from [e] elision ([s] duration: $\beta = -3.07, t = -0.69, p > .1$; closure duration: $\beta = -0.001, t = -0.001, p > .1$). On the other hand, we found a marginal effect of cluster type on the spectral center of gravity of the second half of [s] ($\beta = 292.5, t = 2.01, p < .05$). This measure tended to be slightly higher in underlying /st/ clusters than in reduced *c'était* pronunciations. These spectral differences were confirmed by observation of the dynamics of energy contours in low-pass filtered signals (3.25 KHz) from the beginning of [s] up to the [t] release. The details of this last analysis, which involves Functional Data Analysis (Ramsay and Silverman, 1997), can be found in Gubian et al. (2009).

We then investigated whether the identified spectral differences were caused by the strengthening of low formants (F1-F3) during the last part of [s] in elision *c'était* tokens. Automatic extraction of formant frequencies and bandwidths in this region was attempted through several standard procedures, but turned out to be erratic and did not provide any usable output. We then visualized spectrograms of tokens with low spectral center of gravity values. Only in twelve cases did we observe consistent energy patterns that might be attributable to F2 or F3. Importantly, in the majority of these cases formant-like resonances did not appear during the last part of [s], but were present throughout the whole interval marked as [s]. This pattern was found also in some underlying /st/ clusters, suggesting that it may not always be caused by the articulation of a vocalic gesture between [s] and [t].

We conclude from the comparison of underlying and elision clusters that these two groups do not differ significantly in duration. On the other hand, we observed slight but statistically significant spectral differences consistent with the possibility that elision clusters contain a reduced /e/ vowel.

3.2.5. Intrusive voicing

In most tokens with [e] voicing (94.3%), the initial part of the [t] closure contained some voicing periods attributable to the preceding vowel. We checked if shorter [e] durations were accompanied by longer intrusive voicing durations. If so, it might be argued that reduction in [e] duration could result from increased articulatory overlap between the voicing gesture of [e] and the upcoming stop closure. Our data suggest that this is not the case, since the durations of these two intervals did not appear to be correlated ($r = -0.09, p = .22$).

3.3. Results: Conditioning factors

In the preceding subsections we examined the phonetic characteristics of the word *c'était*. In this section we investigate which factors condition the occurrence of tokens with elided /e/ vowels. In particular, we examined whether the occurrence of [e] elision on the one hand and [e] duration on the other are conditioned by the same factors. If elision is the result of extreme gradient reduction, the occurrence of elision and [e] duration are expected to share at least some conditioning factors.

We used as predictors the prosodic and segmental variables presented in Section 3.2 (i.e. phrasal position, number of syllables in the AP, distance in number of syllables from *c'était* to the start and end of its AP, and the preceding segmental context) as well as speech rate. Speech rate was calculated by dividing the duration in seconds of a given AP, excluding its last syllable, by its number of canonical syllables minus one. The last syllable of the AP was excluded because, due to final lengthening, the inclusion of this syllable would make the estimation of speech rate less precise in phrase-initial and phrase-medial position (where the analyzed *c'était* tokens were always found). Speech rate estimates calculated in this way may be correlated with [e] duration and the occurrence of elision, since [e] duration and elision (defined as [e] duration of 0 ms), were used for the computation of speech rate. For this reason, we also subtracted the duration of [e] from the overall AP duration.

Separate regression models were fitted for the two dependent variables: the occurrence of [e] elision (defined as the absence of voicing between [s] and [t]) and the duration of [e] (defined as the duration of voicing between [s] and [t]) in tokens with positive values (non-elision tokens). Mixed-effects regression with speaker as a random effect was used in all the analyses reported

below. For the analysis of the occurrence of [e] elision, a binary variable, we used mixed-effects logistic regression.

Whenever two potentially significant predictors, say A and B, were significantly correlated, we orthogonalized them by replacing variable A in the corresponding regression model with the residuals of a linear model in which A was predicted by B. These residuals capture the information in A that cannot be attributed to its correlation with B. Furthermore, two prosodic variables, the number of syllables in the AP and the number of syllables from *c'était* to the end of the AP, were highly skewed to the right. These variables were log-transformed so that their distributions better approached a normal shape.

3.3.1. Elision of [e]

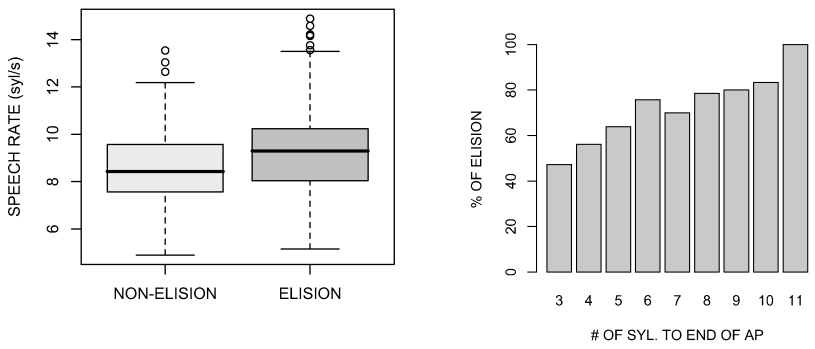
The dataset used in the analysis of [e] elision included 279 elision and 171 non-elision tokens. Only two predictors were statistically significant. First, elision was more likely the higher the speech rate ($\beta = 0.28, z = 3.52, p < .0005$). This does not mean, however, that elision was rare at slow speech rates. Figure 6a shows boxplots of speech rate for the elision and non-elision groups. It can be seen from this figure that, although there was a general tendency for speech rate to be higher in the elision group, the majority of elision tokens occurred at relatively slow and normal speech rates (< 10 syllables per second).

Second, the distance from *c'était* to the end of its AP had a positive effect on the probability of elision ($\beta = 2.17, z = 5.13, p < .0001$): elision is more likely the farther away *c'était* is located from the end of the AP. Figure 6b illustrates the effect of this factor on the probability of [e] elision.

Elision was not affected by phrasal position (IP-initial vs. AP-initial vs. phrase-medial), by the presence of a silent pause before *c'était*, nor by the preceding segmental context (vowel vs. consonant) ($p > .1$ in all cases). The number of syllables in the AP was not a significant predictor when an orthogonalized version of the number of syllables to the end of the AP, a factor with which it was highly correlated ($r = .94$), was included in the model.

3.3.2. Duration of [e]

As explained in section 3.1.3, the interval defined as [e] duration extended from the onset of voicing after the voiceless [s] segment to the beginning of the [t] closure. The label [e] duration was used for the sake of simplicity



(a) (b)

Figure 6: (a) Boxplot of speech rate values for elision and non-elision realizations of *c'était*. (b) Barplot of percentages of elision as a function of the number (#) of syllables (syl.) from *c'était* to the end of the AP.

(rather than 'periodicity duration between voiceless [s] and closure onset'), and should not be understood as vowel duration in a general sense. Only non-elision tokens ($n = 171$) were used in the analysis of [e] duration. The durations approached a normal distribution and ranged from 6 to 72 ms. Contrary to [e] elision, [e] duration was not sensitive to speech rate ($\beta = 0.04, t = 0.06, p = .95$) or to the distance from *c'était* to the end of the AP ($\beta = -1.34, t = -0.33, p = .73$). Instead, it was shorter when *c'était* followed an IP boundary than when it was in phrase-medial position ($\beta = -7.08, t = -2.1, p < .05$). This suggests that elision and the shortening of [e] duration are qualitatively different phenomena, since they appear to be conditioned by different factors.

Segments may be expected to be longer in IP-initial position than in phrase-medial position as a consequence of initial prosodic strengthening. For this reason, we wondered if the shortening effect of a preceding IP boundary on [e] also held for [s]. Contrary to [e] duration, [s] duration was found to be longer in IP-initial position than in phrase-medial position ($\beta = 7.05, t = 2.19, p = .05$). Furthermore, we found that [e] duration was inversely correlated with the duration of [s] ($r = -.28, t = -3.96, p < .0005$). This raised the question whether the effect of phrasal position on [e] duration was not actually an effect of [s] duration. We fitted a final model predicting

[e] duration with phrasal position and an orthogonalized version of [s] duration (the residuals of a linear model predicting [s] duration with phrasal position) as predictors. The results showed that only [s] duration was a significant predictor of [e] duration ($\beta = -0.28, t = -4.03, p < .0001$). This suggests that the shortening of [e] was caused by the strengthening of [s], regardless of phrasal position, rather than by phrasal position per se, and that the shortening of [e] is not due to general reduction and hypoarticulation.

4. Discussion

This study has shown that the word *c'était* /setɛ/ is frequently reduced in conversational French. More than half of the tokens of this word extracted from the Nijmegen Corpus of Casual French presented no apparent (or at most very weak) traces of vowel /e/ between [s] and [t]. Moreover, over a third of these tokens also lacked a complete [t] closure. These findings are in agreement with previous observations that reduction phenomena are more pervasive in everyday connected speech than is usually thought (Ernestus, 2000; Kohler, 2000; Shockey, 2003; Johnson, 2004).

In order to investigate whether vowel /e/ in cases of vowel elision was categorically absent or only gradiently reduced, we measured several acoustic parameters in the region in which it is expected to manifest itself. All durational measurements support the hypothesis that elision involves the categorical absence of a segmental slot corresponding to vowel /e/ between /s/ and /t/. First, the distribution of [s(e)] durations was bimodal, with each of its modes roughly coinciding with the modes of the distributions of elision and non-elision tokens. Since gradient reduction can be expected to result in a unimodal distribution of durations, this bimodality can be seen as evidence of the categorical absence of /e/ in a subset of *c'était* pronunciations. Second, it was found that the duration of [s] and [t] closures were shorter in elision cases than in non-elision cases (also after covariates such as speech rate were controlled for). These shorter durations of [s] and [t] in elision cases may be attributed to the fact that [s] and [t] form a genuine consonant cluster, rather than an apparent /st/ cluster containing a reduced /e/ vowel. In languages like English and Dutch, for instance, consonants [s] and [t] have been found to be shorter when part of a complex onset than when part of a simple onset (Klatt, 1974; Crystal and House, 1988; Waals, 1999). This interpretation was further supported by the lack of durational differences between [st] clusters arising from elision and underlying /st/ clusters. Finally, the shortening of

[s] in elision *c'était* tokens is at odds with an interpretation of elision as the result of gradient vowel reduction and devoicing. If the interval marked as [s] contained a vowel devoiced as a consequence of articulatory reduction, [s] in elision tokens should have been longer, not shorter, than in non-elision tokens.

Also in support of a categorical view of [e] elision, it was found that the probability of elision and [e] duration were conditioned by different factors. The occurrence of [e] elision was favored by fast speech rates and in utterances in which the word *c'était* occurred far from the end of its AP, suggesting that elision could be used by speakers as a strategy to reduce the overall duration of an ongoing utterance. On the other hand, the duration of [e] appeared to be inversely correlated with the duration of [s]. Importantly, in the case of non-elision tokens, [s] was not longer at faster speech rates and when far from the end of the AP, as would be expected if the occurrence of elision and the shortening of [e] were the same phenomenon. All of this suggests that the shortening of [e] durations is due to articulatory overlap between [e] and strengthened realizations of [s], rather than to general reduction and hypoarticulation.

The findings reviewed so far, all based on durational measures, clearly support a categorical account of the elision of /e/ in the word *c'était*. However, we also found that the spectral balance in the last part of the interval marked as [s] tended to be lower in reduced *c'était* pronunciations than in underlying clusters. This may be taken as a sign that a reduced or overlapped vocalic gesture is present between the gestures of /s/ and /t/ in reduced *c'était* pronunciations, since either a relaxation or a slight opening of the [s] constriction before the closing gesture of [t], or a retracted place of articulation of [s] as a consequence of coarticulation with vowel [e], can be held responsible for the observed downward shift in spectral balance.

However, it should be noted that the spectral differences between elision and underlying clusters may have been caused by uncontrolled factors, rather than by the reduced articulation of an underlying /e/ vowel in elision clusters. Underlying /st/ clusters, which in our dataset were always found at the beginning of content words, may tend to be pronounced with more articulatory effort than [st] clusters in semantically weak words like *c'était*. The spectral differences between elision and underlying clusters may be due to such a difference in articulatory effort (van Son and van Santen, 2005) rather than to the presence of an underlying /e/ vowel. Moreover, it should be noted that the identified spectral differences were relatively small (around

300 Hz according to our regression model), and that spectrographic analysis revealed traces of reduced and devoiced vowels such as those in Figure 2 only in a few cases. For these reasons, we believe that these differences in spectral balance between elision and underlying clusters should be interpreted with caution.

If these spectral differences are indeed caused by an underlying vowel, the question arises then how a simple mechanism of gradual articulatory reduction or overlap can account for the duration data discussed earlier above, which indicate that, at the temporal level, elision and vowel shortening are qualitatively different phenomena. One possibility is through gestural reorganization. Speakers might sharply reduce the word *c'était* by changing the phase relationship between the gestures of /s/ and /t/ into one similar to that of an underlying /st/ cluster. Such change in intergestural coordination would result in a gestural sequence with the durational properties of underlying clusters, but with a nonetheless different phonological and articulatory specification containing a vowel /e/. Under this hypothesis, the gestures associated to /e/ would never be categorically deleted or absent, but only downsized as a consequence of extensive overlap with the consonantal gestures of /s/ and /t/, hence the slight spectral differences between elision and underlying clusters. We are not aware of any previous research showing that such discrete changes in intergestural coordination can be the cause of reduction phenomena in everyday speech. Further research is needed to investigate this possibility.

One goal of the present study was to test the usefulness of acoustic analysis for investigating the nature of reduction phenomena in casual spontaneous speech. An advantage of acoustic analysis is that it can be applied to recordings of spontaneous conversations, which exhibit natural speech behavior, and moreover, which document speech phenomena that cannot be easily elicited under laboratory conditions. Our findings show that valuable insights into casual speech reduction phenomena can indeed be gained by using quantitative methods, rather than by providing detailed descriptions of incidental observations. In particular, analyses of the frequency distributions and conditioning factors of relevant acoustic phonetic parameters proved to be informative tools.

On the other hand, we acknowledge that acoustic data are often difficult to interpret in articulatory terms, making conclusions about the production mechanisms behind reduction phenomena unwarranted. Another limitation of acoustic data drawn from corpora of spontaneous speech is that they are

often unbalanced. For instance, only 73 underlying /st/ tokens were available for our comparison of underlying and elision [st] clusters. These limitations point to the necessity of developing new experimental paradigms allowing for the use of invasive instrumentation (e.g. electromagnetic articulography, ultrasound) which do not seriously inhibit the production of casual spontaneous speech.

To sum up, our results have shown that vowel /e/ is frequently reduced in casual renditions of the French word *c'était*. Phonetic and articulatory analyses have shown that many casual speech phenomena are gradient in nature. In our case, however, all analyses based on durational data clearly pointed towards a categorical absence of vowel /e/ in a subset of *c'était* tokens. Further research is needed to explain the slight spectral differences observed between elision and underlying /st/ clusters. Overall, our findings have shown that acoustic analysis can provide valuable information about the mechanisms behind reduction phenomena in casual speech.

5. Acknowledgements

We would like to thank Lou Boves, Louis ten Bosch, Vitoria Piai and two anonymous reviewers for their comments and help. This work was supported by a European Young Investigator Award to the second author.

References

Barnes, J. and Kavitskaya, D. (2002). Phonetic analogy and schwa deletion in French. In *Proceedings of the Twenty-Seventh Annual Meeting of the Berkeley Linguistic Society*, pages 39–50.

Bates, D. and Sarkar, D. (2006). lme4: Linear mixed-effects models using S4 classes. URL: http://CRAN.R-project.org, R package version 0.9975-10.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, volume 17, pages 97–110.

Browman, C. and Goldstein, L. (1990). Tiers in Articulatory Phonology, with some implications for casual speech. In Kingston, J. and Beckman, M., editors, *Papers in Laboratory Phonology 1: between the grammar and the physics of speech*, pages 341–376. Cambridge University Press, Cambridge.

Bürki, A., Fougeron, C., and Gendrot, C. (2007). On the categorical nature of the process involved in schwa elision in French. In *Proceedings of Interspeech 2007*. Interspeech.

Bürki, A., Ernestus, M., and Frauenfelder, U. H. (2010). Is there only one "fenêtre" in the production lexicon? on-line evidence on the nature of phonological representations of pronunciation variants for french schwa words. *Journal of Memory and Language*, 62:421–437.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24(2):209–244.

Cho, T. and McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33(2):121–157.

Côté, M.-H. and Morrison, G. S. (2007). The nature of the schwa-zero alternation in French clitics: experimental and non-experimental evidence. *Journal of French Language Studies*, 17:159–186.

Coveney, A. (2001). *The Sounds of Contemporay French*. Elm Bank Publications, Exeter.

Crystal, T. and House, A. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, 83:1553–1573.

Davidson, L. (2006). Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica*, 63:79–112.

Duez, D. (1995). On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics*, 23:407–427.

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

Ernestus, M. (2000). *Voice assimilation and segment reduction in Dutch: A corpus-based study of the phonology-phonetics interface*. LOT, Utrecht, The Netherlands.

Fagyal, Z., Kibbee, D., and Jenkins, F. (2006). *French: a Lingustic Introduction*. Cambridge University Press.

Farnetani, E. and Busà, M. G. (1994a). Consonant-to-consonant interactions in clusters: categorial or continuous processes? *Quaderni del Centro di Studio per le Ricerche di Fonetica CNR Padova*, 13:220–245.

Farnetani, E. and Busà, M. G. (1994b). Italian clusters in continuous speech. *Quaderni del Centro di Studio per le Ricerche di Fonetica CNR Padova*, 13:248–251.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29:109–135.

Fougeron, C. and Steriade, D. (1997). Does deletion of French schwa lead to neutralization of lexical distinctions? In *Proceedings of Eurospeech97*, volume 2, pages 943–946.

Fougeron, C. and Jun, S.-A. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics*, 26:45–69.

Gubian, M., Torreira, F., Strik, H., and Boves, L. (2009). Functional Data Analysis as a tool for analyzing speech dynamics: a case study on the French word c'était. In *Proceedings of Interspeech 2009*.

Hallé, P. and Adda-Decker, M. (2007). Voicing assimilation in journalistic speech. In *Proceedings of 16th International Conference on Phonetic Sciences*, pages 493–496.

Honorof, D. (1999). *Articulatory gestures and Spanish nasal assimilation*. PhD thesis, Yale University.

Johnson, K. (2004). Massive reduction in conversational American English. In Yoneyama, K. and Maekawa, K., editors, *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, pages 29–54, Tokyo, Japan. The National International Institute for Japanese Language.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. S. (2003). Domain-initial strengthening in four languages. In Local, J., Ogden, R., and Temple,

R., editors, *Phonetic interpretation: Papers in Laboratory Phonology VI*, pages 143–161. Cambridge University Press.

Klatt, D. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research*, 17:51–63.

Kochetov, A. and Pouplier, M. (2008). Phonetic variability and grammatical knowledge : an articulatory study of Korean place assimilation. *Phonology*, 25:399–431.

Kohler, K. (2000). Investigating unscripted speech: Implications for phonetics and phonology. *Phonetica*, 57:85–94.

Kuzla, C., Cho, T., and Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics*, 35(3):301–320.

Manuel, S. (1992). Recovery of "deleted" schwa. In *Perilius: Papers form the Symposium on Current Phonetic Research Paradigms for Speech Motor Control*, pages 115–118), Stockholm. University of Stockholm.

Nolan, F. (1992). The descriptive role of segments. In Docherty, G. J. and Ladd, R., editors, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pages 261–280. Cambridge University Press.

Nolan, F., Holst, T., and and, B. K. (1996). Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics*, 24:113–137.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, http://www.R-project.org.

Ramsay, J. O. and Silverman, B. (1997). *Functional Data Analysis*. Springer-Verlag.

Shockey, L. (2003). *Sound Patterns of Spoken English*. Blackwell.

Torreira, F., Adda-Decker, M., and Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52:201–221.

Tranel, B. (1987). *The Sounds of French*. Cambridge University Press.

van Son, R. and van Santen, J. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47:100–123.

Waals, J. (1999). *An Experimental View of the Dutch Syllable*. LOT, Utrecht, The Netherlands.

Zsiga, E. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Connell, B. and Arvaniti, A., editors, *Papers in Laboratory Phonology 4: phonology and phonetic evidence*, pages 282–302. Cambridge University Press, Cambridge.

Zwicky, A. (1972). Note on a phonological hierarchy in English. In Stockwell, R. and Macaulay, R., editors, *Linguistic Change and Generative Theory*. Indiana University Press.