Schwa Elision in Fast Speech: Segmental Deletion or Gestural Overlap?

Lisa Davidson
New York University, New York, N.Y., USA

Abstract
Pretonic schwa elision in fast speech (e.g. potato → [pt]ato, demolish → [dm]olish) has been studied by both phonologists and phoneticians to understand how extralinguistic factors affect surface forms. Yet, both types of studies have major shortcomings. Phonological analyses attributing schwa elision to across-the-board segmental deletion have been based on researchers’ intuitions. Phonetic accounts proposing that elision is best characterized as gestural overlap have been restricted to very few sequence types. In this study, 28 different [#CəC-] sequences are examined to define appropriate acoustic criteria for ‘elision’, to establish whether elision is a deletion process or the endpoint of a continuum of increasing overlap, and to discover whether elision rates vary for individual speakers. Results suggest that the acoustic patterns for elision are consistent with an overlap account. Individual speakers differ as to whether they increase elision only at faster speech rates, or elide regardless of rate. Phonotactic legality per se does not affect elision rates, but speech rate may affect the phonological system by causing a modification of the standard timing relationships among gestures.

1 Introduction
Pretonic schwa deletion has attracted attention in the phonological literature as an example of how varied registers or styles allow phonotactic structures different than those found in the canonical grammar of English (e.g. potential → [pt]ential, tomorrow → [tm]orrow) [Hooper, 1978; Kaisse, 1985; Zwicky, 1972]. These accounts tend to be impressionistic, so there has been little quantitative data regarding whether deletion is equally likely in all environments, how often it occurs, and whether there is a tendency for deletion to be more frequent in environments that would lead to legal onsets [see Glowacka, 2001, for British English].

Pretonic schwa deletion has also been investigated by researchers working within articulatory phonology, who claim that many surface characteristics of casual speech production can be accounted for by changes in gestural overlap or magnitude instead of traditional rule-based phonological processes [e.g. Browman and Goldstein, 1990,
1992; Fokes and Bond, 1993; Fougeron and Steriade, 1997; Manuel et al., 1992]. For example, nasal assimilation might occur if the release of a nasal is substantially overlapped by a following stop, so that its own place of articulation is acoustically obscured. Similarly, increased overlap between adjacent sounds may effectively render one of the sounds absent from the acoustic record. For articulatory phonology, the specification of gestural overlap is considered a phonological process; should increased overlap lead to the absence or modification of a sound on the acoustic record, it would be due to changes in coordination patterns, not segmental deletion or featural change.

Pretonic schwa elision is investigated in order to examine three main issues: (a) whether ‘elision’ is actually deletion of the schwa gesture or rather substantial gestural overlap, (b) the role of individual variation in schwa elision, and (c) if elision is a result of gestural overlap, how does speech rate affect the planning of gestural coordination?

1.1 Pretonic Schwa Elision

1.1.1 Weak Vowel Deletion in Phonology

The claim that English speakers delete schwa in pretonic initial syllables (abbreviated as /#CəC-) in fast speech has been incorporated into analyses of diachronic change, syllable structure, sonority scales, and the phonology of connected or casual speech [Hooper, 1978; Kaisse, 1985; Roca and Johnson, 1999; Zwicky, 1972]. In these proposals, schwa deletion may result in the formation of both legal or illegal word-initial consonant clusters such as semester → smester or fatigue → ftigue. Generally, these accounts have not been concerned with whether the resulting cluster is found in the English cluster inventory per se. Hooper [1978] frames the question of schwa deletion in terms of a competition between a tendency to delete unstressed vowels and a pressure to maintain universal constraints on syllable structure [based on Greenberg, 1965], claiming that rapid speech provides adequate motivation for overriding syllable structure (similar to related claims that speakers prefer to minimize articulatory effort [Boersma, 1998; Flemming, 1995/2002, 2001; Kirchner, 1998/2001; Lindblom, 1963, 1990]).

Despite the general acceptance of Hooper’s claim, empirical evidence demonstrating the frequency and nature of schwa deletion in fast speech is sparse. There are two corpora studies that investigate schwa deletion in English speech. Dalby [1986] used transcriptions to examine schwa deletion in fast and slow read speech and speech from television news broadcasts. In read speech, for example, Dalby found that schwa was deleted in 2% of the tokens in slow read speech and 44% in fast read speech. A closer analysis of Dalby’s data shows that while deleted forms may contain consonant sequences not found in English, other restrictions, such as the number of consonants in the onset or the sonority profile of the clusters, mostly conform to English phonological patterns. Patterson et al. [2003] carried out a similar study of deletion rates in conversational speech from the Switchboard corpus, showing that pretonic schwa deletion occurred more often for high-frequency words (15.4%) than for low-frequency words (6.2%). Further examination of their stimuli reveals that nearly all of the candidates for pretonic schwa deletion would result in an English-legal onset cluster if the schwa were deleted (such as believe → blieve).

Other research has suggested that perhaps schwa elision resulting from a phonological deletion rule should be questioned regardless of the resulting phonotactics.
Recent work has shown that processes traditionally thought of as a result of a phonological rule, such as nasal assimilation or sandhi, are very much like processes that are considered phonetic, such as coarticulation or nasalization [e.g. Cohn, 1993; Ellis and Hardcastle, 1999; Flemming, 2001; Zsiga, 1995, 1997]. Along those lines, pretonic schwa deletion may perhaps be characterized as the extreme endpoint of a phonetic reduction process [Beckman, 1996; Browman and Goldstein, 1990]. One characteristic of casual speech or faster speech rates is greater overlap of sequential gestures [Byrd and Tan, 1996]. If this is the case, then apparent deletion may actually reflect overlap which obscures the schwa, even though it is actually produced. This issue is explored in the next section.

1.1.2 Acoustic and Articulatory Evidence: Elision as Overlap

Acoustic studies of schwa elision in /#CəC/- sequences have suggested that in cases where the schwa appears impressionistically to be deleted, there is still phonetic evidence on the surface, indicating that the form contains a vowel at some level of analysis [Fokes and Bond, 1993; Fougeron and Steriade, 1997; Manuel et al., 1992]. For example, in French, Fougeron and Steriade [1997] used electropalatography to examine differences between de rôle [dəʁol] ‘of role’, the elided form d’rôle [dʁol], and the monomorphemic drôle [dʁol] ‘funny’. Results indicated that for all participants, the [d] in d’rôle contained significantly longer linguopalatal contact, a longer lingual occlusion, and was less likely to be lenited than the [d] in drôle. The acoustic examination of Manuel et al. [1992] of the status of schwa in support as compared to speakers’ production of sport in casual versus careful speech indicated that while some tokens of support in casual speech provided no acoustic evidence of a schwa, other utterances contained either a vowel, a voice bar, or aspiration between the [s] and the [p]. Both Fougeron and Steriade [1997] and Manuel et al. [1992] concluded that the glottal gesture for the underlying schwa is present in casual speech, and that speakers also retain the same glottal-oral timing found in careful speech.¹

Beckman [1996] attempts to provide an analysis of these types of findings in terms of gestural coordination. Following similar analyses in Browman and Goldstein [1990], Beckman [1996] argues that apparent deletion of schwa in various environments in English could be the extreme endpoint of a continuum of vowel reduction and overlap. If so, she notes that ‘the continuously variable values of overlap in the gestural score are a better representation than a categorical phonological rule of schwa deletion’ (p. 100). In addition to English schwa deletion, Beckman [1996] also discusses similar types of phenomena from Japanese, Korean, and Montreal French.

Gestural overlap can be schematized as in figure 1. In articulatory phonology, the representation of a gesture not only specifies the active articulator and the location of constriction, but also the temporal duration of the gesture and how two sequential gestures are coordinated with respect to one another. The timeline of a gesture is shown in figure 1a, and overlapping coordination is shown in figure 1b. Because consonants and vowels typically contain more than one gesture (e.g. /p/ is a labial gesture plus a laryngeal opening gesture), the term ‘constellations’ will be used to refer to the gestures that

¹ A reviewer points out that a drawback of these studies is that speakers may have been aware that they were producing minimal pairs, and therefore may have articulated them distinctly. This problem does not arise in the current study, which does not use minimal pairs and embeds the target words in stories.
comprise a segment. The schematics in figure 1 represent the duration of the oral gesture of a phoneme.

Neither the traditional phonological accounts of schwa deletion nor empirical studies of schwa elision have addressed whether it interacts with English phonotactic restrictions [e.g. Hammond, 1999]. Specifically, it is plausible that while deletion may occur in many types of /#CəC-/ environments, those environments which would result in phonotactically legal onset clusters might show higher rates of deletion. This is a secondary issue addressed by this study.

1.2 The Influence of Speaking Rate on Gestural Coordination

The effect of speaking rate on the coordination relationships between different types of gestures is important because it sheds light on the way that gestural patterns can be manipulated. To the extent that “paralinguistic” factors may influence phonetic and/or phonological regularities [Gafos, in press; Liberman, 1983], examining the changes that occur as a result of speech rate (or registers of formality, or sociolinguistic variables such as group membership) may indicate what types of modifications are possible and how they are implemented.

Several variables that are affected by speech rate have been noted in the literature, including segmental duration [Gay, 1981], velocity of the articulators [e.g. Adams et al., 1993; Gay, 1981; Gay et al., 1974], and reduction in the magnitude of articulatory gestures produced [Byrd and Tan, 1996; Shaiman, 2001]. Another variable relevant to the study of pretonic vowel deletion in fast speech is the amount of gestural overlap found at various speech rates, and whether it changes as a function of rate.

Several studies have provided both acoustic and kinematic measures indicating that rate does induce greater overlap in various environments [Boyce et al., 1990; Byrd and Tan, 1996; Engstrand, 1988; Munhall and Løfqvist, 1992; Shaiman, 2001; Solé, 1995; Tjaden and Weismer, 1998; Zsiga, 1994]. For example, in their study of voiceless consonants at the word boundary in the phrase kiss ted, Munhall and Løfqvist [1992] found that at slow rates, each consonant retained its own glottal opening gesture, but only one glottal opening motion was observed at quicker rates of speech. They conclude that the single opening gesture results from greater consonantal overlap and blending of the glottal gestures.

**Fig. 1.** a Timeline of a gesture. First, the motion of the articulator is initiated toward the constriction location (onset), the motion reaches the intended articulator position (target), holds that position for a period of time (until the release), and is then moved away to finish the production of the sound (offset). b Overlapping gestures. The articulation of the /p/ begins before the constriction for the /k/ has been released. The articulator schematized for /k/ represents the duration for the tongue dorsum, while the schematic for /p/ represents the duration of the closure of the lips. The overlap of the release of the first consonant by the target of the second consonant means that the first consonant will not be audibly released.
While some studies present similar conclusions regarding increased overlap as a result of faster speech rates [Byrd and Tan, 1996; Shaiman, 2001; Tjaden and Weismer, 1998; Zsiga, 1994], other studies report on substantial variability among speakers regarding how speech rate affects overlap [Adams et al., 1993; Kuehn and Moll, 1976; Lubker and Gay, 1982; Matthies et al., 2001; Nolan, 1992; Shaiman et al., 1997]. Tjaden and Weismer [1998] point out that such wide variability has two consequences. First, it is risky to make conclusions based on data from one or two speakers, since two different speakers may use entirely different mechanisms for implementing fast versus slow speech. Second, the fact that there may be numerous mechanisms employed by speakers to vary speech rate is interesting in itself and deserves further study. Individual speaker variability will be discussed in the context of the experimental data from section 2.

In the next section, an experiment investigating pretonic schwa elision in fast speech in a large number of phonotactic environments is presented. It is hypothesized that elision likely results from one of two causes: (1) either speakers delete the schwa on the surface, or (2), they produce an overlap between the consonant gesture and the adjacent schwa, which may give rise to apparent deletion in the acoustic record. A metric is developed to begin to tease apart the acoustic correlates of deletion from those for gestural overlap. Following the basic tenets of articulatory phonology, it is assumed that gestural coordination is phonologically specified, and that speech rate or other extralinguistic variables can affect the specification and implementation of gestural coordination. In section 3, the experimental findings are discussed in terms of how speech rate may affect the coordination of gestures in English.

2 Experiment

In this study, a large number of /#C1əC2-/ sequences are investigated in order to explore the nature of schwa deletion in English. Two distinct measures are applied in this study to garner the most information about the process underlying deletion (or overlap): elision, which refers to the complete absence of voicing, formant structure, and aspiration which may signal a voiceless vowel, and aspiration/devoicing of the schwa, which takes into account the cases where there is no voicing or voice bar, but where there is a substantial period of aspiration or devoicing. Note that this use of ‘elision’ is not meant to have any theoretical import; it is simply meant as a descriptive term to denote the lack of acoustic vocalic information. The results of the elision measure and the aspiration/devoicing measure will be discussed in sections 2.5.2 and 2.5.3, respectively.

2.1 Participants

The participants were 9 Johns Hopkins University undergraduates who received course credit for their participation. All of them were native speakers of English. There were 3 male and 6 female participants. No speaker reported any hearing or speech impairments.

2.2 Materials

Target sequences in this experiment are words with an initial pretonic /Cə/ syllable (i.e. semester, petitioned, Venetian) where elision of the vowel would lead to the creation of an initial two-member
consonant cluster. The set of target words contains 3 tokens each of 28 different /#C1 ə C2-/ sequences, for a total of 84 words. The target words were incorporated into 4 reading passages. The passages ranged in length from 200 to 400 words, and each passage contained between 13–27 target items. There was no effort to control the prosodic position of the words, as there are no a priori assumptions in this study as to what environments lead to more or less deletion. Instead, this is a general study examining overall rates of schwa deletion. The stories are given in Appendix 1.

The target words can be grouped into categories for analysis based on both phonetic and phonological hypotheses regarding the consonants in the /#C1 ə C2-/ sequence. On both aerodynamic and perceptual grounds, it is hypothesized that there should be differences in deletion rates depending on whether a sequence is stop-initial or fricative-initial. Because the burst of the stop and formant transitions into a following sonorant are critical cues for identifying a stop [e.g. Blumstein and Stevens, 1978; Dorman et al., 1977; Kingston, 1990], it is hypothesized that schwa is less likely to be deleted in this environment. Thus, the first category is (voiceless) stop-initial.

For fricatives, it is hypothesized that there will be more deletion relative to stops. This is because most fricative-initial target words begin with /s/, which is a common first member of initial clusters. In addition, fricatives have more internal cues during the constriction phase, so they are more acoustically identifiable than stops [Kingston, 1990; Wright, 1996], making a following vowel less important for recognizing the consonant. Thus, the second category is (voiceless) fricative-initial.

It is also hypothesized that the voicing of the first consonant could influence whether the schwa is deleted. Because initial voiced obstruents are not aspirated, the aerodynamic conditions for the schwa to be devoiced are not present. Consequently, there should be less deletion in the third category, voiced-initial.

The fourth category, /l/-second, is based on the hypothesis that elision should be greater when the second consonant of the /#C1 ə C2-/ sequence is /l/. Like /s/-initial clusters, /l/-second clusters are legal in English, and this could affect deletion rates. The fricative-initial and /l/-second categories include sequences that would lead to both legal (e.g. /#səm- → /#sm- and /#bəl- → /#bl-/) and illegal clusters (i.e. /#səb- → /#sb- and /#bəl- → /#dl-/). In addition, since /l/ is an approximant, it is a better second segment for an initial consonant to release into [Côté, 2000; Kingston, 1985; Steriade, 1993].

A priori, it is unknown if the potentially illegal sequences are likely to show elision, or if elision will be blocked because a nonlegal cluster would result. Initially, they will be grouped with other fricative-initial and /l/-second sequences, but individual clusters will be further examined in section 2.5.4. The categories established for analysis are shown in table 1. The number of sequences in each category is not equal, because one of the main goals of the experiment was to obtain a broad cross section of /#C1 ə C2-/ sequences that could result in deletion. Rather than limit the number of sequences in some categories, it was decided that it would be better to have an unequal number. The actual words used in the experiment are given in Appendix 1.

2.3 Design and Procedure

Participants were recorded in a quiet room using a Sony ECM-717 microphone and a Sony MZ-R50 MiniDisc recorder. They were given the 4 reading passages each on a separate sheet of paper plus

### Table 1. Experimental /#C1 ə C2-/ stimuli

<table>
<thead>
<tr>
<th>Categorization criteria</th>
<th>Sequences used</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Voiceless) stop-initial</td>
<td>kap-, par-, pat-, tab-, kam-, tam-</td>
</tr>
<tr>
<td>(Voiceless) fricative-initial</td>
<td>far-, səb-, sap-, səf-, səv-, səm-</td>
</tr>
<tr>
<td>Voiced-initial</td>
<td>bag-, dab-, dav-, dan-, vən-, əd-, əp-, baʃ-, daʃ-</td>
</tr>
<tr>
<td>/l/-second</td>
<td>gal-, dəl-, mal-, bal-, səl-</td>
</tr>
</tbody>
</table>

1The sequence /səl-/ was placed into the /l/-second category in order to make the sizes of the categories more equal. This is an arbitrary decision, but it will be shown in the results that it has no crucial effects on the outcome.
a practice passage and told to read them out loud first at a normal reading rate and then a second time, as fast as they could without making mistakes. The order of presentation of the passages was randomized for each participant.

2.4 Analysis

The acoustic data for each participant were digitized at 22 kHz and analyzed using Praat for Windows. In order to ascertain the presence or absence of a schwa in /\#C_1\#C_2/ sequences, the waveform and spectrogram of each target word were examined. For each sequence, 5 duration measurements were taken: the entire /\#C_1\#C_2/ sequence, C_1 burst and aspiration of C_1 (if present), voiced vowel (if present), and C_2. A few points about the measurements should be made. (1) Since these sequences were presented in running texts, the beginning of C_1 when C_1 was a stop was measured starting immediately at the offset of the preceding sound (most often a vowel or a nasal). Otherwise, measurement started at the onset of frication or the nasal formants. (2) The start of C_2 was measured from the offset of the last glottal pulse of the schwa, if there was one; at the end of the aspiration on the preceding stop, if there was aspiration instead of a voiced vowel, or at the end of the silence, frication, or nasal formant of C_1 if there was no vowel present at all. Accounting for the possibility of a devoiced vowel after a fricative will be discussed further in section 2.5.3. (3) The presence of a schwa before /l/ (as in baloney) was determined on the basis of intensity. When a schwa is present in this environment, there is an abrupt decrease in intensity between the last glottal pulse corresponding to the schwa and the first pulse of the /l/. This boundary, usually most obvious in the third and fourth formant, was taken to be the end of the schwa. If no change in intensity was present, the token was coded for elision. (4) Any part of the interconsonantal interval which included a voice bar and/or formant structure was considered part of the vowel. This is consistent with the claim that the voice bar in this environment is a glottal gesture of vowel retention [Ladefoged and Maddieson, 1996]. Total vowel elision was counted only when both the vowel and C_1 aspiration were absent, since aspiration could indicate a devoiced vowel [see also Patterson et al., 2003, for a similar concern]. These measurements are exemplified in the spectrograms in Appendix 2 (which are also used to illustrate points made in section 2.5.3).

The analysis of the results focuses on two measures: elision and aspiration/devoicing. Again, a token is considered to show elision when there is a complete absence of voicing, formant structure, and aspiration which may signal a voiceless vowel. Responses are coded for aspiration/devoicing of the schwa when there is no voicing or voice bar, but a substantial period of aspiration or devoicing follows C_1. This situation has consequences for overlapping coordination, which are discussed in section 2.5.3.

2.5 Results

2.5.1 Speaking Rate

Although a specific fast speaking rate was not induced by any external method other than telling the speakers to read as fast as possible without making errors, an increase in rate in the fast condition was verified by an analysis of variance (ANOVA). In this and all future ANOVAs, participants are treated as a random factor, and type III sums of squares are used to handle unbalanced cells. The dependent variable was the duration of the entire /\#C_1\#C_2/ sequence, and the independent variable was the story rate condition (slow and fast). The mean duration for /\#C_1\#C_2/ sequences in the slow condition was 183 ms, and the mean duration in the fast condition was 160 ms. Results confirmed a significant difference between duration in the slow and fast conditions [F(1, 8) = 77.03, p < 0.001].

2.5.2 Elision

The data for elision in slow versus fast speech were examined using a repeated measures ANOVA. The independent variables were sequence type (voiceless
stop-initial, voiced-initial, fricative-initial, and \(/l/-second, as shown in table 1) and rate (slow and fast). The dependent variable is the mean proportion of elision for each sequence type, which has been arcsine transformed (as have all dependent variables that are proportions in the following analyses).

Mean proportion of elision for sequence type by rate is shown in the graph in figure 2. Results from the ANOVA show a main effect of sequence type \([F(3, 24) = 11.95, p < 0.001]\), but no main effect of rate \([F(1,8) = 1.70, p = 0.23]\). The interaction between sequence type and rate is not significant \([F(3, 24) = 2.48, p = 0.09]\).

With respect to overall deletion patterns across categories, a Student-Newman-Keuls post hoc test indicated that fricative-initial and \(/l/-second sequences show significantly more elision (21 and 19%, respectively, collapsing over rate since there was no main effect of rate) than voiced stop-initial (2%) or voiceless stop-initial (3%) \((p < 0.05)\). Neither fricative-initial and \(/l/-second sequences nor voiced stop-initial and voiceless stop-initial sequences were significantly different from one another. This suggests that while voicing does not have an effect on elision in stop-initial sequences, being fricative-initial or \(/l/-second does. The role of rate will be discussed again in section 2.5.5 when individual speakers are examined.

In this section, the frequency of elision was investigated. In the next section, we explore what acoustic manifestations an overlap account would predict. In addition, acoustic properties beyond the absence of a vowel in the sequences which have been coded as elided are revisited to see if there are any acoustic indications that suggest a schwa may actually have been produced. By doing so, we can begin to address whether the extreme endpoint of increased overlap is apparent deletion.

2.5.3 Acoustic Manifestations of Overlap: Aspiration/Devoicing, Frication, and \(/l/-Duration

To determine whether speakers’ utterances include deletion or are better characterized as gestural overlap, elision is not the only relevant variable that must be measured. That is, deletion of the schwa should lead to the production of a well-formed consonant cluster which has particular acoustic characteristics beyond the absence of a vowel between the consonants. This should be especially evident when deletion would lead to the production of a well-formed consonant cluster, such as \(/sp/- or \(/bl/- For other
sequences, such as stop-initial #CəC, we may not know a priori what a well-formed cluster may be for an English speaker, but it is possible to make some predictions based on information such as the standard voice onset time (VOT) of American English, and how consonant clusters are produced word-medially.

Furthermore, if a single articulatory maneuver like gestural overlap results in a different acoustic manifestation depending on the phonotactic environment, then elision may not be the appropriate measure for each type of experimental sequence. More specifically, if overlap of the consonant and schwa constellations occurs across the board for all sequence types either for fast speech or simply in a speaker’s habitual speaking register, it would likely have a different acoustic consequence for different flanking-consonant environments.

First, in the case of voiceless stop-initial sequences, there is very little elision as defined in section 2.4. A token is only coded for elision if there is no release of C₁ (as would be the case for word-medial sequences) [Catford, 1977; Henderson and Repp, 1982] or only a burst following the closure portion of C₁. However, if there is increased overlap of the stop and the schwa, it may lead to devoicing of the vowel, or the appearance of aspiration after the stop that is more substantial than the aspiration normally found for a stop preceding an unstressed syllable. This would occur because the duration of the glottal opening gesture associated with the initial voiceless stop extends after the stop is released, so if the stop and schwa constellations were overlapped, the production of the laryngeal opening gesture may overshadow or even preclude the production of a voiced schwa. Consequently, we do not expect elision to be frequent, since that measure does not include aspiration/devoicing, but the proportion of cases showing sizeable aspiration/devoicing (but no voice bar) should be high. The distinction between ‘normal’ (or canonical) production and overlapping gestures is illustrated in figure 3.

In the following figures, the solid lines indicate consonants, and the dotted line schematic indicates a vowel. The box underneath the consonant represents the opening of the glottis relative to both consonantal and vocalic gestures. The vertical dotted line marks the consonant-vowel alignment. The boxes underneath the gestural schematics indicate how they would be produced acoustically. Illustrative spectrograms for the conditions in figures 5–8 are shown in Appendix 2.

If there is considerable overlap when the initial consonant is voiced, then there would likely still be a remnant of the schwa present in the acoustic record even if its duration is reduced. This is because there is no glottal abduction gesture eclipsing the production of the vowel, so unless the schwa is completely overlapped, a portion of it should remain on the surface. For this sequence type, it is predicted that elision should be infrequent, and there should be few or no cases with aspiration/devoicing since vocal fold vibration can be maintained from the consonant to the schwa. A schematic for the overlapped voiced-initial case is shown in figure 4.

Considerable overlap when the first consonant in the /#CəC/- sequence is a voiceless fricative may appear as an extended fricative. Since both fricatives and aspiration are characterized by aperiodic noise, there may be little difference in visual appearance between them on a spectrogram. If the glottal abduction gesture for the /s/- which is typically longer than the one for a voiceless stop like /p/ [Yoshioka et al., 1981] - eclipses the schwa and the oral gesture of /s/ overlaps the oral gesture for the schwa, then the surface form may simply resemble a longer /s/. This overlap explanation would predict that the duration of the /s/ when the schwa appears to be deleted should
be longer than when the schwa is still present. If the schwa has instead actually been deleted and an onset cluster has been formed, it would be expected that the /s/ should be shorter than in the tokens where schwa is retained, since /ls/ in onset clusters is typically shorter than singleton /s/ [Crystal and House, 1988; Klatt, 1974]. In addition, if the schwa is not deleted and an initial /ls/-cluster is not formed, then it is expected that there would be aspiration after a C2 that is a voiceless stop. Thus, it is predicted that elision should be frequent and aspiration/devoicing should be infrequent, since the distinction between the fricative and a devoiced vowel may be obscured. The duration of /s/ when /ə/ is elided should be longer than when it is not, and aspiration should be present on the following stop.2 The schematic for overlapping /s/ and /ə/ and a following voiceless stop is shown in figure 5.

Finally, the case of /l/-second sequences is a little more complicated. Except for the /#səl/-token, all of the initial consonants in this category are voiced. Thus, if the consonant overlaps the schwa, a remnant of the vowel should still be present on the surface. Yet, if the portion of the schwa remaining in the acoustic signal is very short, it may be virtually indistinguishable both perceptually and visibly (on the spectrogram) from the following /l/. Since /l/ is vowel-like and contains formant structure that is influenced by coarticulation, it would not be surprising if the remaining schwa portion could not be distinguished from the following /l/ [Ladefoged and Maddieson, 1996]. In this sequence, it is predicted that elision will be frequent, since a schwa may be confused with the acoustic characteristics of /l/, and aspiration/devoicing should be essentially nonexistent since almost all C1s are voiced. However, if the speakers are actually overlapping C1 and

---

2 A reviewer points out that the overlap account could be bolstered by examining the spectrum for /ls/. If the last part of the frication is actually an overlapped /ə/, it might be expected that the spectrum of frication would have peaks below 4,000 Hz. However, because nearly all of the /ls/-initial sequences are followed by a labial C2, the cause of lower spectral peaks in this particular case is ambiguous. That is, it would be difficult to tell if lower spectral peaks were occurring because the /ls/ is overlapping the /ə/, or because the /ə/ was actually deleted and the /s/ was coarticulating with the labial, causing a “labial tail”. An example is shown in Appendix 2e.
the schwa and schwa is being measured as part of the /l/, then the duration of /l/ when /ə/ is elided should be longer than when it is not. This is shown in figure 6.

Given these scenarios for different /#C₁əC₂-/ sequences, it becomes clear that the elision measure used above may be misleading. Elision implies that if there is no voicing, aspiration, or formant structure present between the two consonants, then the schwa is not present either. However, it is now evident that even if none of these characteristics are found on the acoustic record for fricative-initial and /l/-second sequences, the schwa constellation, although overlapped, may still exist in the speaker’s output and may affect other factors, such as aspiration on the second consonant or duration of the fricative or /l/. The predictions based on these expectations about the surface features of overlapping for the different categories of /#C₁əC₂-/ sequences are summarized in table 2.

To confirm whether these predictions are upheld, it is necessary to examine how often the interconsonantal interval in the participants’ productions of the /#C₁əC₂-/ sequences consisted of aspiration/devoicing. The proportion of productions with only aspiration/devoicing can be compared to those with elision and then assessed with respect to the predictions in table 2.

The data for aspiration/devoicing in slow versus fast speech were examined using a repeated-measures ANOVA. The independent variables were sequence type (voiced-initial, voiceless stop-initial, fricative-initial, and /l/-second, as shown in table 2) and rate (slow and fast). The dependent variable is the proportion of tokens containing aspiration/devoicing. Tokens were positively coded as being aspirated/devoiced only if they did not contain any voicing corresponding to a schwa. Aspiration in stop-stop sequences was taken to be the interval from the end of the burst of C₁ to the onset of silence for C₂. For voiceless stop + fricative sequences like /#pəθ-/ , the aspiration after the /p/ was measurable only if there was a clear distinction between the intensity and spectral information for aspiration as opposed to the following fricative (e.g. /θ/ was generally weaker in intensity than the aspiration, as in the example for pathetic in
Appendix 2b). Cases which have neither aspiration nor a schwa are not included in this measure, since they were included in the elision count.

Mean proportion of aspiration/devoicing for sequence type by rate is shown in the graph in figure 7. Results from the ANOVA show a main effect of sequence type \[F(3, 24) = 71.69, p < 0.001\], and a main effect of rate \[F(1, 8) = 6.66, p < 0.03\]. The interaction between sequence type and rate is significant \[F(3, 24) = 3.34, p < 0.04\].

Collapsing over rate, a Student-Newman-Keuls post hoc test indicates that only the voiceless stop-initial sequences have significantly more aspiration/devoicing than the other sequences \((p < 0.05)\), which are not significantly different from one another \((p = 0.50)\). A planned comparison of the voiceless stop-initial categories shows that there is only marginally more aspiration/devoicing in the fast condition \[F(1, 8) = 4.43, p = 0.068\].

A number of other acoustic measurements were also taken to investigate whether the acoustic output is consistent with the coordination patterns outlined in figures 5–8. First, the duration of aspiration following \(C_1\) in voiceless stop-initial stimuli was compared for two types of stimuli: those which had no evidence of a voiced vowel following \(C_1\), and those which did. It was hypothesized that if a schwa is still present in the production but appears as devoicing or aspiration on the acoustic record, then aspiration when there is no voiced vowel should be longer than when a voiced vowel is acoustically present. This was tested with a univariate ANOVA with voiced vowel presence and rate as the independent variables and aspiration duration as the dependent variable. Results show that there are significant main effects for both vowel presence

### Table 2. Predictions for coding the acoustic output of overlapped /C/ and /ə/

<table>
<thead>
<tr>
<th>Sequence type</th>
<th>Likelihood of coding as elision</th>
<th>Likelihood of coding as aspiration/devoicing of (C_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiced-initial (/#dəv/)</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Voiceless stop-initial (/#pət/)</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Fricative-initial (/#fət/)</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>/l/-second (/#bəl/)</td>
<td>high</td>
<td>low</td>
</tr>
</tbody>
</table>

![Fig. 7. Aspiration/devoicing by sequence type.](image)
[F(1, 8) = 63.94, p < 0.001] and rate [F(1, 8) = 5.05, p < 0.04], but no significant interaction [F(1, 8) = 1.32, p = 0.28]. Duration measurements, shown in figure 8, confirm that there is significantly more aspiration when a voiced schwa is not present.

In order to determine whether the prediction about the duration of frication is upheld, the duration of /s/ in sequences with schwa versus CC sequences resulting from schwa elision is examined. The duration of /s/ for both slow- and fast-speech elisions are calculated. The number of tokens for each condition is summarized as follows: slow unelided (n = 115), slow elided (n = 16), fast unelided (n = 91), fast elided (n = 40). Duration measurements are shown in figure 9.

A t test collapsing over speech rate (the third column of the graph in figure 9) indicates that the duration of /s/ is significantly longer in tokens with elision than in tokens which exhibit the vowel [t(15) = 2.18, p < 0.05].

A similar measurement can be made for /l/-second tokens. The number of tokens for each condition is summarized as follows: slow unelided (n = 108), slow elided (n = 25), fast unelided (n = 107), fast elided (n = 26). Durations are shown in figure 10.
The duration of /l/ is significantly longer when the schwa is elided \( t(16) = 3.38, \ p < 0.001 \), collapsing over speech rate. This finding is similar to the results for /s/, and consistent with the possibility that the schwa is in fact being produced but is obscured on the acoustic record.

Finally, in order to determine whether or not the absence of vocalic material corresponding to the /ə/ leads to the formation of a true initial /s/-cluster, all /#səp/- tokens (the only /s/ + voiceless obstruent/ sequences in the experiment) which were coded for elision were examined to see if the /p/ was aspirated. These tokens consisted of the words *superfluous*, *superior*, and *support*. Of the 13 tokens in which schwa was elided (2 in slow speech and 11 in fast speech), all of them exhibited aspiration on the /p/. Furthermore, the range of aspiration, 34–77 ms, is typical of the VOT for voiceless labials in English. Since it is well known that voiceless stops in initial /s/-clusters in English are not aspirated [Browman and Goldstein, 1986; Cooper, 1991; Kahn, 1976], these findings are consistent with the argument that a separate glottal gesture for each consonant is present even though a voiced schwa does not appear on the acoustic record. This would not be expected if the schwa were deleted and a cluster were formed.

Figure 11 combines both elision and aspiration/devoicing on one graph to demonstrate that the predictions in table 2 are upheld. Elision is infrequent for voiceless stop-initial and voiced-initial sequences, and there is frequent aspiration/devoicing for voiceless stop-initial sequences. For both fricative-initial and /l/-second sequences, elision is frequent. These findings are further corroborated by the finding that aspiration is significantly longer when there is no vowel present, and the /s/ and /l/ in tokens with elision are significantly longer than those without elision. In addition, the /p/ in /#səp/- sequences is aspirated, indicating that the absence of a vowel does not lead to the formation of a typical /#sp/ cluster. These results for fricative-initial and /l/-second sequences suggest that there may be an inflated frequency of deletion in these phonotactic categories, since it could simply be the case that there is aspiration/devoicing that is impossible to differentiate from frication, and that there are elements of a schwa that are indistinguishable from the /l/. All of these acoustic results are consistent with an overlap account of coordination, but would be difficult to explain if it were assumed...
that the absence of a voiced vowel were simply due to deletion of the schwa. These results will be discussed in further detail in the general discussion.

### 2.5.4 Individual /#CəC/ Sequences

A look at the individual /#CəC/ environments shows the breakdown of the stimulus categories. Figure 12 shows the proportion of elision for each sequence, and figure 13 contains the proportion of aspiration/devoicing for the individual sequences. A few important points can be made from these graphs. First, the role of phonotactics can be addressed with respect to the individual sequences. For stop-initial stimuli which would give rise to illegal phonotactic sequences, there is little elision overall. However, for the sequences in which elision is relatively high (i.e. fricative-initial and /l/-second), the prospective legality of the resulting /#CC/ does not seem to lead to greater elision. In the fricative-initial category, /#səb/ and /#fət/ sequences show elision rates similar to potentially legal sequences. In the slow condition, /#səv/ does not have any elision, but in the fast condition, there is 11% elision. Likewise, in the /l/-second environment, /#dəl/ shows 11% deletion in the slow condition (and 7% in the fast), and /#məl/ has the second highest deletion rates in the category.

In the devoiced vowel/aspiration measurement in figure 13, the most aspiration occurs for /#kəm/ and /#təm/ regardless of the speech rate. This may be because, like for /l/, it is difficult to distinguish a short, highly overlapped schwa from the following nasal. Alternatively, it could be that greater overlap is allowed in front of a sonorant. That is, if the phonetic environment for the release of the stop is more favorable when the second consonant is a sonorant, then it is possible that the amount of gestural overlap allowed is also sensitive to the nature of the second consonant. The sequence /#gəl/ is also produced with a considerable amount of aspiration at both speech rates (slow: 15%, fast: 31%). A spot check of these tokens reveals that there is a substantial amount of spirantization of the /g/, which leads to both the /g/ and the following schwa appearing as fricative-like on the spectrogram. In this case, then, the spectral portion corresponding to the schwa may be voiced, but higher frequency energy is aperiodic. It is not immediately clear why /g/ is more spirantized than other stops, though one possibility is that speakers can lenite /g/, thereby reducing its articulatory difficulty, since spirantized /g/ cannot be confused with any other sound of English. In the case of /b/ and /d/,
spirantization could lead to confusion with /v/ and /ð/, respectively [Ortega-Llebaria, 2002].

In sum, two points can be addressed regarding the breakdown of performance on individual /#CəC/ sequences. First, the phonotactic environment does not seem to have a large effect on whether or not elision can occur. /#səC/ and /#Cəl/ sequences that would lead to both legal and illegal initial clusters demonstrate similar amounts of deletion. Instead, the articulatory and acoustic properties of /s/ and /l/ seem to have a greater influence on deletion patterns, as explained in section 2.5.3 for /s/-initial sequences.

Second, the information from the amount of aspiration/devoicing in individual sequences suggests that perceptibility of the consonants in the /#C₁əC₂/- sequences may influence the degree of overlap that speakers are able to produce. The voiceless stop-initial sequences have the greatest amount of aspiration before nasals, which is consistent with the notion that the first consonant is most recoverable when it precedes a sonorant. Furthermore, of all /l/-second sequences with an initial voiced stop, only /#gəl/ gives rise to substantial aspiration. Again, it is hypothesized that this aspiration is actually spirantization of the /g/, which occurs more often for that particular voiced stop because it cannot be confused with any other fricative that is in the inventory of

![Fig. 12. Proportion of elision for each individual /#CəC/ environment.](image-url)
English. However, there is also occasional spirantization of a few other voiced-initial sequences, as shown in figure 13.

2.5.5 Individual Speakers

In order to further examine the observation that gestural overlap is implemented to differing degrees by individual speakers, the performance of each individual participant on elision was examined. Elision is used as just one of several possible measures of individual variation that could be examined. A breakdown of each participant indicates that there are two types: those that elide much more in fast speech, and those that elide regardless of speaking rate. This is shown in figure 14.

Participants 1–4 can be classified as rate-dependent eliders (for each, t tests confirm significantly more elision in fast speech, $p < 0.04$, except participant 4, whose difference is marginally significant, $p < 0.07$). Participants 5–9, who show similar amounts of elision in both slow and fast speech (even the patterns of speaker 6 are not significantly different, $p = 0.18$), can be called rate-independent eliders. It should be noted that speech rate increased significantly for both rate-dependent (mean CœC duration: slow = 175 ms, fast = 156 ms, $p < 0.006$) and rate-independent eliders (mean CœC duration: slow = 192 ms, fast = 163 ms, $p < 0.006$). A graph of the individual speakers’ mean CœC durations is shown in figure 15.

Fig. 13. Proportion of devoicing/aspiration for each individual /#CœC/ environment.
Re-examination of the voiceless-initial, voiced-initial, and /s/-initial sequence types by type of elider reveals that interaction between sequence category and elision is respected whether the elision is rate-dependent or not.

An ANOVA for rate-dependent eliders reveals a significant main effect for sequence type \([F(3, 9) = 12.73, p < 0.001]\) and rate \([F(1, 3) = 19.93, p < 0.02]\). The interaction of sequence type and rate was also significant \([F(3, 9) = 19.65, p < 0.001]\). These results indicate that in the slower speech of rate-dependent eliders, there is very little schwa deletion in any /CəC-/ sequences, as shown in figure 16. In the fast speech of these speakers, however, t tests show that there is a significant increase in the amount of elision in fricative-initial sequences \([t(3) = 5.46, p < 0.01]\) and /l/-second sequences \([t(3) = 5.00, p < 0.02]\).

For rate-independent eliders, there is a significant effect of sequence type \([F(3, 12) = 5.49, p < 0.02]\) and no effect of rate \([F(1, 6) = 2.48, p = 0.17]\). The interaction of sequence type and rate was not significant either \([F(3, 12) < 1]\). A Student-Newman-Keuls post hoc test shows that collapsing over rate, the amount of elision in voiced and

Fig. 14. Overall elision patterns by participant.

Fig. 15. Mean /CəC/ duration for individual speakers at slow and fast rates.
voiceless stop-initial sequences is not significantly different from one another, and fricative-initial and /l/-second sequences are both different from one another and from the voiced and voiceless stop-initial sequences (all \( p \) values \(<0.05\). This is illustrated in figure 17.

These results indicate that whether schwa absence is a function of speaking rate or whether it is a general characteristic of a participant’s normal speech patterns, it is limited to fricative-initial and /l/-second sequences. The gestural configurations and amount of overlap posited to account for these findings were discussed in section 2.5.3. The variation due to speaker type and the fact that elision occurs no more than 40% of the time (for the fast speech of rate-dependent eliders) leads to the question of how overlap is implemented in order to give rise to the patterns found in the experiment. Possible analyses are discussed in section 3.1.

3 General Discussion

The results of this study suggest that word-initial pretonic schwa deletion is not consistent with a phonological deletion rule that is applied at fast speech rates.
deletion of the schwa would presumably lead to the formation of an onset cluster that should have no unusual acoustic artifacts, such as consonant durations that are longer than those for singletons, or aspiration of \( C_2 \) in fricative-stop clusters. This is particularly apparent in those sequences that would lead to phonotactically legal onset clusters in English; when speakers apparently elide the schwa in /\( s /l \)-initial or voiced stop + /\( l /\) sequences, they do not produce well-formed consonant clusters. Instead, acoustic residue attributable to schwa remains. These findings are consistent with those reported by Fougeron and Steriade [1997] and Manuel et al. [1992]. By measuring the rates of both elision and aspiration/devoicing, we provide preliminary evidence for an overlap account of pretonic schwa deletion and illustrate how acoustic information might be related to speakers’ articulatory behavior. For example, the following experimental outcomes are better explained in terms of gestural overlap.

(a) The schwa in voiceless stop-initial /\( #C_1 \text{ə}C_2/-\) sequences often surfaces as aspiration of \( C_1 [C_1\text{h}C_2] \) or a devoiced schwa \([C_1\text{d}C_2]\), because the glottal opening gesture of \( C_1 \), which is completed after the \( C_1 \) oral gesture ends, can obscure whatever part of the schwa is not already overlapped by the \( C_1 \) oral closure. Furthermore, when no voiced vowel is present, there is still substantial aspiration after \( C_2 \), which is consistent with the maintenance of a schwa.

(b) There is nearly no elision in voiced-initial /\( #C_1\text{ə}C_2/-\) sequences because overlap of the schwa by the consonant closure only partially obscures the vowel, and there is no glottal opening gesture to conceal the part of the vowel that does remain. Only in very extreme cases does the \( C \) gesture so overlap the \( V \) gesture that it appears absent.

(c) In voiceless fricative-initial /\( #C_1\text{ə}C_2/-\) sequences, the initial fricative is significantly longer in productions which do not also have a voiced vowel than in those which do. This finding is consistent with a situation in which the frication noise of the first consonant and the aspiration from the glottal opening gesture combine to totally obscure the schwa, making it appear as though it has been deleted. Had the schwa actually been deleted, however, it would be expected that the /\( s /l \) would be shorter than singleton /\( s /l \), since it would form a cluster.

(d) The schwa in /\( l /l\)-second /\( #C_1\text{ə}C_2/-\) sequences may appear deleted because overlapping by the initial consonant (either a voiced stop or /\( l /l \) in these data) will cause the schwa to have a very short duration on the acoustic record, which can be difficult to be distinguished from the formant structure present in the following /\( l /l \). Like in /\( s /l\)-initial sequences, schwa absence rates may appear inflated because of this.

The results of section 2.5.4 indicate that when elision rates are high, there does not seem to be an effect of the phonotactic legality of the resulting cluster. Many sequences that would lead to the formation of illegal sequences, such as /\( #d\text{ə}b/-\) or /\( #p\text{ə}t/-\), fail to show significant amounts of elision, but it has been suggested that this is because speakers are manipulating overlap rather than deleting the schwa altogether. Even if elision in categories where rates are high were due to actual deletion, it does not appear that phonotactic restrictions play a significant role in limiting which sequences can undergo deletion. Likewise, the legality of the resulting cluster does not affect the amount of overlap that may be implemented.

A closer look at individual performance indicates that speakers can be divided into rate-dependent eliders and rate-independent eliders. For some speakers, very overlapped
gestures are characteristic of their habitual speech patterns some proportion of the time, whereas other speakers only increase the overlap at faster rates. Regardless of whether speakers’ overlap is rate-dependent or not, however, their productions always follow the predicted pattern for each phonotactic category. This suggests that schwa deletion cannot simply be an optional phonological rule found only in fast speech.

The acoustic data discussed in this study are consistent with a gestural overlap account and are not easily explained by a deletion analysis. However, the acoustic data do not tell the whole story, and in future research, an articulatory analysis of pretonic schwa elision is warranted. For example, the ultrasound methodology used in Davidson [2005] might be one way to investigate elision from an articulatory perspective. Davidson compared English speakers’ production of nonnative sequences like /zg/ (usually produced as [zəg]) to their production of /sk/ and /sək/ to determine whether the inserted schwa could be attributable to inaccurately coordinated consonant gestures. A similar methodology could be applied to a comparison of schwa, elision and the other acoustic manifestations discussed in the present study. An articulatory study would also allow more fine-grained questions about overlap to be asked, such as whether it is more likely when independent articulators are involved than when there are sequential consonants produced with the same articulator (e.g. papaya, tonight). Articulatory data would also help in examining the relationship between overlap and gestural reduction.

3.1 The Relationship between Overlap, Speaking Rate, and Coordination

The elision data raise some important questions about the nature of coordination, how it is implemented, and how speech rate affects the amount of overlap that sequential gestures exhibit. Researchers working in the articulatory phonology framework have proposed that coordination among gestures is defined at a more abstract, phonological level [e.g. Browman and Goldstein, 1986, 1990, 1992; Byrd, 1996; Gafos, 2002; Hall, 2003]. While there is evidence that coordination is phonological and language specific, it is well-known that there is often considerable variability in the amount of overlap as evidenced by both acoustic and articulatory data. Speakers may display some amount of variation in the overlap between gestural constellations [Löfqvist, 1991; Lubker, 1986], less variation among the gestures within a constellation [e.g. Krakow, 1989; Saltzman et al., 2000; Sproat and Fujimura, 1993], and likely even more across word boundaries [e.g. Byrd and Tan, 1996; Zsiga, 2000]. One framework proposed to account for such variability in a principled way is phase windows (Byrd, 1996; Saltzman and Byrd, 2000; see also a similar framework in Keating 1990a, b). According to the phase window proposal, coordination between gestures is not defined between single points on each of the gestures (punctate relative phasing), but rather with respect to constrained ranges in the timelines of sequential gestures or constellation of gestures.

Byrd [1996] argues that if coordination is limited to two single gestural landmarks, then there would be no way to explain the range of variation in the amount of overlap that speakers may produce. Instead, phase windows should be able to account for the variation in overlap due to factors such as stress, focus, speech rate, and phrasal boundaries. On the other hand, it is clear that not all possible overlapping configurations are empirically attested, which is why phase windows are proposed in the first place:
by limiting overlap to a prespecified (language-specific) range, variability can be constrained.

One point to note about phase windows is that every possible coordination within the window is potentially not equally likely, but may rather be probabilistically distributed around some point. This is illustrated in figure 18, adapted from Byrd [1996] with respect to the coordination of adjacent consonants. In this schematic, the coordination relationship between C1 and C2 demonstrates that the target of C2 can vary in its alignment from the center to the offset of C1. However, as demonstrated by the hypothetical probability distribution, the release of C1 will most often be coordinated with the target of C2.

Since the amount of overlap (or regions of coordination within the phase window) is influenced by linguistic and extralinguistic variables in Byrd’s [1996] proposal, the probability distributions depend on the particular conditions under which they are being implemented. In other words, there will be low token-to-token variability when the contextual influences remain similar. Most relevant to this study, Byrd mentions that “for example, a fast speech rate will favor the “more overlapped” end of the window and a slower speech rate the “less overlapped” end” (p. 151).

However, this idea is not as simple as it first seems. Even if Byrd is correct, and a variable like speech rate influences the system by pushing alignment within the phase window toward the more overlapped end (but retaining the same beginning and ending values of the phase), it must be explained how this is implemented. What controls the probability distribution of phase windows so that faster speech rates have the effect of causing greater overlap? Another possibility is that changes caused by speech rate are not reflected in the probability distribution within the window, but rather change the point at which the window is effectively centered. In this scenario, the window and its distribution stay the same, but the starting and ending phase values or landmarks shift so that greater overlap is effected. Building on the example in figure 18, it could be that the range for C1 shifts from its target aligning between the center and offset of C2 to aligning between the target and release of C2. This is illustrated in figure 19.

The possible analyses – whether rate is implemented in the probability distribution or whether the landmark range shifts – likely predict differences in the range of variability that would be observed, and these could be tested empirically with an appropriately designed study. The point that is important for the current discussion is that in general, phase windows are compatible with the proposition that a variable like fast speech can force a categorical change in the amount of overlap, suggesting implementation at a more central or cognitive level.
A number of studies have confirmed that the amount of overlap, or the coordination relationship between two gestures/constellations, varies systematically as a function of rate, though sometimes in a speaker-specific way. For example, Shaiman et al. [1995] and Nittrouer et al. [1988] demonstrated that the phase relation between the jaw closing gesture of a vowel and the upper lip lowering gesture of a following bilabial stop significantly shortened as speakers moved from normal to fast speech. Shaiman et al. [1995] further found that the direction of the effects was not consistent for certain speakers. For 3 of the speakers, as the jaw cycle durations for the vowel decreased, the amount of overlap increased. For 2 other speakers, the onset of upper lip lowering for the bilabial gesture started earlier in the jaw cycle at the slower rates of speech. It is concluded that how speakers implement rate changes may vary, although the notable finding is that these different methods all result in categorical differences between the amount of overlap found for slower versus faster speech. The remaining 3 speakers showed no significant difference in the amount of overlap across speaking rate, which was attributed to the fact that these speakers had particularly large standard deviations for all types of utterances. This is related to the claim of Tjaden and Weismer [1998] that the increase in speaking rate must be sufficiently large in order for overlap to cause a significant, categorical change in the variable of interest.

The issue of individual variability is important in this study of pretonic schwa elision, since 5 of the 9 speakers did not show an increase in elision as speech rate increased. For these speakers, though, the frequency of elision (for fricative-initial and /l/-second sequences) was high at the slow rate. While statistical measures showed that these speakers significantly increased their rate in the fast condition, it is possible that their ‘canonical’ coordination already provides for greater overlap. Following the suggestion of Tjaden and Weismer [1998], the increase in speech rate that they self-selected in this study may not have been great enough to shift them to a different coordination pattern. The remaining 4 speakers, however, increased the rate of elision in the fast condition. For these speakers, it can be hypothesized that the faster speaking rate was sufficient to categorically change the coordination pattern used to determine how much overlap will be tolerated between the gestures in the target words. The distinction between the different kinds of speakers could mean that even at habitual speaking rates, they may have different landmarks or windows designated by their phonological systems. Alternatively, in a phase window account, such a distinction

3 In these studies, phase relation refers to the point in the 360° cycle of a gesture at which the following gesture begins. Shortening the phase angle means that the next gesture begins earlier in the cycle than it did under some other condition(s). This is equivalent to saying that overlap increases.
between speakers could also arise if individual speakers have different probability distributions associated with their windows.

### 3.2 Accounting for Speech Rate at a Central Level of Planning

While there is evidence that increases in speaking rate could lead to a categorical change in the coordination found between gestures, there is still an open question about how that change is specifically executed. It seems that there are two possibilities. One, alluded to above, is that speech rate is governed by a central parameter that directly influences phonological coordination, such as constraints that specify a more overlapping coordination. Presumably, this is done in order to shorten utterance length without having to directly specify what that length should be. Certainly, overlapping is the only factor that speakers manipulate in order to increase rate, but it is the one most directly related to temporal coordination. The other possibility is that greater overlap is a compensatory mechanism that is not planned at a central level, but falls out of the way in which the peripheral motor system organizes itself when faced with the task of producing a greater amount of speech in less time. Though the second possibility is still viable, studies have shown how rate effects can be effectively modeled with a central/cognitive system that controls the temporal organization of the gestures before they reach the motor implementation stage.

Some models of speech production systems have included a central clock which serves as an extrinsic timer. The task dynamics model incorporates two levels that control different kinds of coordination: the intergestural level and the interarticulator level. Saltzman and Munhall [1989] define the roles of these levels as follows:

> The intergestural level accounts for patterns of relative timing and cohesion among the activation intervals of gestural units that participate in a given utterance (e.g. the activation intervals for tongue-dorsum and bilabial gestures in a vowel-bilabial-vowel sequence). The interarticulator level accounts for the coordination among articulators evident at a given point in time due to the currently active set of gestures (e.g. the coordination among lips, jaw, and tongue during periods of vocalic and bilabial gestural coproduction). (p. 336)

From this description, it can be inferred that the intergestural level most closely corresponds to the domain in which phonological coordination specifications would be found. In discussing the types of overlap and 'articulator sliding' presented in Stetson [1951] and Hardcastle [1985] due to increases in speaking rate, Saltzman and Munhall [1989] note that these rate effects could be captured in the dynamical model through 'hypothetically simple changes in the values of a control parameter or parameter set presumably at the model’s intergestural level’ (p. 366).

A later version of the task dynamics model presented in Saltzman et al. [2000] eschews explicit gestural representations with predetermined coordination at the intergestural level, incorporating instead recurrent connectionist networks that determine gestural activation trajectories. The intergestural level in this hybrid model contains a sequential neural network, while the interarticulator level still employs a task dynamics model. Crucially, when the nature of the intergestural level is reconceived, an explicit subnetwork acting as a clock is required to ensure that time flow is accurately represented in the model. Certain parameters in the clocking subnetwork can be fixed to control the speech rate, which ultimately will have an effect on the coordination among gestures. By changing speech rate, gestures can be phased differently with...
respect to one another. In linguistic terminology, it is the job of the phonology – here the network of the intergestural level – to ensure that the gestures are temporally related in a principled, but language-specific way.

The results of the study in section 2 can be tentatively interpreted in terms of the preceding discussion of overlap, temporal coordination and whether speech rate is determined by changes in the phonology. The task dynamics framework assumes that gestures are the underlying units that are given temporal specifications. A phonological framework like articulatory phonology, which is based on the principles of dynamical system models, incorporates the idea that different sequences of gestures have their own specific coordination patterns, and these must be defined and enforced for any given language. In the task dynamics framework, it is proposed that there are functional linkages among gestures, and that these arise as a result of dynamical coupling among gestures in some particular relationship (such as onset cluster, nucleus-coda) [Saltzman and Munhall, 1989]. It has been argued that such linkages are not posited lexically, but are rather assigned by the phonological system of a language [Davidson, 2003; Gafos, 2002; Hall, 2003].

Specifically with respect to the /#CəC-/ sequence examined in this study, the parameters of the phase window framework can be manipulated to give rise to the patterns of elision and aspiration/devoicing found in the results. Using phase windows, it could be hypothesized that the onset of the vowel is aligned from the target to the release of the consonant [as adapted from Browman and Goldstein, 1988, 1995]. At a faster rate, the phonology may specify that the onset of the vowel is aligned from the onset to the center of the consonant. Note that if the onset of V is aligned with the onset of C, the gestures will be fully overlapped and will give the appearance of deletion, or what has been called elision. Alternatively, it could be that the probability distribution shifts so that the onset of the V is more often closer to the target of the C. However, this approach would entail that the vowel gesture must also shrink in size at faster rates, or else there would never be elision. This is because the earliest landmark that the onset of the vowel could be aligned with is the target, and so there would always be some portion of the vowel gesture not overlapped by the consonant. With shifting windows, it is possible to have elision without also shrinking the duration of the vowel gesture, although a decrease in vowel duration is still possible. The two types of coordination constraints possible under a shifting window analysis are illustrated in figure 20.

For the speakers in this study who showed significantly greater elision in the fast condition, their self-selected change in rate was sufficient to induce a change from ‘normal’ coordination to ‘fast’ coordination. For the other speakers, the substantial amount of elision for both rates suggests that they were already speaking at the rate which induces greater overlap even in the slow condition.
4 Conclusions

The main goal of this study was to investigate whether pretonic schwa elision is a phenomenon that is best characterized as a deletion rule or is attributable to increased gestural overlap. Results from the experiment in section 2 suggest that elision is most consistent with overlap. Whereas a true deletion account would have to assume that individual /#CəC-/ sequences selectively allow deletion, overlap can account for all of the acoustic findings by assuming that the same process occurs for all /#CəC-/ sequences.

This analysis is bolstered by the relationship between elision rates and the flanking consonants. Results show that elision rates are dependent on the surrounding consonants in a certain sense, since elision is more frequent in fricative-initial and /l/-second sequences, but this is not related to the phonotactic legality of a resulting cluster. Since greater gestural overlap between the initial consonant and the schwa may occur regardless of the composition of the sequence, the legality of the resulting sequence does not have any effect.

Finally, models of gestural production such as the task dynamics model demonstrate how speech rate may affect coordination patterns at a central level. A change in speech rate is implemented by modifying a coordination parameter, which was hypothesized to be a change in the specifications of the alignment of adjacent gestures. By shifting the coordinating landmarks to be earlier in the time course of the first gestural constellation in a sequence, more overlap is achieved. A phase window approach was also proposed as a way of explaining the amount of variation present in the amount of overlap that is allowed. Whether or not phase windows are ultimately the right approach to capturing the variation requires further research.
Appendix 1: Stimuli and Stories from the Fast Speech Schwa Elision Experiment

<table>
<thead>
<tr>
<th>Voiced-Initial Sequences</th>
<th>db</th>
<th>bg</th>
<th>dp</th>
<th>dk</th>
<th>bf</th>
</tr>
</thead>
<tbody>
<tr>
<td>developers</td>
<td>debatable</td>
<td>begun</td>
<td>depend</td>
<td>Dakota</td>
<td>before</td>
</tr>
<tr>
<td>development</td>
<td>debilitating</td>
<td>beginning</td>
<td>dependence</td>
<td>decanter</td>
<td>buffet</td>
</tr>
<tr>
<td>diversify</td>
<td>debauchery</td>
<td>begonia</td>
<td>departure</td>
<td>decaying</td>
<td>befall</td>
</tr>
<tr>
<td>sn</td>
<td>dm</td>
<td>dj</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Venetian</td>
<td>denounced</td>
<td>democracy</td>
<td>deficient</td>
<td>defeat</td>
<td></td>
</tr>
<tr>
<td>Vanessa</td>
<td>denied</td>
<td>diminish</td>
<td>defense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vanilla</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Voiceless Stop-initial Sequences</th>
<th>kp</th>
<th>pt</th>
<th>pθ</th>
<th>pd</th>
<th>tb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copernicus</td>
<td>potato</td>
<td>pathetic</td>
<td>pedometer</td>
<td>tabasco</td>
<td></td>
</tr>
<tr>
<td>capacity</td>
<td>petunia</td>
<td>pathology</td>
<td>pedestrian</td>
<td>tabouli</td>
<td></td>
</tr>
<tr>
<td>capitulated</td>
<td>petitioned</td>
<td>Pythagorean</td>
<td>pedestrian</td>
<td>Tibetan</td>
<td></td>
</tr>
<tr>
<td>tm</td>
<td>km</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tamale</td>
<td>commercial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tomato</td>
<td>commendable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>timidity</td>
<td>community</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fricative-Initial Sequences</th>
<th>sf</th>
<th>sv</th>
<th>sm</th>
<th>sp</th>
<th>sb</th>
<th>ft</th>
</tr>
</thead>
<tbody>
<tr>
<td>sophisticated</td>
<td>civilian</td>
<td>Somali</td>
<td>superior</td>
<td>suburban</td>
<td>fatigue</td>
<td></td>
</tr>
<tr>
<td>sufficient</td>
<td>Savannah</td>
<td>semester</td>
<td>support</td>
<td>Sebastian</td>
<td>photographer</td>
<td></td>
</tr>
<tr>
<td>suffice</td>
<td>Seville</td>
<td>cement</td>
<td>superfluous</td>
<td>sabbatical</td>
<td>photography</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/l/-Second Sequences</th>
<th>bl</th>
<th>ml</th>
<th>dl</th>
<th>gl</th>
<th>sl</th>
</tr>
</thead>
<tbody>
<tr>
<td>balloon</td>
<td>Malaysia</td>
<td>dilapidated</td>
<td>Galapagos</td>
<td>cylindrical</td>
<td></td>
</tr>
<tr>
<td>believe</td>
<td>malaria</td>
<td>delightful</td>
<td>Galicia</td>
<td>selection</td>
<td></td>
</tr>
<tr>
<td>baloney</td>
<td>Melissa</td>
<td>delectable</td>
<td>galactic</td>
<td>celebrity</td>
<td></td>
</tr>
</tbody>
</table>

Note: Stimulus words are bolded.

Story 1

The food that I ate as a child growing up in South Dakota was the same standard, boring fare that kids typically eat. My lunches often consisted of baloney and tomato sandwiches and a little plastic cup of vanilla pudding for dessert. As I’ve gotten older, I’ve tried to diversify my eating habits. I learned that I like spicy foods, and tabasco sauce is often found on my shopping list. Mexican food is my preferred cuisine, and my favorite dish at the local restaurant is a delectable meal of tamales stuffed with potato and covered in salsa verde and cheese. My next goal is a nearby Somali restaurant that my friend Melissa recommended. She and her sister Vanessa are the best food critics I know. Although take-out from the Chinese buffet is a common alternative to going out, there are even some nights that I cook. I’ve found that good food can be easy to make. One quick but delightful meal is polenta, and even exotic foods like tabouli or falafel can be bought in a box.

Of course, not all of my culinary explorations have been so successful. While it’s easy to be adventurous at home, foreign food can be a little more risky. As a photographer for...
a magazine, I have to travel a lot, which means trying many new types of food. One time on a trip to Malaysia, I tried a pudding made of cassava root, which gave me a debilitating headache for a week. After that, I toyed with depriving myself of local foods, but I quickly learned that refusing the offers of your hosts denotes great ignorance. One time upon my departure from the home of a Tibetan monk, I was given a decanter of a local tea. I was about to politely decline it, when my interpreter quickly informed me that if I did so, a great tragedy could befall me. Not wanting to offend my host, I capitulated and accepted the tea. It turned out to be very tasty, and I think it even sedated me a little. I later used it when I was stressed and fatigued from my job. Then there was the time that I was on assignment in Spain. After enjoying myself immensely at the restaurants in Seville, I was sadly forced to forgo the local delicacy in Galicia because I had run out of pesetas. Maybe I’ll be able to go back and try it someday.

Story 2
Recently, crime in our peaceful collegiate town has been on the rise. Ten years ago, the biggest concern was the defacing of college property or the occasional collision between a car and a pedestrian. Although problems like drunken debauchery by students will always be pervasive, today we’re also faced with the threat of personal injury. Prevention of campus crime requires a collaboration between police and civilians. Furthermore, funding of campus security is often deficient, and officers require more resources if they are to defend us against crime.

However, increased security measures can only go so far. It is imperative that members of the university campus also be trained in self-defense tactics. One example of the success of such measures is the case of Dr. Sebastian Gray. Dr. Gray, a professor of pathology who researches the effects of malaria, was here on sabbatical from Savannah. One day while walking from his office to the parking lot, he was mugged. Although he was forced against a cement wall, Dr. Gray was able to defeat his sadistic attacker as a result of his self-defense training. Dr. Gray, who was known for his timidity and pedantic lecturing style before the attack last semester, has now achieved celebrity status with his students.

Story 3
The problem of urban sprawl has already begun to bedevil American society. Many Americans are seduced by the convenience of the suburban lifestyle. Suburbs first emerged as a means for Americans to escape decaying cities and dilapidated neighborhoods. Rather than demolish and reconstruct devalued property, planners wanted to start anew in the countryside. Both the Atlantic and Pacific coasts, where urban centers abound, were targeted by zealous developers.

Some see ballooning suburbs as a realization of the ‘American dream’ afforded by democracy. Others see the advent of the suburbs as divisive, exemplifying the widening gap between rich and poor. Recently, there has been a backlash against urban sprawl. Many have denounced sprawling development because its sterile feel deprives people of a sense of community. The segregation of residential and commercial zones forces people to depend on their cars. This leads to commuters clogging the highways and causing superfluous traffic jams. Many highways are now burdened with more traffic than their maximum capacity allows. Mass transit in the form of buses and trains is beginning to be implemented, but its ability to diminish car traffic is debatable. Many Americans, who believe the car is a superior mode of transportation, do
not want to be denied the independence the car affords them. Some groups have petitioned for improved mass transit, but because of political struggles, there is often a lack of sufficient funds or support for public transportation. The efforts of the commissions in charge of public transit are far from commendable and sometimes downright pathetic. Until we figure out a way to reduce the dependence on cars in our towns, we may be heading toward a collapse of our current transit system.

Story 4

As I was sitting in my solarium, I was bedazzled by the beauty of the night sky, which seemed to be bedecked with millions of gleaming stars. I wondered if this is how the astronomer Copernicus felt as he gazed at the sky 500 years ago. It’s amazing how much astronomy has advanced since the days of Copernicus, starting as early as his successor, Galileo. It is said that Galileo, a poor professor in the Venetian Republic in 1609, discovered that the rudimentary cylindrical telescope being developed by his contemporaries would not suffice, and set about creating one that would. Since then, we’ve developed space technology capable of galactic travel like the Poseidon satellite which can map the ocean surface and the Cassini spacecraft which will soon make a mission to Saturn.

In fact, all fields of science have advanced tremendously since the days of thinkers like Pythagoras, Galileo, or Newton. One of the most pervasive ideas in modern science is evolution, first described by Charles Darwin in the early 1800s. Darwin traveled to the Galapagos Islands, where he tackled questions such as how the flora and fauna got on the islands to begin with. The variety of plants and animals he found, like petunias and begonias as well as finches, tortoises and sea lions, fascinated him and led him to formulate the theory of natural selection.

The great scientific advances of the likes of Galileo and Darwin have led to technology that has transformed every aspect of our lives. Despite the proliferation of technology in the modern world, few lay people have an understanding about how the technology they use every day works. Just ask somebody if they understand the science behind photography or the video cassette recorder. Even a contraption like a pedometer, which counts the number of steps a person makes, requires sophisticated technology in order to give accurate readings. Fortunately, there is no shortage of people interested in inventing new and better products and we can expect that scientific discoveries will keep improving forever.
Appendix 2: Spectrograms Illustrating Measurement Criteria, Schwas, Aspiration, and Elision

(Note: the low frequency noise present during voiceless consonants in these spectrograms is due to a hum in the room that the recordings took place in.)

a Voiceless consonant-schwa coordination with voiced schwa present (participant 3,wee potato, slow condition)

b Voiceless consonant-schwa coordination with no voiced schwa present (participant 8, wee Pythagoras, fast condition)

c Voiced consonant-schwa coordination with voiced schwa present (participant 6, wee dependence, fast condition)
Schwa Elision in Fast Speech

1. Voiceless consonant-schwa coordination with voiced schwa present (participant 7, *suffice*, slow condition)

2. Voiceless fricative-schwa coordination with no voiced schwa present (participant 7, *support*, fast condition)

3. Voiced consonant-schwa-/l/ coordination with schwa present (participant 2, *delectable*, fast condition)
Acknowledgements

Thanks to Paul Smolensky, Matt Goldrick, Greg Guy, Jason Shaw, Paul de Decker, Mark Tiede, and an anonymous reviewer for helpful comments on this work. This paper is adapted from a chapter in the author’s 2003 Johns Hopkins University dissertation. A much earlier version of this work was presented at LabPhon 8 at Yale University. This research was supported by the IGERT program in the Cognitive Science of Language at Johns Hopkins University, NSF grant 997280.

References


Ladefoed, P.; Maddieson, I.: The sounds of the world’s languages (Blackwell, Oxford 1996).


