Schwa Deletion under Varying Prosodic Conditions: Results of a Pilot Study *

Christian Geng
Christine Mooshammer
Hosung Nam
Philip Hoole

Abstract: Pretonic schwa elision, as in e.g. *support* $\rightarrow$ *sport*, has been conceptualized as a change in intergestural timing by Articulatory Phonology (Browman and Goldstein, 1994). This study aims at a preliminary evaluation of predictions made by the current version of the theory (Saltzman et al., sub) by comparing C_1 C_2 coordination in bisyllabic /C_1əC_2VC/words with monosyllabic /C_1C_2VC/ words: In order to address this question we recorded the tongue and lip movements of two speakers of British English by means of EMA. Both speakers showed the predicted longer lag between the first two consonants in words like *police* than in *please*. The first consonant was not integrated into the syllable in the pretonic schwa condition. Furthermore, we tested whether the schwa in pretonic position has a specified vowel target. The movement of the tongue dorsum suggests that the trajectory is determined by the context, at least for one speaker. This would support a targetless schwa interpretation.

1 Introduction

There has been a growing body of evidence in support of the idea that temporal patterns in speech production are an essential characteristic of phonological organization. Especially, the fact that many phonological processes, such as final devoicing, vowel neutralization and assimilation, seem to be gradual rather than categorical lends evidence to this

* Thanks to our mentor whose nickname has stimulated the design of the current experiment.

view. In this study, we want to explore further whether temporal variation alone can explain the phenomenon of pretonic schwa deletion. Impressionistically, the schwa in the first syllable of words such *police* appears to be deleted in fast speech (see Hooper, 1978; Kaisse, 1985, for a recent overview see Davidson, 2006). For many words, such as *tomorrow* $\rightarrow$ *tmorrow*, pretonic schwa deletion yields consonant sequences that are phonotactically illegal. However, a number of studies found that schwa deletion is not a categorical process with the schwa being either deleted or not, but rather varies continuously in duration and quality (e.g. Jannedy, 1994, after Beckman 1996).

Apart from the question of the categoricalness versus the graduality of the process that relates representations with and without a schwa or not, the exact nature of representation has also been debated. A related topic is whether the perceived schwa has its own specification for place of articulation, or whether it results from the timing relations between the surrounding consonants and their articulatory specifications. In the latter case the quality of this so-called transitional schwa would be determined completely from the surrounding consonants, and - even more interestingly from a phonological point of view - would not be there underlyingly.This approach, which runs counter to most phonological theories, accounts for the fact that schwa deletion is not a categorical process but a continuous process varying on a hyperarticulation continuum. One pole of this continuum is occupied by a hyperarticulated version with a substantial delay between the two consonants, which is transformed into the other pole by a gradual increase in overlap in such a way that the first syllable eventually loses its syllabicity.

The variation between these two extreme forms of timing was modeled within the framework of Articulatory Phonology (Browman and Goldstein, 1990a) as a continuous stepwise change in the timing between the two consonants. Timing relations between gestures account for e.g. reduction phenomena (Browman and Goldstein, 1990b), syllable related allophony in laterals and nasals (for an overview see Krakow, 1999) and the properties of consonant sequences in different syllable positions (Browman and Goldstein, 2000). In earlier versions of Articulatory Phonology, intergestural timing was specified manually or based on rule; however, more recently, a planning model of intergestural timing has been developed (Browman and Goldstein, 2000; Nam and Saltzman, 2003; Goldstein et al., 2010). In this model, gestures are associated with planning oscillators and pairwise coupled to one another, which constructs a coupling network of gestures, called coupling graph. Steady-state relative phases from coupled oscillators are used to trigger relative timing among gestures, which is timing between gestural onsets. The coupling graphs (see Saltzman et al., 2008) can be organized on several prosodic levels which control the organization of linguistic material in a more principled way than was possible in the older versions of AP. In the model, vowel and consonant gestures are coupled with one another in two basic modes (in-phase and anti-phase), which are spontaneously available without difficulty (Haken et al., 1985) or the necessity for repetitive practice. In CV sequences, the consonant and vowel gestures are coupled to each other synchronously or in phase, and in VC sequences the consonant and vowel gestures are sequentially coupled with a 180°phase-lag or so-called anti-phase coupling. Note that this kind of coupling is also assumed for consonant sequences. Based on studies by Byrd (1996), Honorof and Browman (1995), Browman and Goldstein (2000), and others, the timing of consonant sequences differs with respect to their position in the syllable. The organization of /C_1C_2VC/ material is assembled by the onset consonants' competing tendencies to be coupled in-phase to the vowel and anti-phase to each other. The resulting compromise shifts the phasing of the onset material from 180° in consonant sequences in codas or across boundaries to 120° (C-centre timing, see Figure 1 on the following page). This is in contrast to /C_1C_2VC/ with a reduced initial syllable. The timing between C_1 and C_2 is sequential, which corresponds to 180°. In absolute terms it depends on the reduction introduced by the prosodic modulation gesture (Saltzman et al., 2008). Additionally the fact that schwas are colored by the consonantal context to a much greater degree than unreduced vowels can be modeled via shortening due to a prosodic modulation gesture. For the bisyllabic case, as in *police*, two different variants have been suggested, with and without an underlying schwa. The first one to discuss here assumes that the schwa in the pretonic syllable is specified for TBCL (tongue body constriction location) and TBCD (tongue body constriction degree), i.e. the schwa has a target of its own (for the assumption of a lexical schwa see e.g. Smorodinsky, 2002). As shown in Figure 2 the timing relation between the initial stop and the lateral in e.g. *police*

<please>

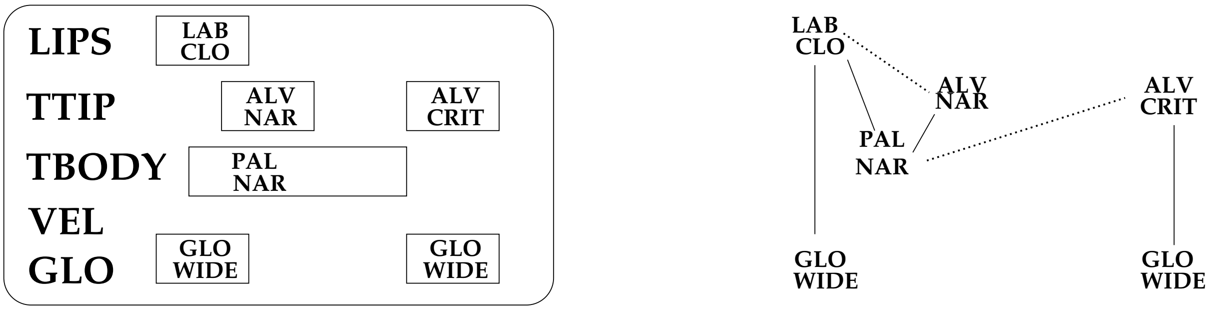

Figure 1. Illustration of C-centre timing in complex onsets (120°). In-phase couplings are shown as solid lines, anti-phase couplings as dotted line. Explanations of Gestural CL/CD specifications: LAB CLO - Labial Closure, PAL NAR - Palatal Narrowing, ALV NAR/CRIT - Alveolar narrowing / critical constriction, GLO WIDE - Glottis wide.

is anti-phase even though they are not directly linked. The initial /p/ is initiated synchronously with the schwa, i.e. a 0° in-phase coupling; the schwa is timed sequentially with the following syllable, i.e. a 180° anti-phase coupling. Evidence for an underlying schwa with its own TBCL and TBCD gestures comes from Browman and Goldstein (1994); they found tongue dorsum troughs during the schwa in sequences such as /pipəpip/. One possible explanation for the apparent contradiction between Browman and Goldstein's papers (1990a and 1994) could come from the effects of the segmental context (stop-liquid versus stop-stop environments) as hypothesized in Browman and Goldstein (1990a).[1]

The second modeling variant with a transitional schwa has been suggested by several researchers, e.g. Browman and Goldstein (1990a) and Davidson (2006) for American English. According to their hypothesis, *bray* and *beret* are distinguished only by their gestural organization but not by the number of gestures. As shown in Figure 3, even without an explicit vowel gesture in the first syllable the timing between /p/ and /l/ is sequential, which is the default cross-syllable phasing. Again, shortening of the pretonic syllable can be modeled by applying a prosodic modification gesture to the reduced syllable.

Therefore, the two representations in Figures 2 and 3 do not differ in their phasing specification but in the fact that for the first representation

1 The paper that appeared later (Browman and Goldstein, 1994) actually seems to predate the older work in Browman and Goldstein (1990a).

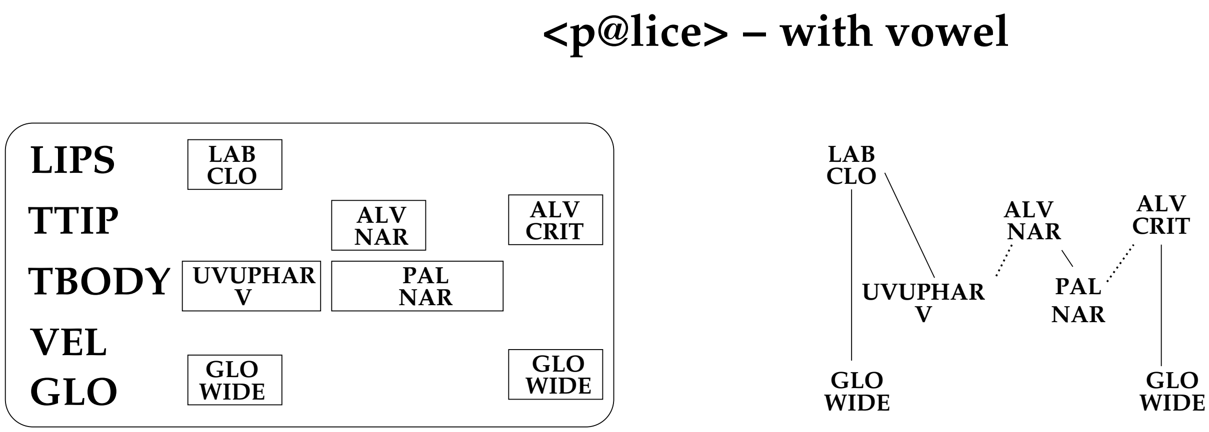

Figure 2. Gestural Score and coupling graph in /C_1əC_2VC/. The example word is represented with a reduced initial syllable. This representation contains a lexical schwa but is otherwise identical to Figure 3.

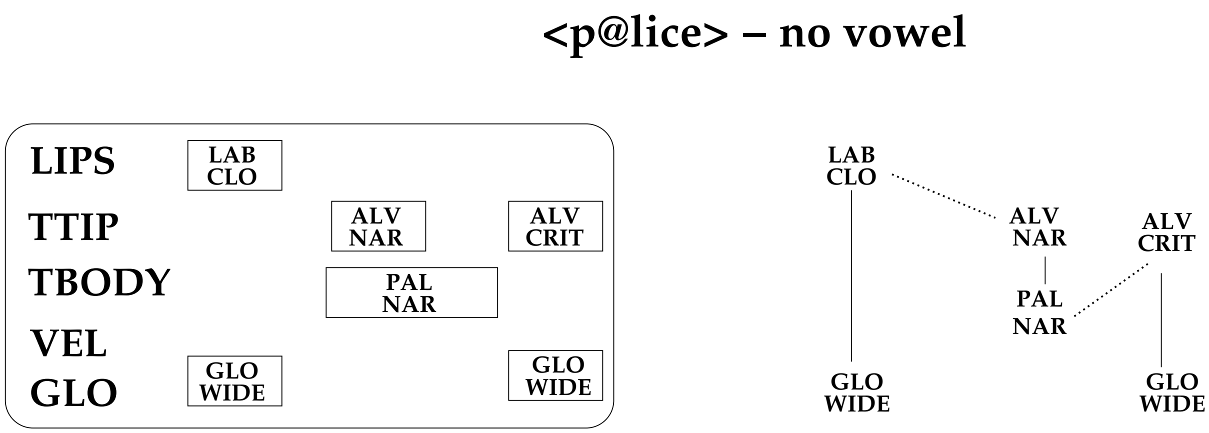

Figure 3. Gestural Score and coupling graph in /C_1əC_2VC/ with a reduced initial syllable. This variant has no lexical schwa.

we expect the tongue dorsum to move towards a vocalic target whereas for the targetless schwa the tongue dorsum trajectory during the labial and tongue tip gestures is completely determined by the consonantal context. Browman and Goldstein (1990a) informally stated that they did not observe differences in tongue dorsum movement shape during the interval between the bilabial and /r/ in words like *beret* and *bray*.
In the current study we aim at testing empirically whether speech production data give evidence for a targetless schwa or for a specified vowel gesture in pretonic position. Furthermore, the timing between the first consonants in word pairs such as *police* and *please* is investigated. Our assumption is that we will find a longer lag between the first two consonants in the pretonic reduced syllable than between the consonants

when forming an onset cluster. We expect this difference in timing to be stable across prosodic conditions or speech rates. Temporal variation in this study was introduced by varying sentence accent which should shorten the delay for reduced syllables to a greater degree than for onset clusters which are known to be very stable (see Byrd, 1996).

2 Method

2.1 Experimental Procedure

We recorded articulatory motions of two native speakers of British English by means of articulography (EMA AG500, Carstens Medizinelektronik, c.f. Hoole et al. 2003, Hoole and Zierdt 2010 for more detailed descriptions of the method). The system is designed to track articulatory motion over time by attaching sensors to various locations in the vocal tract. Three sensors were attached midsagittally to the tongue: the front-most sensor (TT) was positioned roughly 1cm behind the actual tongue tip, the rearmost sensor (TB) as far back as possible without creating discomfort for the participant, and the third sensor (TM) such that its distance from TM and TB was roughly equal. Articulatory data were acquired at a sample rate of 200 Hz. Position estimation was done with the help of the TAPAD toolbox (see Hoole and Zierdt, 2010). The data were low-pass filtered before and after position estimation with FIR filters (Kaiser window): First, the raw amplitude data were filtered, and, after position estimation, the same filter was applied again to the position estimations before head correction. For the reference sensors (right and left ear, bridge of nose and maxilla), the filter cutoff was set to 5 Hz, for the TT to 40 Hz and to 20 Hz for all the other sensors (including the two rear tongue sensors, lips and jaw). After position estimation, the contributions of head movement were removed from the measured data on a frame-by-frame basis. Horizontal, vertical and tangential velocities were calculated using a filter that was obtained by convolving a differentiation kernel with filter coefficients of a lowpass filter with 20 Hz passband and 30 Hz stopband edges respectively (again using a Kaiser window). Audio data were simultaneously captured at a sampling rate of 32 kHz with an AKG CK98 shotgun microphone.

2.2 Stimuli

Currently we have collected data from two native speakers of British English, with no known history of speech, hearing or language problems reported. Three series consisting of real words were recorded (S1: *balloon bloom boon loon*, S2: *police please peace lease*, S3: *collapse claps caps laps)*. The main focus is on /C_1əC$_2$/ in comparison to the complex onset condition /C_1C_2/, whereas the monosyllabic words with singleton C_1 and C_2 onsets are used as control conditions in the C-centre analyses.

Table 1. Experimental items. Material belonging to the onset is shown in bold face; the consonant used as anchor point is printed in *italics*.

Test Items		
Series	Condition	Item
b/l	Elision	**ball**oo*n*
	Cluster	**bl**oo*m*
	C_1	**b**oo*n*
	C_2	**l**oo*n*
p/l	Elision	**pol**i*c*e
	Cluster	**pl**ea*s*e
	C_1	**p**ea*c*e
	C_2	**l**ea*s*e
p/l	Elision	**coll**a*p*se
	Cluster	**cl**a*p*s
	C_1	**c**a*p*s
	C_2	**l**a*p*s

In contrast to most previous studies weakening of the neutral vowel was elicited by prosodic manipulation: The target words were manipulated by an accent variation, i.e. they were recorded in an accented condition and a postnuclearly deaccented condition. For elicitation, two sentences were simultaneously displayed on the prompting screen, with the first one read in silence, and the second - the target sentence - read aloud. For example, the screen presentation for an accented item was "What did you say? I said a POLICE again" and for an deaccented item "Was it the larger police? No, it was the NICER police again." In total, six items were recorded in each condition. In the unaccented condition, the order of the words "nicer" and "larger" that conditioned contrastive stress was balanced: We recorded three repetitions in which "larger" occurred in the (unread) prompting question and "nicer" occurred in the test sequence itself, and also three repetitions per condition in which this order was interchanged. Table 1 gives an overview over the stimulus material acquired.

2.3 Data labelling

For each constriction movement, a number of elementary temporal landmarks were extracted. They are displayed in Roman numerals in Figure 4.

CLOSING MOV ON (I)	The onset of the movement towards the constriction was extracted using a 20% velocity criterion.
CONSTR ON (II)	The beginning of the constriction phase.
MAX CONSTR (III)	The point of maximum constriction was operationalized as the maximum in the vertical displacement signals.
CONSTR OFF (IV)	The end of the constriction phase.

These elementary landmarks were used to define additional point landmarks:

C1MID/C2MID:	were defined as the middle of the closure phase of each consonant, i.e. the timepoint temporally equidistant to the CONSTR ON and CONSTR OFF landmarks of the respective constriction gestures. This particular point was chosen because it turned out that it is more stable than the point of maximum constriction.
C-centre	was in turn defined as the duration between the midpoint of the interval [C1MID to C2MID] and the anchor point.
ANCHOR:	is an alias for the constriction onset of the coda consonant (CONSTR ON).

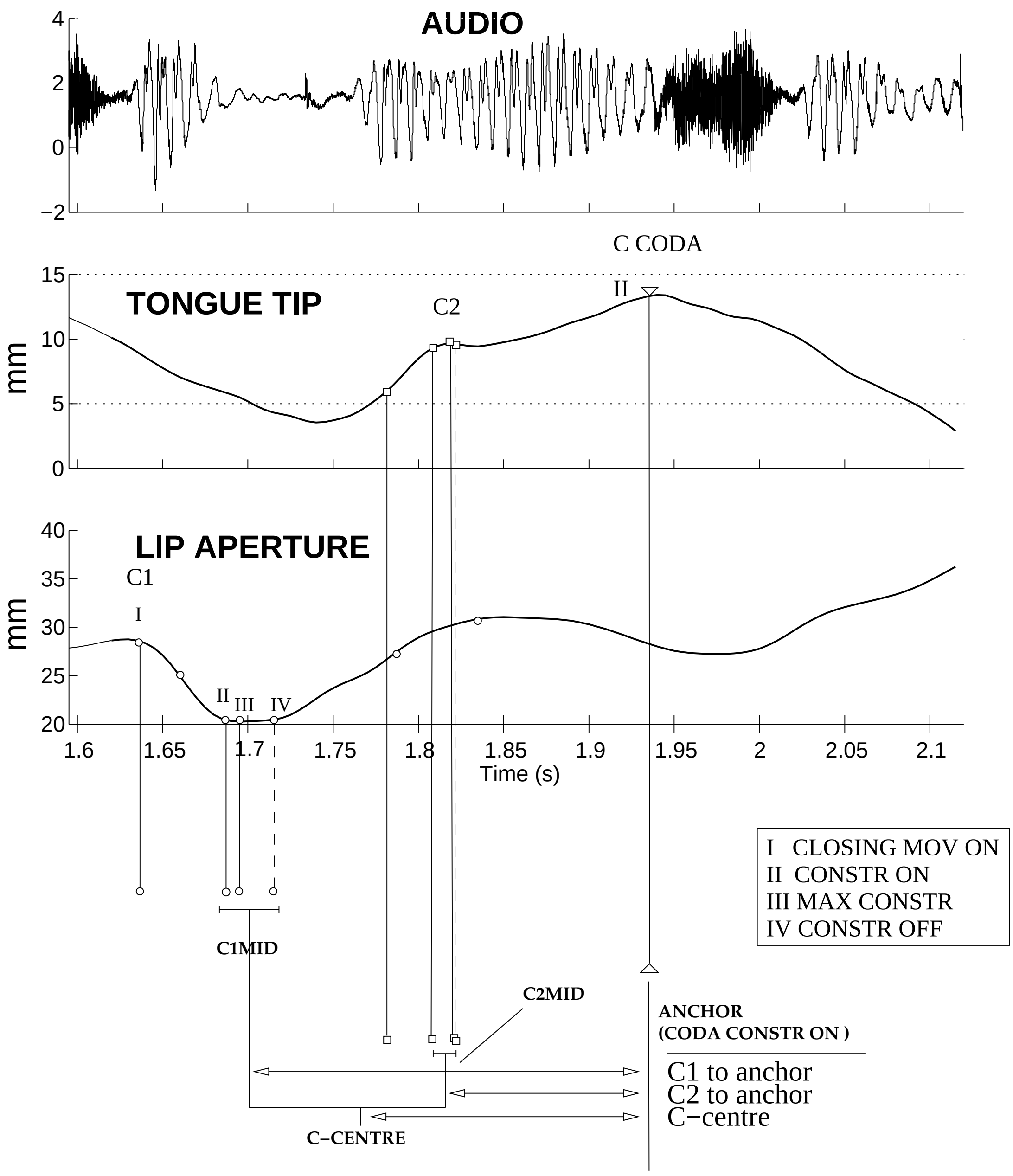

Figure 4. Illustration of the data labelling procedure an instance of the testword *police*. For each constriction, landmarks I to IV were extracted semi-automatically: CLOSING MOV ON (I), CONSTR ON (II), MAX CONSTR (III), CONSTR OFF (IV). The definitions of composite intervals are illustrated below the figures, (i) the duration between the first onset consonant and the consonantal anchor ("C_1 to anchor"), (ii) the duration between the second onset consonant and the anchor ("C_2 to anchor") and (iii) the C-centre.

These will serve to compute the dependent variables analysed in this contribution, i.e. measures of constriction overlap of initial consonants and durations of various consonantal landmarks to ANCHOR. The measure of constriction overlap analysed here is adopted from Hoole et al. (2010) and consists in the time between the end of the first consonant's constriction (C1 CONSTR OFF) and the time of onset of the nucleus of the second consonant (C2 CONSTR ON). This time interval can be analysed in absolute terms (in ms). Alternatively, it can be normalized by the time between the beginning of the first consonant's nucleus constriction (C1 CONSTR ON) and the end of the second gesture's constriction (C2 CONSTR OFF), and in the letter case yields a percentage. The interpretation of this measure is that increasingly positive values correspond to increasing overlap of C_1 and C_2, and negative values convey the information that the constriction phases show a lag between the consonants. As was addressed in the introduction, onset consonants show neither in-phase nor anti-phase timing with the following vowel but a compromise timing, called C-centre. This measure was calculated as the duration between the midpoint of the interval [C1MID to C2MID] and the anchorpoint later in the syllable. The C-centre of onset clusters should be aligned with the centre of a singleton in a CVC word.

3 Results

The following description will describe the differences between consonant sequences with and without an intervening schwa. This section is organized such that first temporal and then spatial patterns will be assessed (sections 3.1 and 3.2). Section 3.1 is further subdivided: In the first part we will describe the temporal organization of the onset consonant sequence material itself (section 3.1.1). After that the C-Centre will be introduced and discussed (section 3.1.2). Section 3.2 will investigate whether and how tongue dorsum movements differ for the two conditions.

3.1 Transcononantal timing

3.1.1 Constriction overlap

Results for the patterns of absolute constriction overlap are shown in Figure 5. As described in section 2.3, constriction overlap is defined as the time between the end of the constriction nucleus of the first consonant (C1 CONSTR OFF in Figure 4) and the time of onset of the constriction of the second consonant (C2 CONSTR ON in Figure 4). As can be seen in Figure 5, the two consonants are always separated, i.e. the two constriction target regions do not overlap. In the case of the intervening schwa the absolute lag is consistently longer than for the cluster. The accent condition did not consistently affect the patterns of constriction overlap. This finding is somewhat unexpected, since the manipulation of acccent was introduced in order to trigger schwa deletion – instead of the speech rate manipulation used in previous research on this topic. This issue will be considered again in the discussion section.
Interestingly, the variability of intergestural timing, measured as standard deviations in Figure 5, shows no tendency to be lower for the onset cluster than for the pretonic schwa syllable. This is the opposite to what could have been expected from previous studies on the behaviour of onset clusters (e.g. Byrd, 1996).
In summary, it was found that onset clusters were produced with a consistently shorter lag than consonant sequences with an intervening schwa which gives evidence against a complete homophony of the onsets of e.g. *police* and *please*.

3.1.2 Durations to anchorpoint: C-centre

This section describes observed patterns of timing of the consonantal onset material with respect to a consonantal anchor point in the rhyme. The consonantal anchor point was the rhyme consonant's constriction onset, i.e. the CONSTR ON marker. The CONSTR ON marker was chosen empirically by a stability analysis comparing candidate anchor points involving the centre of the rhyme consonant and the timepoint of the rhyme consonant's maximal constriction in addition to the CONSTR ON marker. The method of stability analysis was adopted from Shaw et al. (2009) and consisted in the the comparison of the coefficients of

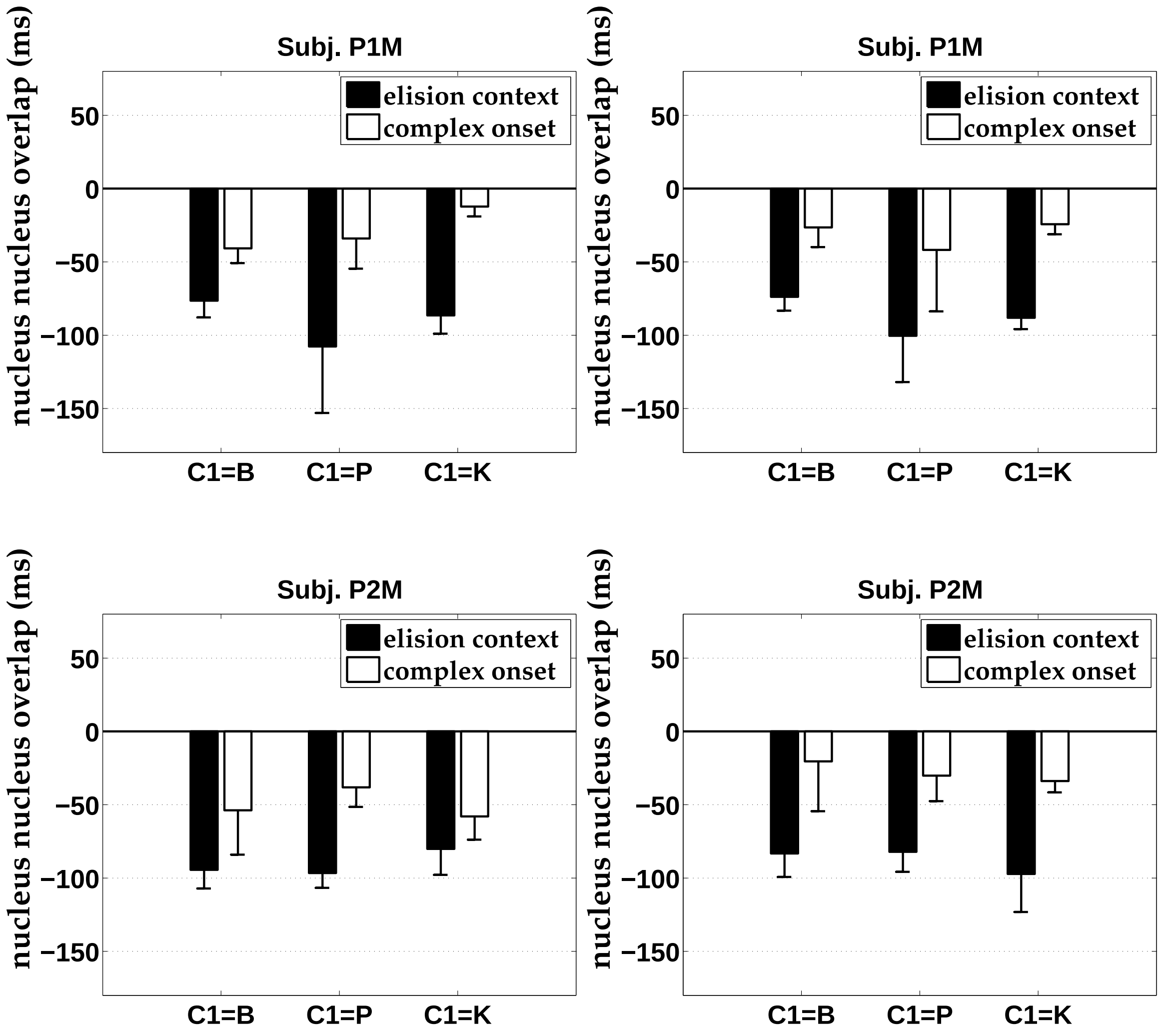

Figure 5. Plots of absolute constriction overlap of complex onset consonants and elision context respectively, both for the accented condition (left panels) and unaccented conditions (right panels). Speaker P1M is shown in the top row, speaker P2M in the bottom row. More negative values correspond to longer lags between C_1 and C_2.

variation (also called relative standard deviations) across the candidate anchors. It turned out that the constriction onset was marginally more stable according to this analysis. For singletons, the graphs in Figures 6 and 7 display the durations between the onset consonant's temporal midpoint to the anchor point in the rhyme as defined above. For elision contexts and consonant clusters, both C_1 and C_2 consonants' temporal midpoints, and in addition the C-centre are displayed relative to the consonantal anchor.

Figures 6 and 7 show the syllable timing relationships for participant P1M and P2M. Missing data for speaker P1M (C2 singletons for the test word *lease*, series *police*) are caused by annotation problems. Movement amplitudes of the tongue tip in the syllable onset consistently were too small to extract all necessary landmarks.

For the data to conform to the C-Centre hypothesis, the C-Centres in cluster contexts (shown as plus-signs in Figures 6 and 7, second row) should align with the singletons (shown as circles in Figures 6 and 7). This effect should be independent of the particular identity of the onset consonant in the singleton, i.e. C_1 and C_2 should show the same duration to the anchorpoint which means that the circles for the singletons in Figures 6 and 7 should align. For the elision context, the triangles (i.e. the second consonant) should be aligned with the singletons because C_2 /l/ is also the onset of the second syllable, at least in the absence of effects of polysyllabic shortening: An interpretation in terms of polysyllabic shortening would make two interrelated predictions. First, the interval between the second onset consonant and the anchorpoint should be shorter for the elision context than both for the cluster conditions and the singleton conditions. Second, it would make the same predictions for the C-Centre for the singletons considered alone, i.e. there should be no difference between C_1 and C_2 to anchorpoint durations. Deaccentuation should have a uniform effect of shortening the interval between the C-Centres and the anchorpoint because of the potential of the stressed vowel to undergo shortening.

The observations can be summarized as follows: The probably most eye-catching pattern in our data is that the C-Centre of the cluster and the singletons (the circles) do not align well in most cases. This is mainly because the C_2-singleton to anchorpoint durations are shorter in most cases (except in P2M's unaccented *police* series). The alignment with the

C_1-singleton is slightly better in accordance with the C-Centre hypothesis, but still the data do not warrant a fully consistent interpretation. The C-centres of the pretonic syllables precede the ones of clusters for most series. This effect is significant if a t-test is calculated over all data ($t = 3.05$, $df = 111.6$, $p < 0.01$). It fails to reach the level of significance though if speaker P1M's data are considered alone. This is due to the accented items of the *police* series and the unaccented items of the *collapse* series for which the C-centres of the elisions are actually closer to the anchorpoint than for the clusters. The patterns for speaker P2M are more consistent, but interestingly the pattern is attenuated for the same experimental condition (the accented *police* series, see Figure 7, top left panel). Still, in general this result can be interpreted as the expected failure to integrate the initial stop into the syllable in words with pretonic syllables.

A second observation is that there is a tendency for the durations between the second consonant – i.e. the /l/– and the anchorpoint - to be shorter for the lexically bisyllabic elision contexts (as denoted by the triangles) than for the other conditions. Again there are exceptions consisting in the C_2 accented clusters of *balloon* (P1M) and the *collapse* series (P2M). The shorter C_2 to anchorpoint interval for the pretonic condition in comparison to the clusters could reflect shortening of the following stressed vowel due to polysyllabic (foot) shortening (Turk and Shattuck-Hufnagel, 2000; Saltzman et al., 2008).

3.2 Spatial organization

In addition to the analysis of temporal patterns we also attempted to analyze the spatial patterns in order to compare the lingual behaviour during pretonic syllables and clusters in the different accent conditions. In order to arrive at a qualitative description, we calculated ensemble averages for the tongue mid sensors. For deaccentuated cases, we averaged over the different carrier phrases (see the stimulus description in section 2.2). The choice of the tongue mid sensor was motivated by the fact that it should, in contrast to tongue back and tongue front sensors, minimize segment-specific coarticulatory contributions from neighboring sounds. It is therefore considered the most appropriate sensor for the analysis

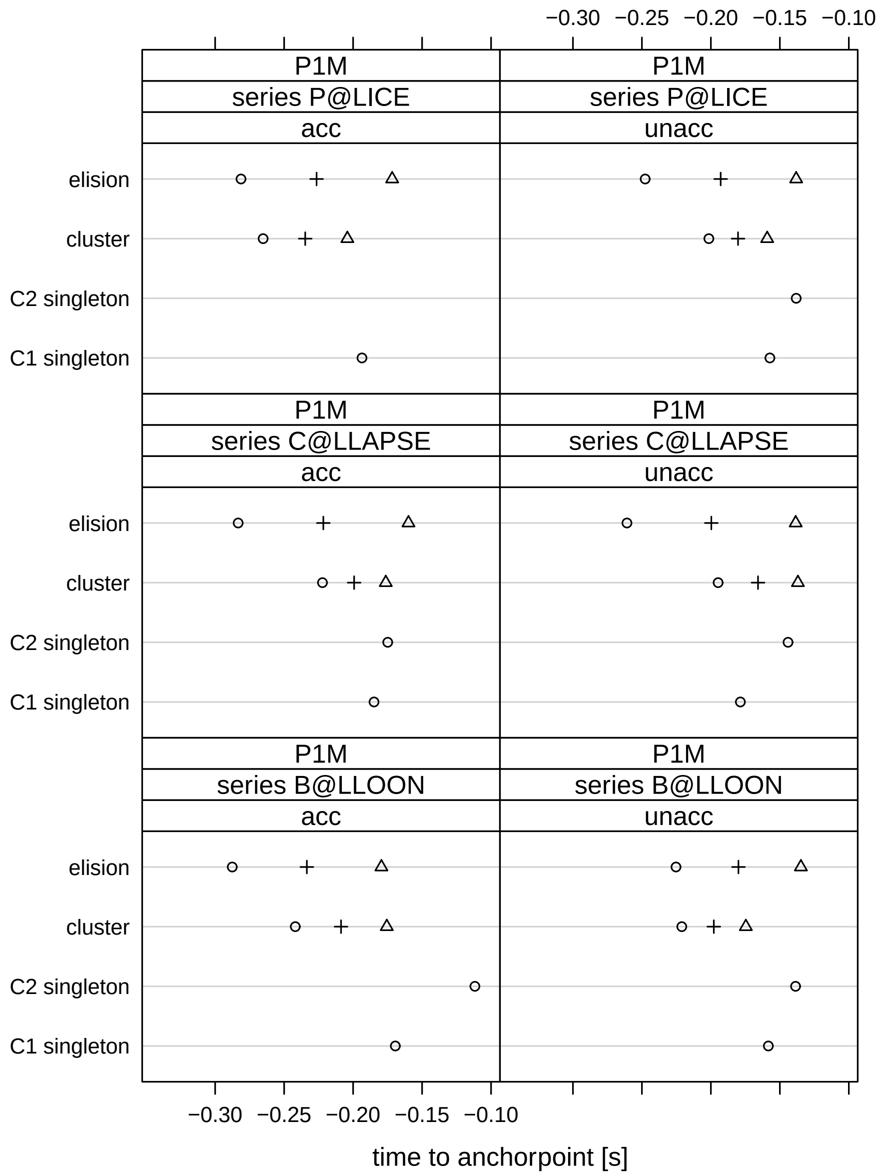

Figure 6. Lags to the rhyme anchor point for onset consonant material, speaker P1M. For singletons the analysis comprises the onset consonant's temporal midpoint. For elision context and consonant clusters, both onset consonants' temporal midpoints are displayed, and also the additional C-centre. The circles represent the initial consonants' midpoints. For elision and cluster contexts, plus-signs and triangles represent the C-center and the rightmost consonant's midpoint. Note there are missing data in the *police* series (see text for explanation).

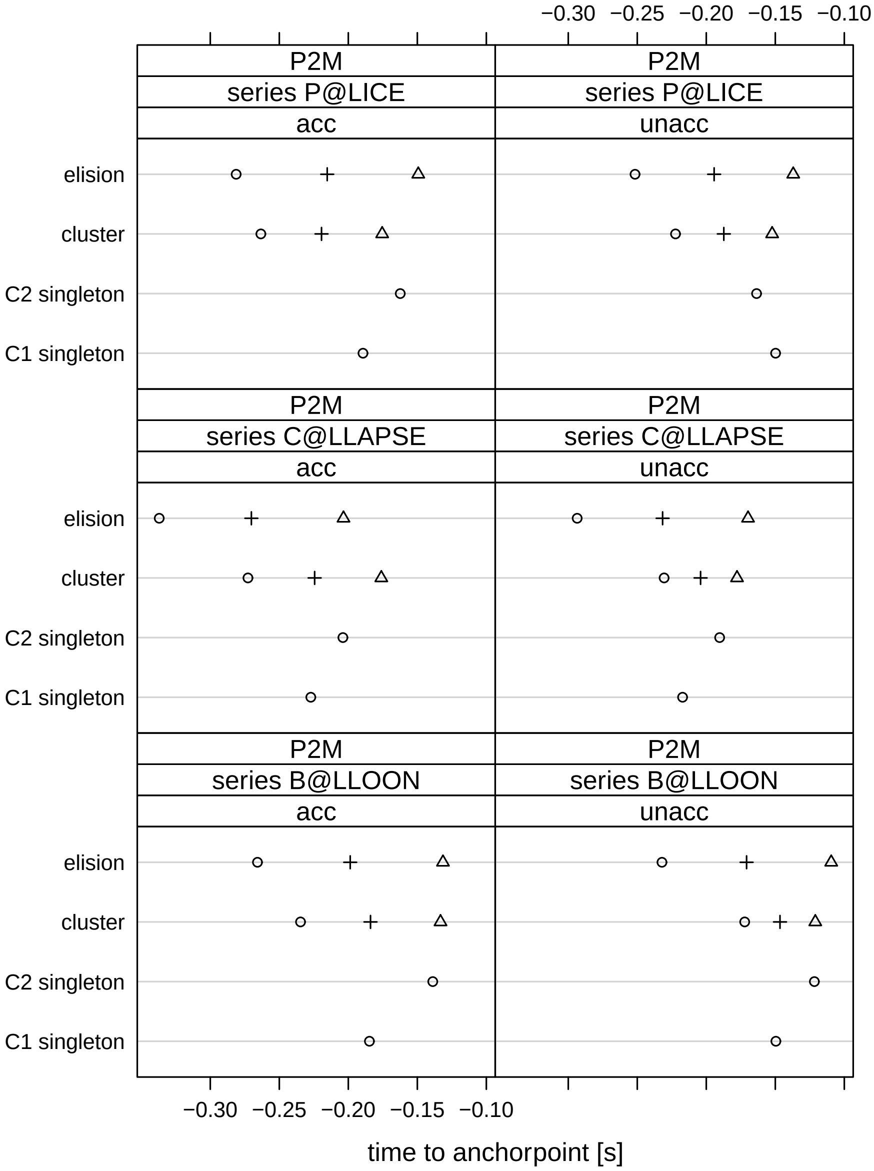

Figure 7. Lags to the rhyme anchor point for onset consonant material, speaker P2M. See Fig. 6 for details.

of any articulatory specification for schwa. The ensemble averages are shown in Figure 8 with the left panels displaying the ensemble averages for speaker P1M, and the right panels those for speaker P2M. The top panels display the *police* series, the middle panel the *balloon* series, and the bottom panel the *collapse* series. For the *balloon* and *police* series, the plotted interval for the bilabial-/l/-sequences is the interval that starts 100 ms prior to the CONSTR ON of the bilabial closing movement and ends 100 ms later than CONSTR OFF of the tongue tip movement. In addition, the interval between CONSTR ON of the bilabial movement and CONSTR OFF of the tongue tip constriction is shown in bold face; the start of the movement is shown by the capital S. Similar intervals were defined for the *collapse*-series accordingly. The main finding is that regardless of the accentuation condition there is evidence for the "targetlessness" of the schwa insofar as the visual inspection of the tongue dorsum trajectories suggests that in most cases their shape is controlled by the consonantal environment and by accentuation. Accented items usually show more extreme turning points towards the stressed vowels than deaccented items, i.e. the trajectories in accented *balloon/bloom* are lower and slightly more retracted than in deaccented items. In contrast, there are no substantial differences between items with a lexical schwa (*police*, *balloon*, *collapse*) versus the complex onset condition (*please*, *bloom*, *claps*). The fact that trajectory shape is mainly determined by the accentuation seems to confirm earlier results by Browman and Goldstein (1990a) on the targetlessness of schwa in pretonic position and suggests that in most cases the tongue dorsum trajectory in the interval between the two consonants is determined by the surrounding consonants. There are two deviances, both for P1M, the first concerns the *police* series: The unaccented elision condition does not group well with the unaccented cluster condition. Still, this does not result in a regrouping of clusters and elisions. Such a regrouping is observed for the *collapse* series though: Here, speaker P1M indeed has similar trajectory shapes for pretonic syllables and cluster conditions respectively.

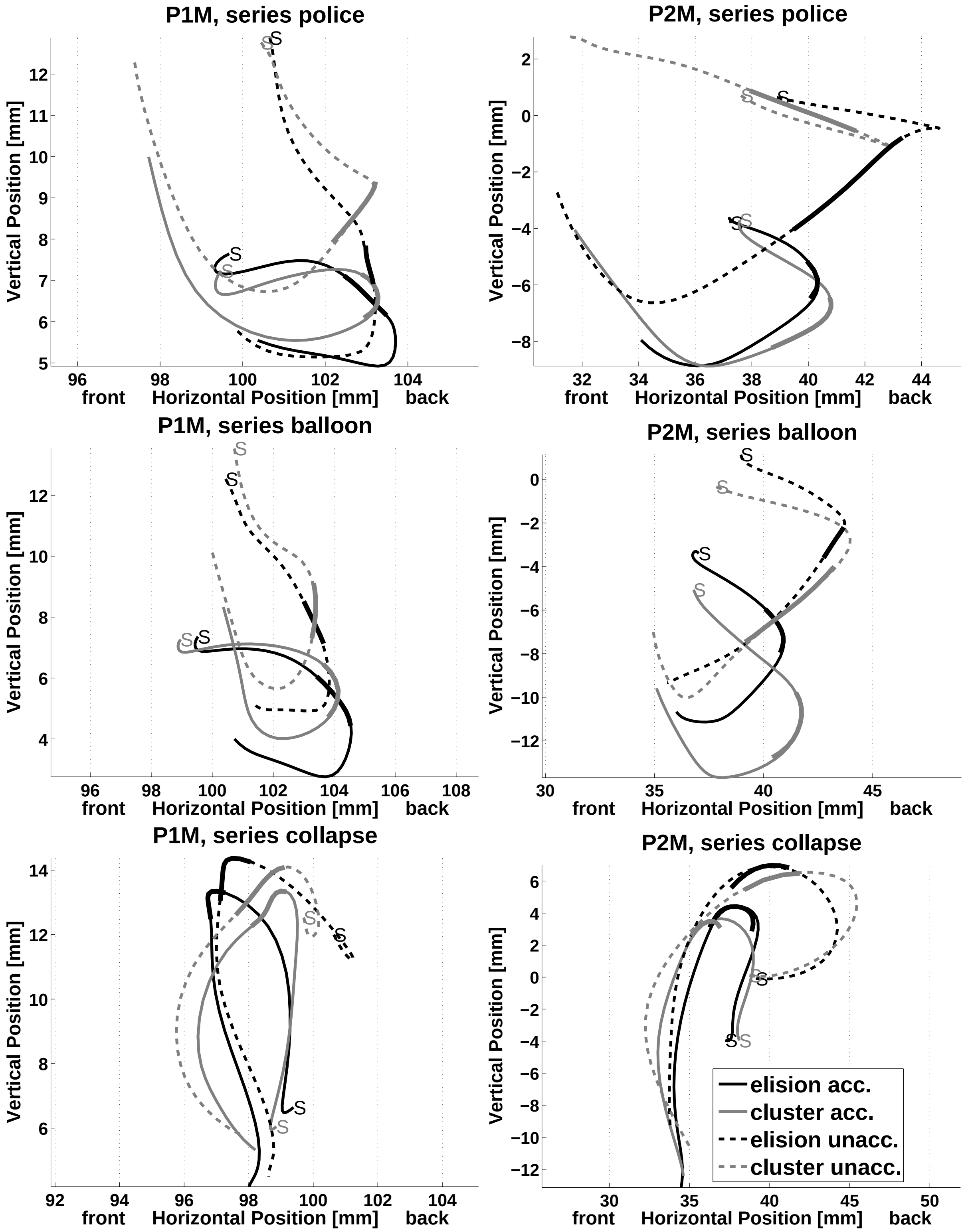

Figure 8. Ensemble Averages of tongue mid sensor for speaker P1M (left panels) and speaker P2M (right panels). The first row shows data for the *police* series, the middle panel contains the data for the *balloon* series. The *collapse* series is shown in the bottom left panels. For further explanation see text.

4 Discussion

This pilot study addressed the behaviour of complex onset clusters and pretonic reduction syllables in English. Results consisted in the findings that pretonic syllables and clusters do not converge into homophones resulting from the prosodic manipulation applied here (see section 2.2). This was shown by patterns of intergestural timing between the consonants, and the analyses of C-centres. Our speakers show a consistent lag between the consonants in the onset cluster. This finding is in agreement with the overlap pattern for 3 French speakers, reported by Hoole et al. (2010), but differs from their German speakers who produced C/l/ clusters with considerable overlap. Nevertheless, the interconsonantal lag for the pretonic case is still longer than the same lag for the cluster case. Results of the spatial patterns are less consistent for our speakers: While one speaker (P2M) consistently lacks a separate target for the pretonic schwa if the tongue dorsum sensor is considered alone, the picture for P1M is less conclusive. However, limitations of the analyses performed so far do not yield a complete picture yet. We plan to extend the analyses to other oral articulatory structures like the lips and the jaw in order to clarify whether there is a target for the schwa or not. In addition, we plan to carry out additional acoustic analysis which could be indicative of the targetlessness of the schwa in the pretonic syllables.

The results presented here are at least in partial agreement with Browman and Goldstein's (1990b) interpretation of the difference between consonant sequences in pretonic schwa syllables and in underlying clusters as a pure difference of temporal coordination. While fully consistent for one of our speakers, the results presented at least do not disconfirm such an analysis for our second participant.

In other words, in the light of the spatial data the scenario in Figure 3, i.e. the representation of the initial CV syllable without vowel seems to be better compatible with the data that we have observed, but is likely to be at odds with most speakers' intuitions. Anyway, this representation bears some similarity with the Schwa-Null-Alternation often assumed as the representation of Schwa in German Phonology (Hall, 1992, Wiese, 1988, both after Hall 2000). According to these views, the underlying representation does not contain the schwa; rather, the surface realization is obtained by epenthesis or, within the framework of Articulatory

Phonology, by a transient schwa generated by a lag between the two consonants.
Rather than furthering these maybe even accidental associations it is in our view more productive to speculate on the reasons for the slight inconsistency of grouping in the spatial patterns that have been observed. The grouping of ensemble averages, which follows by accent condition in the vast majority of cases, is broken for P1M in the *collapse* series. This finding might actually point to a shortcoming of the current experimental design: The current experiment did not make an attempt to more explicitly control speech rate or utilize an increase for eliciting more extreme reductions. In contrast to Davidson (2006) who used speech rate increases for eliciting reductions, Browman and Goldstein (1990a) and the current experiment aimed at using a more natural prosodic variation.
The absolute lag between the two consonants was not affected by accent in a consistent way. Two reasons might be responsible for this: first of all, it has been found that the timing in onset clusters is relatively insensitive to variations in boundary strength (see Byrd and Choi, 2010) as well word stress (see Bombien et al., 2010). This might at least explain why the underlying onset clusters showed very little temporal variation for deaccentuation. The second reason might apply to the pretonic case: as was shown by Turk and White (1999) for American English, deaccentuation affected stressed syllables to a much greater degree than following unstressed syllables. The effect on pretonic syllables was even smaller and also less reliable across speakers. This high degree of speaker specificity might explain why our two speakers showed much smaller effect sizes of deaccentuation in comparison to Browman and Goldstein's (1990a) speaker.
As just suggested, speech rate manipulations might be more effective for further reduction of the pretonic syllable. However, as Davidson (2006) found for American English, only some speakers showed higher elision rates when speaking faster. Furthermore, the study by Jannedy (1994, after Beckman 1996) on German gave evidence for a complete temporal neutralisation only in posttonic syllables, e.g. /n/ in **CC. *Kannen* (i.e. /kˈanən/ → kʰann̩)** approached the same duration as /n/ in *kann* only at the fastest speech rate. Similar results were found by Pompino-Marschall and Janker (1999) whose speakers retained durational dif-

ferences between *ein* and bisyllabic *einen*. These findings were further corroborated by accompanying perceptual tests. Interestingly, this tendency was less consistent for the pretonic items *beraten* versus *braten* in Jannedy (1994)'s study. Therefore, complete neutralisation of the difference in temporal coordination patterns between underlying onset clusters and consonants in pretonic syllables is not to be expected for Standard German, British English and American English.

Nevertheless, other languages or dialects, such as Bavarian, could well show the same timing in underlying clusters and the Bavarian equivalent of consonant sequences in pretonic syllables. Bavarian is of special interest for this issue because the realization of the prefixes *ge-* and *be-* as in *gsuffa* (Standard German: *gesoffen*, 'drunken' past tense) and *bsuffa* (Standard German: *besoffen*, 'drunk', adjective) is not induced by a change in speaking style or speech rate but a crucial feature of this dialect (see Rowley 1990 for North Bavarian and Wiesinger 1989). The perfect participle /ge-/ is realized as unvoiced unaspirated [k] preceding stems with initial fricatives, nasals, laterals, trills and vowels. Before stops it is completely deleted, e.g. *gekonnt* → *kenna*. Even though this latter restriction seems to imply that phonotactically illegal clusters are avoided, several counterexamples exist, such as *gefragt* → [*kfr*]*ogt*, *geschlagen* → [*kʃl*]*ong*, which do not occur as onsets in simplex words of Bavarian. Furthermore, lexicalized forms can be found frequently, such as *Klump*, derived from *ge+lumpe*, 'junk'. Similar regularities apply to the derivational prefix *be-*. On the one hand, because past tense forms in Bavarian never show a schwa, there may be a stronger tendency towards C-Center timing in these forms than in the standard language. On the other hand, the timing relationships may be sensitive to the morphological complexity of the clusters, so that, for example, lexically given /gl/ clusters may show a stronger C-center effect than /gl/ clusters that emerge from past participle formation. Anyway, Bavarian should allow the possibility of experimentally manipulating morphological complexity, and it therefore would present an interesting test case.

For British English, the difference in C-centre in our and other studies speaks against an integration of the first consonant in e.g. *police* in the stressed syllable and can be viewed as the main property distinguishing the two words from each other since we could not detect a target for the schwa. The absence of the C-centre suggests that these two word forms are distinguished by their coupling graphs, as exemplified in Figures 2 and 3. Further evidence for unsyllabified initial consonants, which are hypothesized not to be in-phase coupled directly to a vowel, were presented for the impure /s/ in Italian by Hermes et al. (2008), Moroccan Arabic by Shaw et al. (2009), and for Tashlhiyt Berber by Goldstein et al. (2010).

However, the C-centre analyzes presented in Figures 6 and 7 of the current study might have suffered from problems to generate test materials with well-defined anchor points. For example the problem in the *police* series is that the vowel in *please* may be inherently longer than in *police*, related to the voiced vs. voiceless fricative in the coda (see Chen, 1970). This effect might have shifted the whole cluster further to the left than expected. However, this does not explain the short durations in *lease*, so this effect might not ultimately be separable from the effects of polysyllabic shortening that were discussed in 3.1.2. Another potentially interfering effect depending on the choice of the anchor consonant is the contrast between *balloon* versus *bloom*, but here it remains unclear how this could have affected the results. The choice of the anchor consonant cannot be responsible for the differences in alignment between *boon* and *loon*. Still, in a future version of the corpus it would advisable to record items like *boom* and *loom* in addition. Apart from these more specific effects of corpus design, the paper by Goldstein et al. (2010) addresses the possibility that there are examples in which the same syllable structure is associated with different timing patterns depending on factors like place of articulation, manner or language studied. Rather than offering an extension of the task dynamic model that could account for deviations of such a constant C-centre, the analysis focused on whether there is the predicted left-shift of the first onset consonant and a simultaneous right shift of the second onset consonant C_2. A gross examination of our data does show a lack of right-shift for C_2 but the left shift for C_1 and the longer lag between the two consonants are extremely consistent and speak for fundamentally different timing constellations depending on syllable structure and onset complexity.

5 Acknowledgement

This research was funded by grants to the EPSRC (EP/E01609X/1, EP/-E016359/1), the NIH (NIDCD DC008780) and the German Research Council (HO3271/3-2). We also wish to thank Pascal Perrier and Susanne Fuchs for invaluable comments on a draft version of the paper.

References

Beckman, M. (1996). When is a syllable not a syllable? In Otake, T. and Cutler, A., editors, *Phonological structure and language processing: cross-linguistic studies*, pages 111–146. Mouton de Gruyter, Berlin.

Bombien, L., Mooshammer, C., Hoole, P., and Kühnert, B. (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics*.

Browman, C. and Goldstein, L. (1990a). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18:299–320.

Browman, C. and Goldstein, L. (1990b). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J. and Beckman, M. E., editors, *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, pages 341–376. Cambridge, Cambridge, UK.

Browman, C. and Goldstein, L. (1994). "Targetless" schwa: an articulatory analysis. In Docherty, G. and Ladd, D., editors, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pages 26–56. Cambridge University Press: Cambridge.

Browman, C. P. and Goldstein, L. M. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, 5:25–34.

Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology*, 13:139–169.

Byrd, D. and Choi, S. (2010). At the juncture of prosody, phonology, and phonetics – the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In Fougeron, C., Kühnert, B., d'Imperio, M., and Vallée, N., editors, *Papers in Laboratory Phonology 10: Variation, Detail and Representation*. Mouton de Gruyter.

Chen, M. (1970). Vowel length variation as a function of the variety of the consonant environment. *Phonetica*, 22:129–159.

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

Goldstein, L., Nam, H., Saltzman, E., and Chitoran, I. (2010). Coupled oscillator planning model of speech timing and syllable structure. In Fant, G., Fujisaki, H., and Shen, J., editors, *Frontiers in Phonetics and Speech Science, Festschrift for Wu Zongji*, pages 239–250. Commercial Press, Beijing.

Haken, H., Kelso, J., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51:347–356.

Hall, T. (2000). *Phonologie. Eine Einführung*. Berlin, New York: De Gruyter.

Hall, T. A. (1992). *Syllable structure and syllable-related processes in German*. Linguistische Arbeiten 276. Niemeyer-Varlag, Tübingen.

Hermes, A., Grice, M., Mücke, D., and Niemann, H. (2008). Articulatory indicators of syllable liason in word initial consonant clusters in Italian. In *Proceedings of the 8th International Seminar on Speech Production*, pages 433–436.

Honorof, D. and Browman, C. (1995). The centre or edge: how are consonant clusters organised with respect to the vowel? In Elenius, K. and Branderud, P., editors, *Proceedings of the XIIIth International Congress of Phonetic Sciences*, pages 552–555.

Hoole, P., Bombien, L., Kühnert, B., and Mooshammer, C. (2010). Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In Fant, G., Fujisaki, H., and Shen, J., editors, *Frontiers in Phonetics and Speech Science, Festschrift for Wu Zongji*, pages 281–291. Commercial Press, Beijing.

Hoole, P. and Zierdt, A. (2010). Five-dimensional articulography. In Maassen, B. and van Lieshout, P., editors, *Speech Motor Control: New developments in basic and applied research*, pages 331–349. Oxford University Press, Oxford, U.K.

Hoole, P., Zierdt, A., and Geng, C. (2003). Beyond 2D in articulatory data acquisition and analysis. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 265–268.

Hooper, J. (1978). Constrains on schwa-deletion in american english. In Fisiak, J., editor, *Recent Developments in Historical Linguisitics.*, pages 193–207. The Hague: Mouton.

Jannedy, S. (1994). Rate effects on german unstressed syllables. *Working Papers in Linguistics, Ohio State University*, 44:105–124.

Kaisse, E. (1985). *Connected Speech: The Interaction of Syntax and Phonology*. San Diego: Academic Press.

Krakow, R. A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics*, 27:23–54.

Nam, H. and Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In Solé, M.-J., Recasens, D., and Romero, J., editors, *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 2253–2256, Barcelona. Futurgrafic.

Pompino-Marschall, B. and Janker, P. M. (1999). Production and perception of syllabic [n] in German. In *Proceedings of the 14th International Congress of Phonetic Sciences (San Francisco)*.

Rowley, A. (1990). North Bavarian. In Russ, C., editor, *The Dialects of Modern German.*, pages 417–437. Routledge, London.

Saltzman, E., Nam, H., and Goldstein, L. (sub.). Intergestural timing in speech production: the role of graph structure.

Saltzman, E., Nam, H., Krivokapic, J., and Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Barbosa, P. A., Madureira, S., and Reis, C., editors, *Proceedings of the Speech Prosody 2008*, pages 175–184.

Shaw, J., Gafos, I. A., Hoole, P., and Zeroual, C. (2009). Syllabification in Moroccan Arabic: evidence from temporal stability in articulation. *Phonology (Yearbook)*, 26:187–215.

Smorodinsky, I. (2002). *Schwas with and without active gestural control.* PhD thesis, Yale University.

Turk, A. and Shattuck-Hufnagel, S. (2000). Word-boundary-related durational patterns in English. *Journal of Phonetics*, 28:397–440.

Turk, E. A. and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27:171–206.

Wiese, R. (1988). *Silbische und lexikalische Phonologie*. Linguistische Arbeiten 211. Niemeyer, Tübingen.

Wiesinger, P. (1989). *Die Flexionsmorphologie des Verbums im Bairischen.* Verlag der Österreichischen Akademie der Wissenschaften.