



Saying consonant clusters quickly

Dani Byrd* and Cheng Cheng Tan

UCLA Department of Linguistics

Received 27th July 1994, and in revised form 11th November 1995

Articulatory variation as a function of speech rate is investigated experimentally. Specifically, two strategies for increasing rate are considered: shortening the duration of each component of a sequence and increasing the relative overlap of these components. Reduction in the magnitude of the articulations is also reported. Four consonant sequences spanning word boundaries were produced by five talkers at a variety of rates, and electropalatographic data were collected. Consonant duration, displacement, and temporal overlap and latency between the consonants are evaluated with respect to speaking rate. The results evidence both mechanisms of faster speech—individual consonants shorten in duration *and* a relatively linear increase occurs in the overlap of the articulations. The consonant sequences, however, do not behave identically. A consonant's place, manner, and syllabic position are also found to affect the way in which speech rate is increased.

©1996 Academic Press Limited

1. Introduction

What do we do when we talk faster? That is, which aspects of articulation vary as speech rate increases? We know that faster speech rates cause a succession of phonological units to occupy less total time (see for example, Adams, Weismer & Kent, 1993)—presumably, that is what we mean by “faster.” Individuals could, of course, accomplish this in a variety of ways. Some researchers (Ostry & Munhall, 1985; Adams, Weismer & Kent, 1993), for example, have found a consistent relation of an articulatory unit's duration to its ratio of peak velocity to displacement. Here, we consider two conceivable ways in which talkers might increase their speaking rate. First, the *duration* of each component articulatory unit might shorten (with some units being shortened more than others). As a consequence of decreasing duration, the spatial magnitudes of these articulations might also reduce (Lindblom, 1963, 1964). However, researchers have also suggested another view of fast speech in which an increase in the *relative overlap* of units might yield the overall shorter duration of a sequence (Fowler, 1977; Gay 1981; Browman & Goldstein, 1990). Such changes have been shown for sequences of laryngeal gestures by Löfqvist &

* Now at Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, U.S.A., where correspondence should be addressed.

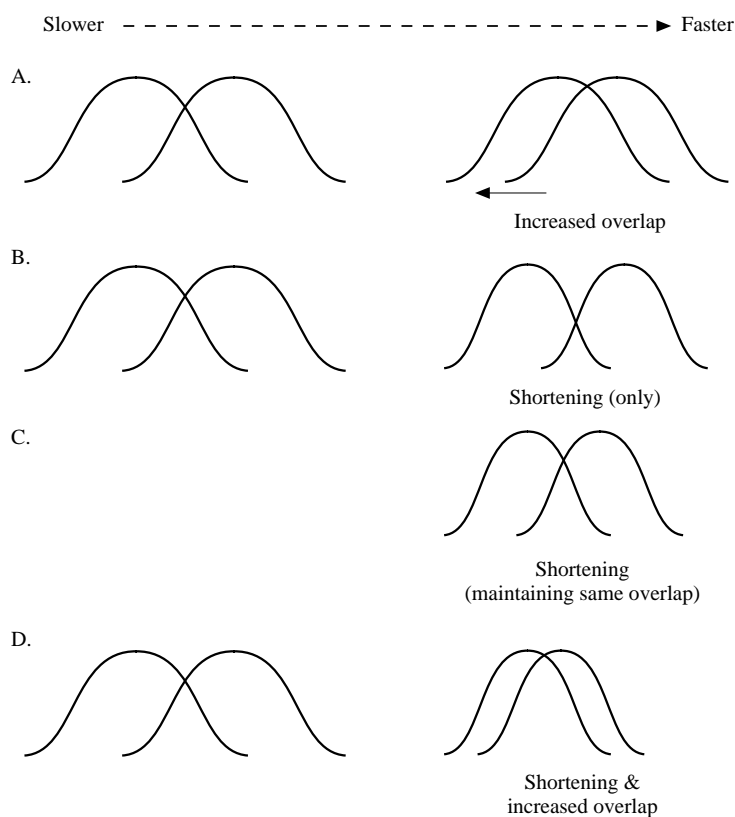


Figure 1. Schema of possible ways in which speech rate may be increased.

Yoshioka (1981) and Munhall & Löfqvist (1992). This experiment compares these two possible mechanisms, shown schematically in Fig. 1, for talking faster. Notice that shortening could occur with no change in the *absolute* temporal latency of the consonants (Fig. 1B); this would result in less overlap with increased rate. However, shortening could also occur with no change in the *relative* timing, which would yield no difference in overlap but shorter temporal latencies (Fig. 1C). Of course, increased overlap and decreased component duration are not mutually exclusive (Fig. 1D), and individuals may use each strategy to a different degree.

Three issues are addressed experimentally. First, we determine whether the relative timing, i.e., overlap, of a C#C sequence (where # indicates a word boundary) changes as a function of rate. Increased temporal overlap in fast speech has been demonstrated between vowels and consonants (Gay, 1981) and has recently been hypothesized more generally to account for certain casual speech processes (Browman & Goldstein, 1990). Gay found that “the duration of segmental units, the displacement and velocity of articulatory movements, and the temporal overlap between individual segments undergo nonlinear transformations during changes in speaking rate” (1981: 158). Gay interpreted this as reflecting a restructuring of the temporal pattern of an utterance rather than a simple change in

the spacing of motor commands. Additionally, we want to know whether rate affects articulatory timing in the same way for *different* consonant sequences.

Secondly, we also consider whether individual consonants in a C#C sequence *shorten* as rate increases. Recall that Gay (1981), like Kozhevnikov & Chistovich (1965), determined that duration changes were not distributed proportionally across consonant and vowel segments. We ask the parallel question of whether timing changes in a consonant sequence are instantiated in the same way across the consonants regardless of their place, manner, and syllabic position. That is, do speakers control rate by decreasing the duration of both consonants in a C#C sequence in the same way?

Thirdly, speaking rate may affect spatial reduction. (By spatial reduction, we mean simply a decrease in linguapalatal contact.) Adams, Weismer & Kent (1993) noted that some researchers have observed decreases in movement amplitude in faster speech (citing Lindblom, 1964; Kent & Moll, 1972; and Flege, 1988) while others have not seen such amplitude changes (citing Abbs, 1973; Gay & Hirose, 1973; Benguerel & Cowan, 1974; Hughes & Abbs, 1976; Engstrand, 1988). Dixit & Flege (1991) looked at one Hindi speaker's [t] productions using EPG and found that rate did not substantially affect amount of linguapalatal contact. This could result from many factors—including changes in effort, coarticulatory demands, or velocity—which we cannot measure independently. We test whether the consonants in a sequence reduce in linguapalatal contact as speech rate increases. Specifically, we consider whether an increase in speech rate causes all the consonants in a sequence to reduce or whether this reduction is limited to certain consonants or certain syllabic positions.

Having measured both duration and displacement, we can address the relationship between these variables. Gay (1981) compared Lindblom's (1963, 1964) reasoning that the degree of undershoot is directly proportional to duration with Gay's 1968 and 1974 work and Kent's (1970) work suggesting that rate may cause changes in articulatory effort or velocity, signifying a reorganization in muscle forcing function along with a temporal reorganization. Gay says, "while target undershoot commonly accompanies an increase in speaking rate, it is by no means ubiquitous" (1981, p. 152–3). Based on the findings of much previous research, it seems clear that a number of variables might be used by a speaker to implement rate change and that they may interact in a complex fashion. We discuss the relations between the duration and displacement of individual consonants and the degree to which prosodic and articulatory variables affect the changes that occur with increased rate.

This examination of these potential articulatory timing changes as a function of rate makes no claim, implicit or otherwise, as to whether the motivation for these changes is limited to increasing articulatory efficiency or whether such adjustments serve the purpose of minimizing changes to the acoustic/auditory output patterns. Asking people to speak faster may yield a variety of articulatory strategies depending upon each speaker's interpretation of instructions. (In fact, we observe some individual differences below.) Flege (1988) believes that a model of rate-induced articulatory variation must take into account "speaker-dependent variables such as the desire to speak clearly and the extent to which speaking rate is increased" (p. 914). It is quite possible that the acoustic/auditory consequences of increasing overlap, shortening durations, and reducing spatial displacements constrain the degree to which each strategy is employed by a particular speaker. (See

Byrd (in press) for a discussion of a model of interarticulator timing that explicitly allows such influences on phasing relationships.) Our purpose here is to describe the changes in consonant cluster articulation produced by five individuals as a function of their speaking rates without prejudice to the motivations that underlie these changes. Clearly, before we can address the goals of and influences on the motor control of speech rate, we must understand the degree to which various phonetic forms are affected by rate and the nature of those effects.

2. Method

2.1. Data collection

2.1.1. Electropalatography

We use the Kay Elemetrics Palatometer system for electropalatography (EPG) to record information about speakers' tongue-palate contact over time. This system uses an artificial palate of thin acrylic that extends around the teeth and is embedded with 96 electrodes. The centers of the electrode contacts range from 2 mm apart in front to 10 mm apart in back but of course differ depending on palate size. The Kay Palatometer scans the palate at a 100 Hz sampling rate with a scan time of 1.7 ms to acquire all 96 values in a sample. The EPG data is recorded by computer and indicates the presence or absence of linguapalatal contact at each electrode.¹

EPG data are generally quantified using specifically designed software for data reduction. Most temporal displays show the total number or percent of electrodes contacted at each frame over time. These displays are called "contact profiles" (Hardcastle, Gibbons & Nicolaidis, 1991). In the present work, we use measurements made from such displays. (A similar method was used by Barry, 1991).

2.1.2. Subjects

Five speakers were recorded and paid at a standard compensation rate. These include two men and three women who have grown up and been educated in Southern/Central California. All were monolingual English speakers with a dialect typical of this area. All speakers reported no speech or hearing pathology. Speakers will be referred to as Speakers A, B, K, M, and S.

¹ There are concerns regarding the use of electropalatography. First, the presence of the artificial palate might interfere with normal articulation. Research has shown no significant difference in patterns of linguapalatal contact between direct palatography and electropalatography (Hardcastle, 1972, see also Fletcher, McCutcheon, Wolf, Sooudi & Smith, 1975 and Flege 1976; but see Hamlet & Stone, 1978). Kozhevnikov & Chistovich (1965) found no difference in intelligibility with versus without the palate. Second, the electrode coverage is limited mostly to the hard palate area making it possible that velar closure contact is under-represented. This problem is not serious for front velars. Hardcastle & Roach (1979), using a smaller pseudopalate, observed complete velar closures in the phrase "catkin." In the present study using the Kay Palatometer with 96 electrodes, we examined ten repetitions of the phrase "Type bag gab again" for every subject and determined that each token had a complete seal across the back of the palate for the velar closure in [g#g]. Some tokens showed up to five electrodes contacted along the mid-sagittal plane. In general, one should note that, depending on the method of analysis, EPG data may not reflect sliding of tongue against the palate. Lastly, electropalatography, like movement tracking of articulators, of course offers no means of identifying the onset/offset of neural activation.

2.1.3. Recording set-up

Before each experimental recording, the speakers wore their artificial palates for an hour or normal activity to accommodate.² Before recording, the palate electrodes were calibrated using the Palatometer software. For each recording, one practice page (i.e., one block) of material was read by the speaker before the experiment began. The nonsense word *sab* was pointed out to the speaker at this time. A simultaneous voice recording was made via a head-mounted, directional microphone connected to the external Computer Speech Lab hardware that interfaces with the computer. Both the EPG and audio signals were recorded directly into a single computer file. Speakers were cued for each sentence by the experimenter, and there was a pause after each one. The recording session lasted for an hour and a half. If a speaker paused, hesitated, or otherwise had a false start, he or she was prompted for a repetition of that sentence. Such repetitions were infrequent.

2.2. Data analysis

2.2.1. Region definition

Articulatory regions on the pseudopalate are often used in EPG data analysis. These regions are groupings of electrodes in a subsection of the palate that generally corresponds to a “place of articulation.” Previous EPG work has employed *predetermined* regions on the pseudopalate which are identical for all experimental subjects and may even be hardwired into the pseudopalate itself. In contrast, we determined articulatory regions empirically for each speaker based on control utterances having no lingual coarticulation with another consonant. This approach allows for differences across speakers in the electrode placement with respect to anatomical configurations or articulatory patterns. (See Byrd, Flemming, Mueller & Tan, 1995 for a discussion of this approach to EPG data analysis.)

Ten repetitions of [d#d], [s#s], and [g#g] sequences in the phrase *Type “baC Cab” again.* were used to establish front and back regions on the palate independently for each speaker. Crucially, no electrodes that were contacted at the minimum for the [æ] vowel contact were included in the consonantal region. This ensures that the moment of initial contact measured in the sequence will in fact be the concomitant of the formation of a consonant constriction rather than normal vocalic contact. All electrodes contacted after the vowel minimum, until and including the frame of maximum contact during the consonant, were designated as belonging in the relevant region: front for [s] and [d], back for [g]. Any electrodes that were marked in this way as members of both regions were also excluded. These cases were generally few and always adjacent to the excluded vocalic region. Thus the measurements made are conservative in identifying the frame of initial contact for a consonant but there is a high degree of confidence that the contact measured is actually attributable to the upcoming consonant in that region. All other (i.e., uncontacted) electrodes were also included in one of the two regions. If an electrode was never contacted during the control sequences, it was included in the region to which it was physically closest. This was determined by measurements made with a flexible ruler on the acrylic palates themselves. The resulting regions for each

² Flege (1986 cited in Flege, 1988) found that a snugly fitting pseudopalate produced no perceptible interference with speech after five minutes of adaptation on the part of the speaker.

subject are shown in the appendix where an effort has been made to represent the dimensions of the pseudopalates, as palate size and depth differ substantially from person to person.

2.2.2. *Experimental stimuli*

This experiment considers four heterosyllabic sequences: [d#g], [g#d], [s#g], and [g#s]. They were recorded in the frame sentence “Say *baC Cab* again.” The goal was to create as much variation in speaking rate as possible without allowing particularly slow speech. To accomplish this, blocks of the four sequences were created with a rotating order (with the constraint that no two instances of the same C1 abutted). Thirty-two such blocks were recorded with the speakers instructed to increase their speech rate through each sentence in a block of four. The speakers were instructed specifically to read the first sentence in each block at a normal speaking rate. The experimenter cued the speakers for each sentence with the words “Normal,” “Medium,” “Faster,” and “Fastest” successively. It should be emphasized that these rate instructions do not serve as variables in a categorical analysis, but rather, served only to engender a wide range of rate variation. Rate is considered a continuous variable in our analyses. This approach was adopted because we do not feel that speakers make a unique transform of their articulation in response to a specific categorical instruction as to speech rate. We hoped rather to collect data across a wide range of each subject’s speaking rates. A total of 640 utterances was recorded for the study. (Of these, 11 are not included in the analysis, eight due to complete lack of contact for a consonant and three due to data processing error.)

2.2.3. *Measurements*

Our measurements were made from contact profiles plotting time in frames of 0.01 s against the percentage of lingual contact (rounded to the nearest integer) in each pseudopalate region. For example, a measure of 50% in the front region at a particular frame means that in that speaker’s designated front region on their palate, 50% of the electrodes registered contact at that time. For the measure of speech rate, it is important to use an interval that does not include the consonant sequence itself (to avoid artificially high, part-whole correlations, see Barry, 1983) but is close to the consonant sequence so as to adequately reflect its rate of articulation. As the measurement of speaking rate, we evaluate the time interval in the carrier sentence from the frame after the end of any contact for the CC sequence through the peak contact for the [g] in “again,” the last word of the carrier. This interval immediately following the cluster is assumed to accurately reflect the cluster’s rate. Additionally, its endpoints are clearly defined by the EPG data. This rate measure is regressed against the following measures of the consonant sequences:

- the temporal interval between C1 and C2 onsets (Δ_{ONSETS});
- the temporal interval between C1 and C2 peaks (Δ_{PEAKS});
- relative overlap as indicated by the percent of C1 contact during which C2 contact also occurred (C1 OVERLAP (%));
- individual consonant duration as indicated by the duration of contact in a region; and
- individual consonant displacement as indicated by peak contact amount (maximum percentage of a region contacted).

TABLE I. Mean of rate measurements and ranges (in seconds) split by speaker and sequence; the lower the rate measure, the faster the speech

| Speaker | A | B | K | M | S | Sequence | [s#g] | [g#s] | [g#d] | [d#g] |
|---------------|------|------|------|------|------|----------|-------|-------|-------|-------|
| Mean Rate (s) | 0.23 | 0.25 | 0.25 | 0.17 | 0.19 | | 0.21 | 0.21 | 0.23 | 0.23 |
| Range (s) | 0.16 | 0.25 | 0.17 | 0.16 | 0.19 | | 0.23 | 0.26 | 0.27 | 0.23 |

The range of rate variation observed for each sequence and each speaker is shown in Table I. Note here and throughout that a higher value for the rate measure indicates a longer interval in the carrier sentence, hence, a *slower* speaking rate.

3. Results

Decreases in absolute latency (i.e., Δ_{ONSETS} and Δ_{PEAKS}) and increases in overlap (C1 OVERLAP) are generally observed as rate increases, although not every combination of speaker and sequence shows this pattern.

3.1. Effects of speaking rate on timing

Table II details the goodness of a linear fit in the regression analyses (r^2) and the slope of the fitted line for each pair of speaker and sequence, as well as for each sequence with the speakers pooled. The r^2 is the proportion of the dependent variable's variability that is explained by the independent variable, here speech rate.

TABLE II. All significant r^2 values ($p < 0.05$)— $F(1, 31)$ for individual cells, $F(1, 157)$ for pooled cells; * indicates $p < 0.01$; and slope values (m) for a linear fit for Δ_{ONSETS} , Δ_{PEAKS} , and the percentage of C1 overlapped by C2 as a function of measured speech rate

| Seq. Speaker | Measure | [s#g] $r^2(m)$ | [g#s] $r^2(m)$ | [g#d] $r^2(m)$ | [d#g] $r^2(m)$ |
|-----------------|--------------------------|-------------------|-------------------|-------------------|-------------------|
| A | Δ_{ONSETS} | *0.603 (0.357) | *0.0354 (0.19) | | |
| | Δ_{PEAKS} | *0.0519 (0.299) | *0.213 (0.24) | | |
| | C1 OVERLAP | | | *0.258 (1.355) | |
| B | Δ_{ONSETS} | | | | 0.18 (0.129) |
| | Δ_{PEAKS} | 0.117 (0.112) | 0.182 (0.236) | | |
| | C1 OVERLAP | | | | |
| K | Δ_{ONSETS} | *0.429 (0.253) | | | |
| | Δ_{PEAKS} | *0.335 (0.207) | *0.507 (0.267) | *0.233 (0.166) | |
| | C1 OVERLAP | | | | |
| M | Δ_{ONSETS} | *0.258 (0.23) | 0.124 (0.102) | 0.166 (0.15) | |
| | Δ_{PEAKS} | 0.179 (0.165) | | | |
| | C1 OVERLAP | *0.247 (−1.212) | | | |
| S | Δ_{ONSETS} | *0.284 (0.111) | *0.29 (0.095) | *0.443 (0.122) | |
| | Δ_{PEAKS} | | | | 0.138 (0.182) |
| | C1 OVERLAP | | | | |
| pooled | Δ_{ONSETS} | *0.265 (0.241) | *0.255 (0.174) | *0.32 (0.306) | *0.152 (0.109) |
| | Δ_{PEAKS} | *0.285 (0.167) | *0.3 (0.189) | *0.077 (0.073) | *0.203 (0.175) |
| | C1 OVERLAP | *0.101 (−1.096) | *0.07 (−0.826) | *0.134 (−1.116) | *0.061 (−0.904) |

In Table II only the significant ($p < 0.05$) fits and slopes are shown, and an additional * indicates a $p < 0.01$. A significant fit means that as rate changed, the timing measure also changed in a relatively linear fashion. The sign of the slope (m) indicates whether the two measures are positively or negatively related. The higher the r^2 , the more linear the relationship.

Perhaps the first thing one notes in Table II is that the correlation of these measures with rate is not very high. In part, this is due to the limited resolution of the data (100 Hz sampling rate), but it is also the case that rate is only one of many factors influencing articulatory timing in these sequences.

Rate has a significant effect on the articulation of each consonant sequence such that absolute latency between consonants decreases *and* overlap increases with faster speaking rates.³ However, not all speakers and sequences show equivalent effects. By comparing the columns in Table II, we see that rate has the least influence on the [d#g] sequences. This presumably is due to a ceiling effect, whereby [d#g] is so overlapped even at slower rates (see also Byrd, this issue) that only a minimal additional increase due to faster rate is evidenced.

Table II does not detail the predicted effect of rate on overlap for each individual. Speaker M shows a strong increase in overlap with rate for the [s#g] sequence. One speaker (A) for one sequence ([g#d]) shows a significant effect in the opposite of the direction, and Speaker S tends ($p = 0.08$) to decrease overlap for the [g#s] sequence as rate increases. In order to determine whether the effect of rate on overlap indicated in the bottom row of Table II for speakers pooled is in fact robust *across* speakers and sequences, we removed the exceptional cells from the analysis. When the sequences, *excluding* the [d#g] sequence and these three cells (M's [s#g], A's [g#d], and S's [g#s]), are pooled, Speakers B and M show significant effects ($p < 0.01$) of rate on overlap such that overlap increases with rate, Speaker A shows a trend ($p = 0.08$) in this direction, and Speaker S has an effect ($p = 0.04$) in the opposite direction. Rate has no effect on overlap for Speaker K. Speaker S is discussed further below.

We conclude that the results with speakers pooled do indicate a relatively linear increase in temporal coarticulation for each sequence as speech rate increases, particularly with respect to absolute latency. This can be seen in the consistent significant influence of rate on timing for every sequence in the pooled data, which considers many more tokens than each speaker-by-sequence cell considered separately. Given a linear relationship, speaking rate accounts for up to 32% of the variation in the timing measures with the speakers pooled, as demonstrated by the r^2 values shown in Table II. Additionally, although not surprisingly, speakers with an overall faster speaking rate have a higher mean degree of overlap. (See Table III). This indicates that not only did speakers increase temporal coarticulation as they talked faster but that individuals with an overall faster speech rate also have greater overlap. That is, the relation between overlap and rate holds both across speakers and, although less robustly, within speakers.

Having considered each individual's behavior, the regression plots for the absolute latency (Δ_{ONSETS} and Δ_{PEAKS}) and for overlap (C1 OVERLAP) between the consonants are shown in Figs. 2–5 with speakers pooled. These plots serve to

³ An examination of the 95% confidence bands for slope in the regression plots for each sequence (speakers pooled) exhibits no change in the sign of the slopes.

TABLE III. Median rate (s) for each speaker—the lower the value, the faster the speech—compared to mean C1 Overlap (%)

| Speaker | Median rate (s) | Mean C1 overlap (%) |
|---------|-----------------|---------------------|
| M | 0.17 | 78 |
| S | 0.19 | 59 |
| A | 0.22 | 55 |
| K | 0.25 | 56 |
| B | 0.26 | 54 |

illustrate differences among the four sequences. Contributions of individual speakers to these plots can be assessed by considering Table II and the discussion above.

Several aspects of these plots are noteworthy. First, as rate decreases, overlap generally increases and absolute latency decreases. These changes are relatively linear. However, some differences among the sequences are apparent. One can see for [d#g], particularly in the case of C1 OVERLAP (Fig. 5c), the ceiling effect described above. Regardless of speaking rate, there is a strong tendency for this sequence to be produced in a completely overlapped fashion. The data points tend to cluster around 100% overlap, and all speakers produced tokens with this degree

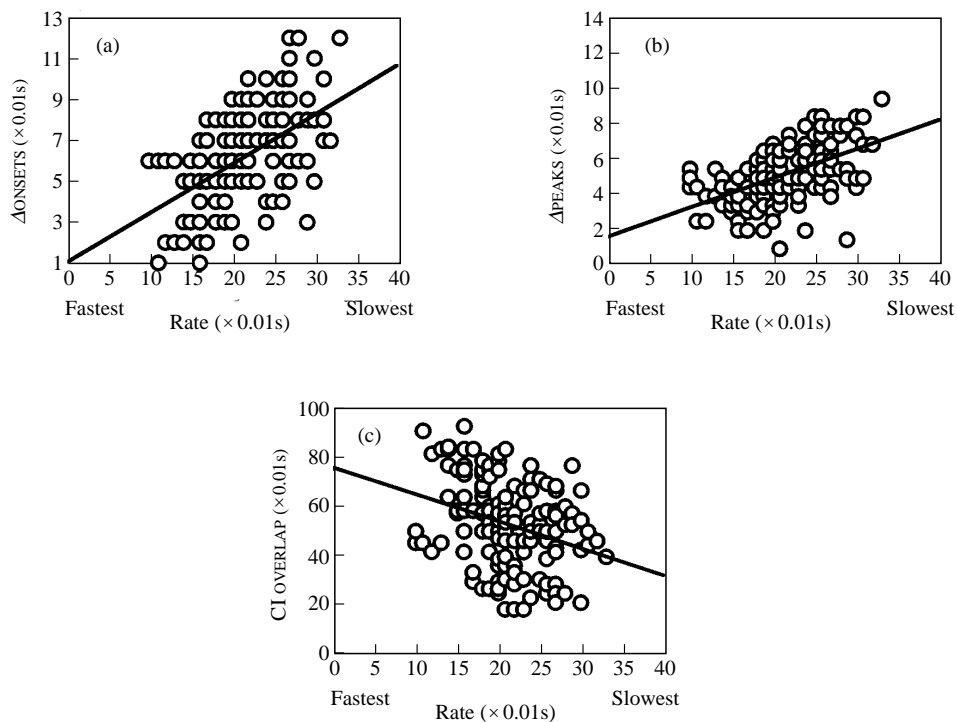


Figure 2. Plots for [s#g] for Δ ONSETS (a), Δ PEAKS (b), and C1 OVERLAP (%) (c) regressed against rate (x-axis; fastest closest to the origin).

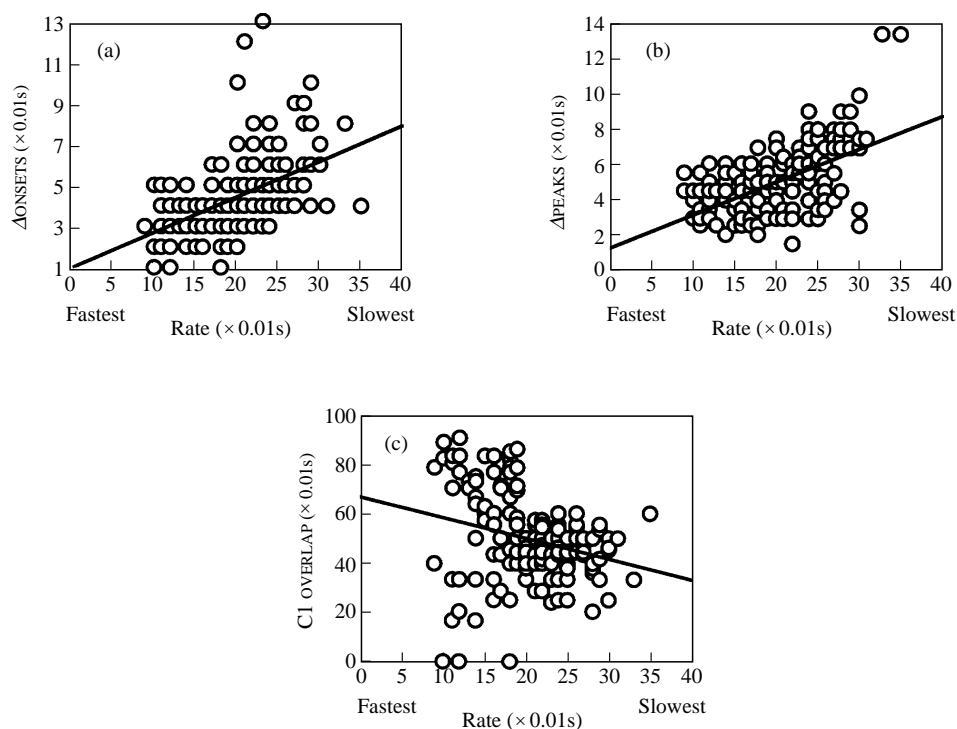


Figure 3. Plots for [g#s] for ΔONSETS (a), ΔPEAKS (b), and C1 OVERLAP (%) (c) regressed against rate (x-axis; fastest closest to the origin).

of overlap. In this case, the particular sequence appears to have a greater influence on timing than the rate at which it is spoken.

We noted above that Speaker S differed from the other speakers in terms of overlap patterns. In a closer examination of Speaker S, it seems that his front-back sequences show little effect of rate on overlap, although absolute latency always shortens with increased rate. [s#g] also shows relatively little rate effect (but does have a linear fit with a positive slope for C1 OVERLAP), and, as for the other speakers, [d#g] has almost complete overlap in most instances. Speaker S's, back-front sequences, however, tend to show *less* overlap at faster rates. Thus, it appears for these sequences that something else about the articulation of the back-front sequences was more influential in their timing than any advantage gained by increasing overlap at faster speech rates. Recall that Speaker A also displayed this pattern for the [g#d] sequence, although not for the [g#s] sequence.

In summary, this experiment provides articulatory evidence that the coproduction of two consonants across a word boundary increases as rate increases. Figure 6 plots the percentage of the way through C1 contact that C2 contact was initiated, against speaking rate. The exceptional behavior of [d#g] and Speaker S outlined above have been omitted from Fig. 6 to present a picture of the overall change in relative timing with respect to rate. We see that as speaking rate increases, C2 starts earlier in C1, the r^2 for the linear fit being 0.24.

These results indicate that, for these sequences, overlap does increase with rate

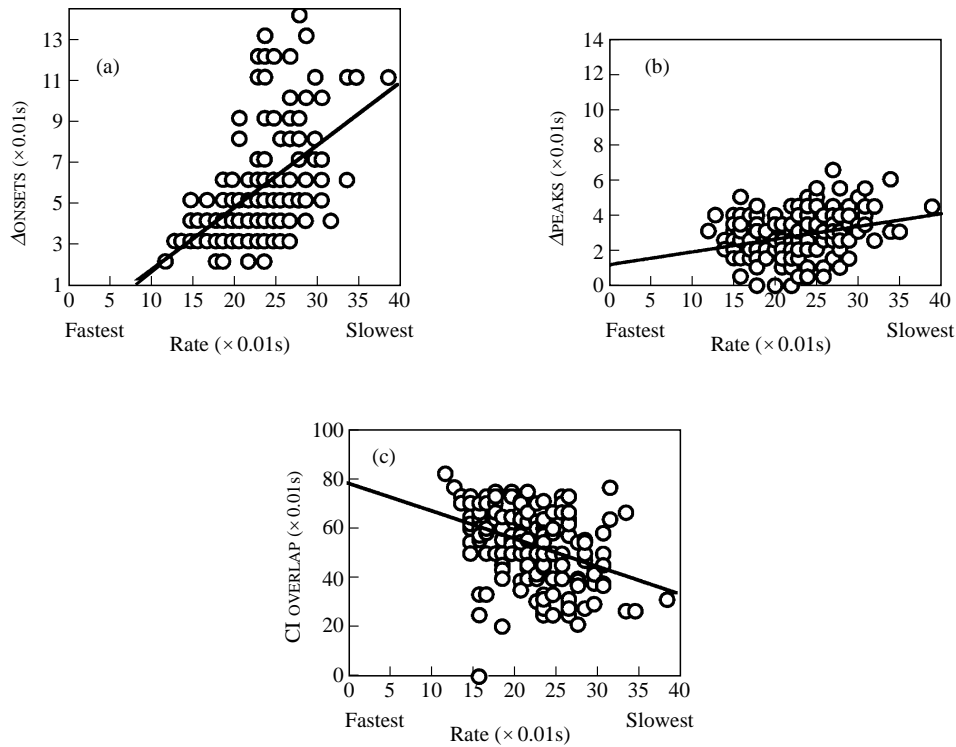


Figure 4. Plots for [g#d] for ΔONSETS (a), ΔPEAKS (b), and C1 OVERLAP (%) (c) regressed against rate (x-axis; fastest closest to the origin).

and absolute latency decreases with rate, and that these changes are generally linear.

3.2. Effects of speaking rate on individual consonants

3.2.1. Effects of speaking rate on duration

Next, recall the questions regarding shortening and reduction at faster speaking rates. The linear regressions (with speakers pooled) of contact duration against rate are evaluated for each of the three consonants in each syllabic position. These plots are shown in Fig. 7.

In general, all the consonants become shorter as speech rate increases. Even the extremely overlapped [d] shortens with rate. The one exception to this pattern is the [d] in onset position which shows no effect of rate on its duration. However, three of the five speakers (A, K, and S) do show significant shortening here with r^2 's for the linear fit as high as 0.573. Shortening is less robust for Speakers B and M than for the other speakers. Speaker M doesn't shorten either consonant in [d#g] and, of the onset consonants, shortens only the [g] in [s#g]. Speaker B shortens only the [d] in coda position and only the [s] in onset position. The other three speakers shorten all consonants in all clusters as speaking rate increases.

Thus, overall, we find support for the shortening of individual consonants as

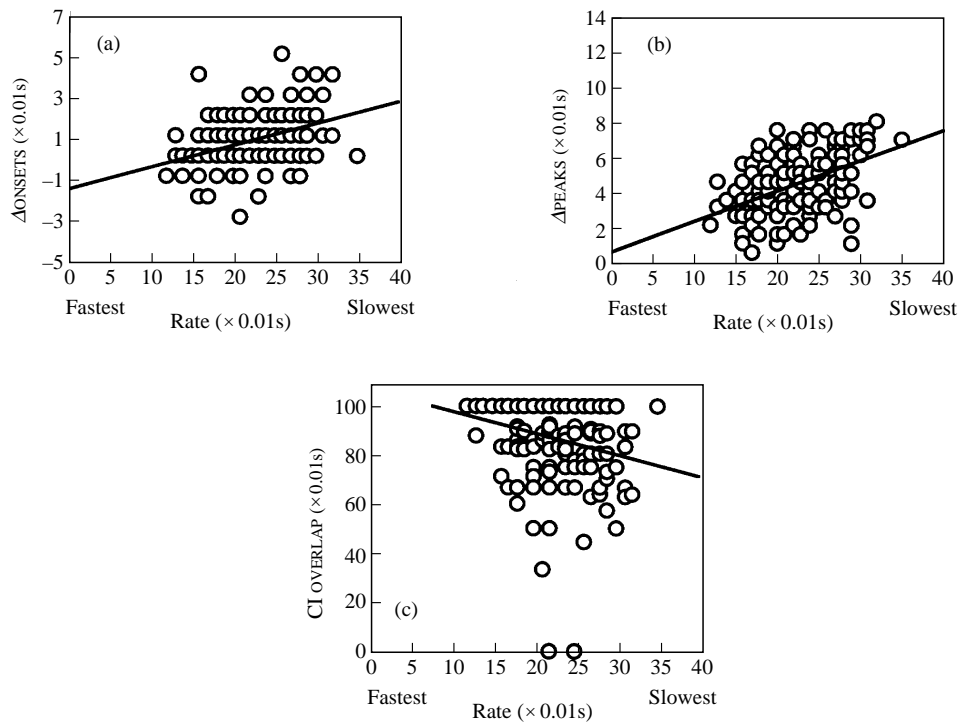


Figure 5. Plots for [d#g] for $\Delta ONSETS$ (a) (note the scale change here for plot a), $\Delta PEAKS$ (b), and CI OVERLAP (%) (c) regressed against rate (x-axis; fastest closest to the origin).

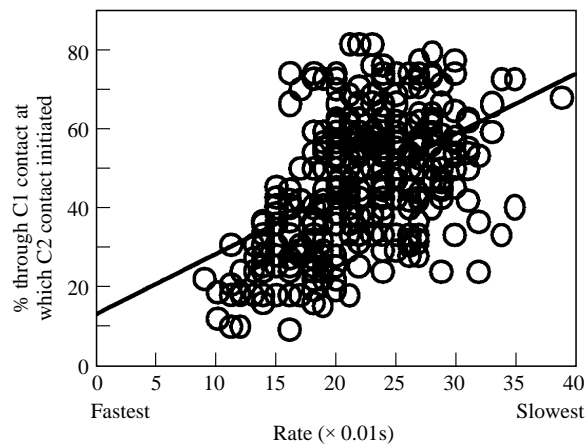


Figure 6. Plot for the percentage of the way through C1 contact at which C2 contact initiates (y-axis) regressed against rate (x-axis; fastest closest to origin); [d#g]'s and Speaker S omitted. $y = 12.179 + 1.556x$; $r^2 = 0.24$; $p < 0.001$.

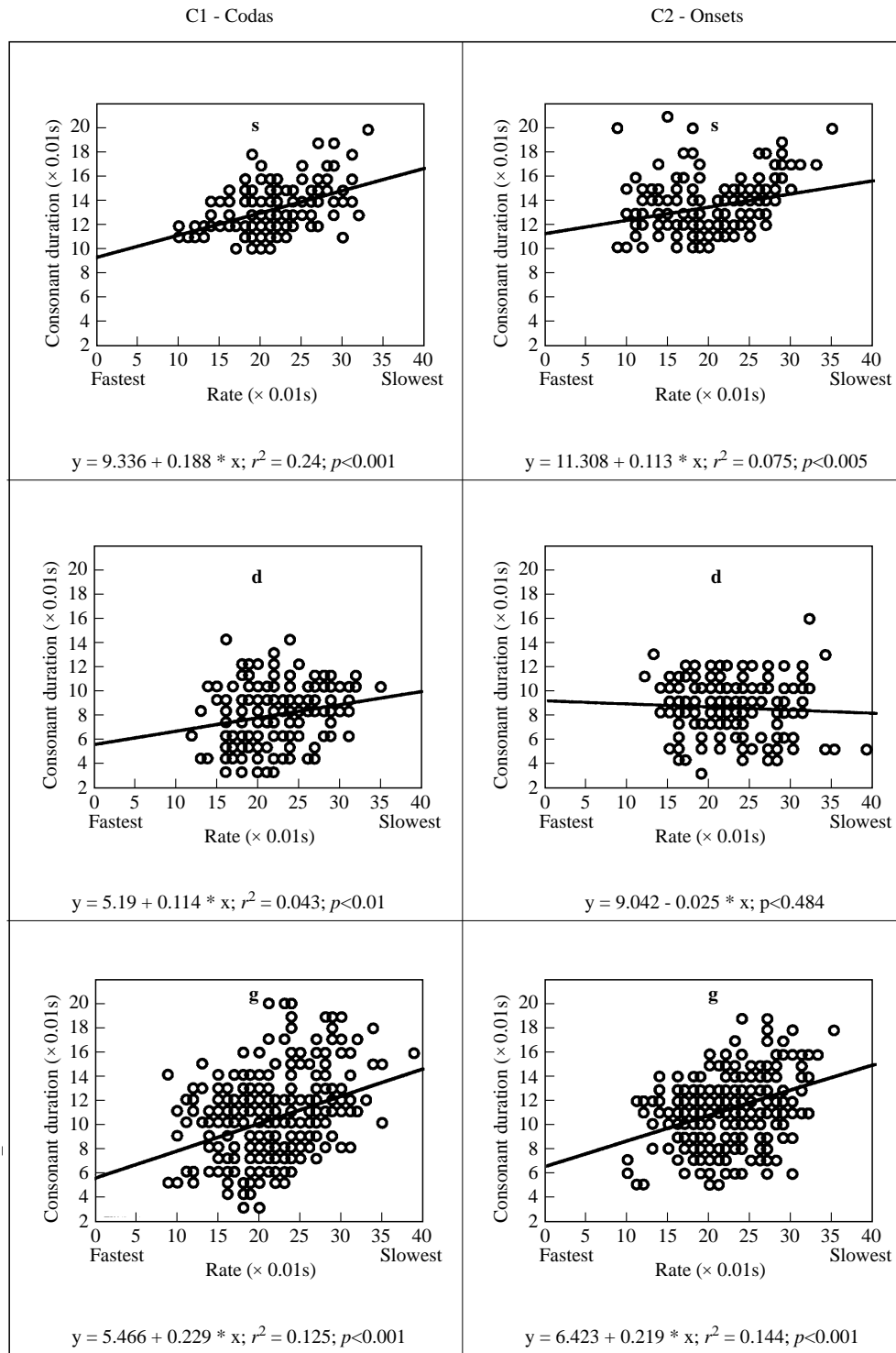


Figure 7. Plots for onset and coda durations (first contact to last contact in region) regressed against rate (x-axis; fastest closest to origin).

speaking rate increases. For most speakers, this shortening takes place regardless of the place and manner of the individual consonant or its syllabic affiliation.

3.2.2. *Effects of speaking rate on reduction*

The maximum contact percentage in the front and back regions for both consonants in each sequence was examined to determine if speaking rate affects displacement. The linear regressions for each of the three consonants in each syllabic position are shown in Fig. 8, with separate regression lines shown for each speaker. In two cases (coda [g] and onset [d]), four speakers have significant effects ($p < 0.02$) in an opposite direction from that of another speaker (Speaker B). The four significant r^2 for these two cases range from 0.15 to 0.27 for coda [g] and from 0.17 to 0.37 for onset [d]. In the other cases the r^2 's for the pooled data are given.

Rate has a significant effect on reduction of all consonants in coda, i.e., C1, position, although the effect on [s] is very small. The change in contact for coda [d] as a function of rate is probably under-represented due to a floor effect. There is no overall effect of rate for onset [s] and [g], but two speakers do reduce these consonants significantly as a function of rate (Speakers B and S for [s]; Speakers K and S for [g].) Four of the five speakers exhibit a strong effect of rate on the reduction of [d] in onset position as well. It appears that the consonants reduce somewhat more consistently in coda position than in onset position, as shown by the positive slopes in the left column of Fig. 8. Overall, it appears that stops are more likely to reduce than the fricative [s], and that coronals show greater reduction than the velar [g]. These results accord with results on positional (coda) reduction reported in Byrd (this issue).

4. Discussion

Early studies on speech variation found rate change to be largely implemented by varying pause length (Goldman-Eisler, 1968; Grosjean & Deschamps, 1975). However, further work determined that local variation in articulatory rate also exists (Miller, Grosjean & Lomanto, 1984). Our experiment assumed this to be the case. We did in fact find differences in latency, overlap, duration, and displacement in consonant sequence spanning a word boundary as a function of speaking rate.

The results evidence two articulatory mechanisms used in increasing speech rate. As speaking rate increases, individual consonants shorten in duration *and* a relatively linear increase occurs in the temporal coarticulation of the articulations. The sequences we studied, however, do not all behave identically.

With respect to overlap, rate has only a minimal effect on [d#g], which remains almost completely overlapped at all rates. These results provide support from articulatory data for Zsiga's (1994) acoustic study in which she found evidence for increased gestural overlap at fast rates for certain consonant sequences and for certain speakers. Other researchers using articulatory data have also found evidence of changes in articulatory overlap as a function of rate. Hardcastle (1985) reported an increase in overlap in [kl] sequences at faster rates. Shaiman, Adams & Kimelman (1995) found that interarticulator timing relationships were not constant across rates. They emphasized that the nature of their observed rate effect differed among individuals.

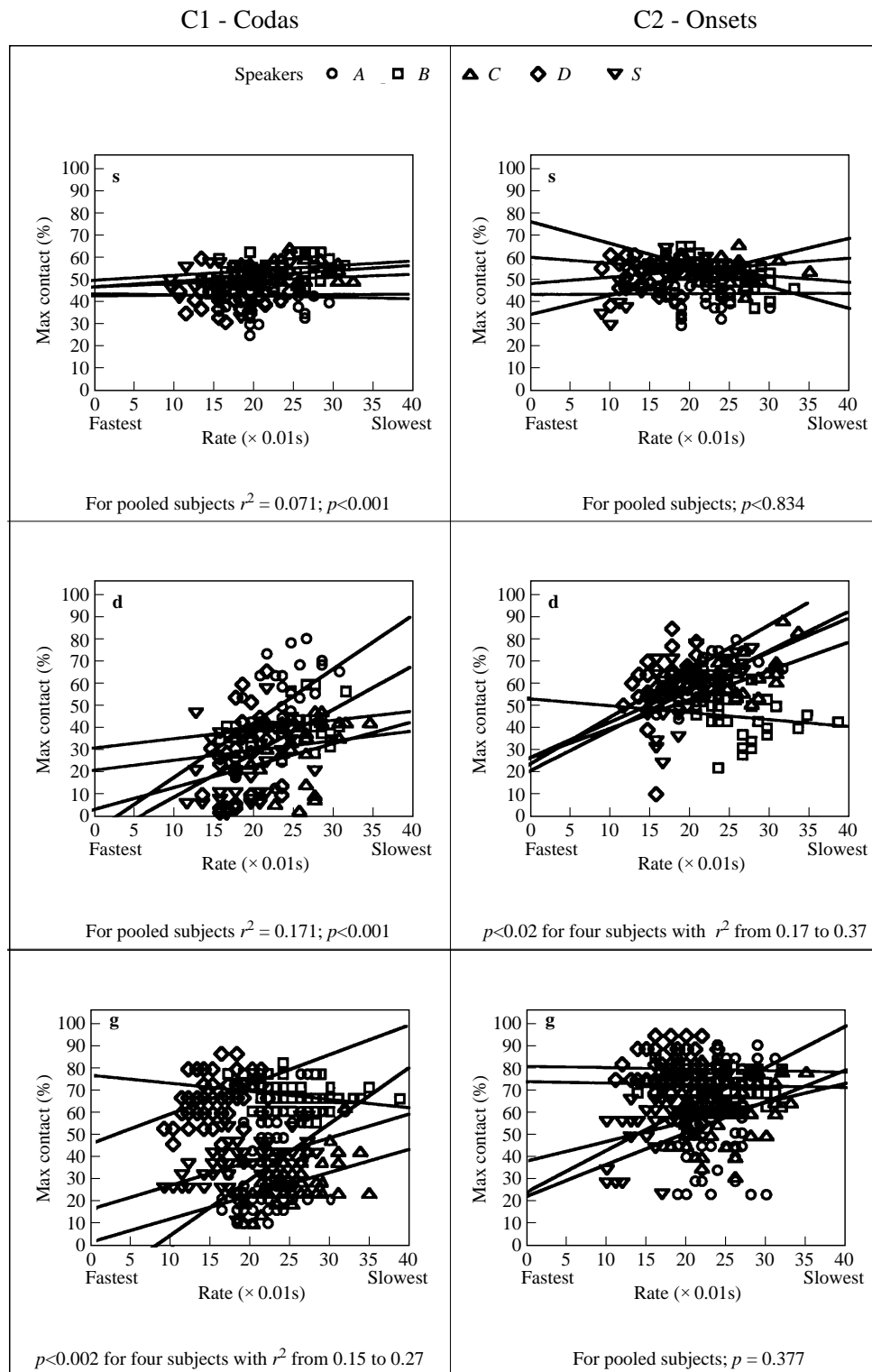


Figure 8. Plots for peak displacement (maximum percent contact in region) regressed against rate (x-axis; fastest closest to origin).

With respect to duration, we find that individual consonants shorten in duration as speaking rate increases, generally regardless of position, place, or manner. This is in accord with Gaitenby's (1965) finding that most segments have approximately the same ratio of segment-to-utterance length. But with respect to spatial reduction, we find spatial differences in the occurrence of reduction and in the degree of reduction depending on a consonant's manner of articulation and syllable position. In general, we find spatial reduction to be the less consistent concomitant of fast speech, occurring for fewer speakers, for fewer consonants, in fewer prosodic positions than does shortening. Researchers (Ostry & Munhall, 1985; Adams, Weismer & Kent, 1993) have found that while shortening occurs generally as rate increases, average displacement (or average peak velocity) may or may not change for any individual speaker as rate increases since it is the ratio of peak velocity to displacement that appears to be relevant in implementing the rate change. It could be the case that for the fricative, which was less susceptible to spatial reduction in our data, velocity differences were paramount, while for the stops reduction was more likely to occur. This makes sense in that reduction of [s] might be constrained for perceptual reasons (for a discussion see Kohler, 1992). If the displacement of the tongue tip were to decrease substantially, the narrow channel necessary for the production of fricative noise would cease to exist, thereby making the recovery of this consonant by the listener more at risk.

The reduction results also bear on a hypothesis about why coronals reduce more than non-coronals. Barry (1991) also found alveolar + velar clusters across morpheme and word boundaries to have alveolar reduction. In considering why alveolars are more susceptible to reduction, Barry recently (1992) proposed that the tongue tip be modeled in a task dynamics approach (Saltzman & Munhall, 1989) as a massless articulatory subsystem, unlike the tongue body. This predicts a facilitation of lenition in coronals because of a capacity for more rapid changes in direction of movement when confronted with competing demands. However, this approach alone does not account for the observed importance of prosodic affiliation in reduction or the difference between coronal stop and fricative in amount of reduction. Here, the different behavior of [d] versus [s] and onset [s] versus coda [s] shows that this reduction cannot be due *only* to the low mass of the tongue tip. If it were, the articulator would of necessity be different masses depending on manner of articulation and prosodic affiliation. This would keep "mass" from having any obvious physical interpretation. We should emphasize that Barry doesn't consider this to be the only cause of coronal reduction. In fact, he suggested that low-level processes such as reduction are to a degree under the "cognitive" control of the speaker. He proposed this in part due to the language-specific differences he found in alveolar reduction in Russian and in English. In order to resolve the question of whether alveolar reduction is purely a mechanical phenomenon, dynamic articulatory data from a variety of languages must be considered.

Finally, recall that we discussed the possibility that perceptual and/or acoustic goals could play a role in determining how a speaker increases speech rate. We agree with Gaitenby and others that many factors can influence speech rate, some "physiological, others linguistic, and still others... specific to the individual talker and occasion" (Gaitenby, 1965, p. 3.9). Some (Ladefoged, DeClerk, Lindau & Papçun, 1972; Johnson, Ladefoged & Lindau, 1993) have argued that speech movements are directed by auditory goals and have suggested that speaking is the

crucial determinant of the *organization* of speech articulation” (Johnson *et al.*, 1993, emphasis added). Others emphasize the importance of balancing listener-oriented efficiency and speaker-oriented efficiency (Lindblom, 1990 and elsewhere, Flege, 1988). Flege (1988) points out that:

“[a] balance of two countervailing forces influences how phonetic segments are articulated: the need to maintain sufficient distinctiveness between segments to ensure that words are recognized correctly, and the need to minimize effort while rapidly interleaving the multistructural movements that characterize successive phonetic segments.” (p. 901)

Surely, a primary goal of the speech process is for the listener to understand the speaker’s utterance. It is not unreasonable to assume that this goal constrains the motor strategies adopted for talking quickly and the degree to which each is employed. This leaves unexplored the issue of “listening quickly;” i.e., what, if any, adjustments are made by the listener to understand rapid speech.

We have seen for these data that absolute latency and the shortening of individual consonants are the most frequent means by which rate is reduced. Increased overlap and spatial reduction occur as concomitants of faster speech rate, but less frequently. It could be the case that decreasing latency and duration yield an acoustic signal in which perceptual cues used for recovering the consonants are less obscured than those available in a signal produced by consonantal gestures that are greatly overlapped and/or reduced. This may make decreasing consonant latency and duration more favored strategies for increasing rate. It seems clear, however, that speakers do generally employ a variety of means of changing their speech rate.

4. Conclusion

This experiment demonstrates that for a set of obstruent consonants in sequences spanning a word boundary, talking faster can mean decreasing articulatory durations *and* increasing coproduction between successive articulations. At the same time, articulatory specifics such as spatial reduction and temporal overlap may depend on linguistic structure and differ across individuals.

This research was completed in the Department of Linguistics at UCLA and is largely excerpted from a Ph.D. dissertation by Dani Byrd entitled *Articulatory Timing in English Consonant Sequences*. Some of this work, “Rate and reduction in consonant sequences” by Dani Byrd, was also presented at the 1994 meeting of the Linguistic Society of America. The authors are grateful to Patrice Beddor, J. E. Flege, Edward Flemming, Keith Johnson, Pat Keating, Peter Ladefoged, Björn Lindblom, Ian Maddieson, and an anonymous reviewer for their insightful comments and assistance. We also thank our speakers for their generous participation. This research was supported by an NSF Graduate Fellowship, by an NSF Dissertation Improvement Grant #DBS 9213604 to the first author and Patricia Keating, by the Department of Linguistics at UCLA, and by NIH grant DC-00016.

References

- Abbs, J. H. (1973) The influence of the gamma motor system on the jaw movements during speech: a theoretical framework and some preliminary observations, *Journal of Speech and Hearing Research*, **16**, 175–200.
- Adams, S. G., Weismer, G. & Kent, R. D. (1993) Speaking rate and speech movement velocity profile, *Journal of Speech and Hearing Research*, **36**, 41–54.
- Barry, W. (1983) Some problems of interarticulator phasing as an index of temporal regularity in speech, *Journal of Experimental Psychology: Human Perception and Performance*, **9**, 826–828.

- Barry, M. (1991) Temporal modelling of gestures in articulatory assimilation. In *Proceedings of the XIIth international congress of phonetic sciences*, pp. 14–17. Aix-en-Provence: Université de Provence.
- Barry, M. (1992) Palatalisation, assimilation and gestural weakening in connected speech, *Speech Communication*, **11**, 393–400.
- Benguerel, A. & Cowan, H. A. (1974) Coarticulation of upper lip protrusion in French, *Phonetica*, **30**, 41–55.
- Browman, C. & Goldstein, L. (1990) Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology I: between the grammar and the physics of speech*. (J. Kingston & M. E. Beckman, editors), pp. 341–376. Cambridge: Cambridge University Press.
- Byrd, D. (this issue) Influences on articulatory timing in consonant sequences, *Journal of Phonetics*, **24**, 209–244.
- Byrd, D. (in press) A phase window framework for articulatory timing, *Phonology*.
- Byrd, D., Flemming, E., Mueller, C. A. & Tan, C. C. (1995) Using regions and indices in EPG data reduction, *Journal of Speech and Hearing Research*, **38**, 821–827.
- Dixit, R. P. & Flege, J. E. (1991) Vowel context, rate and loudness effect of linguopalatal contact patterns in Hindi retroflex [ɭ], *Journal of Phonetics*, **19**, 213–229.
- Engstrand, O. (1988) Articulatory correlates of stress and speaking rate in Swedish VCV utterances, *Journal of the Acoustical Society of America*, **83**, 1863–1875.
- Flege, J. E. (1976) Plasticity in adult and child speech production, *Journal of the Acoustical Society of America*, **79**, S54.
- Flege, J. E. (1988) Effects of speaking rate on tongue position and velocity of movement in vowel production, *Journal of the Acoustical Society of America*, **84**(3), 901–916.
- Fletcher, S. G., McCutcheon, M. J., Wolf, M. B., Sooudi, I. & Smith, T. L. (1975) Palatometric-gnathometric study of speech articulation. Final Report. (NIH Grant No. NS10540).
- Fowler, C. A. (1977) *Timing control in speech production*. Ph.D. dissertation, University of Connecticut.
- Gaitenby, J. (1965) The elastic word, *Haskins Laboratories: Status Report on Speech Research*, **SR-2**, 3.1–3.12.
- Gay, T. (1968) Effect of speaking rate on diphthong formant movements, *Journal of the Acoustical Society of America*, **44**(6), 1570–1575.
- Gay, T. (1974) A cinefluorographic study of vowel production, *Journal of Phonetics*, **2**, 255–266.
- Gay, T. (1981) Mechanisms in the control of speech rate, *Phonetica*, **38**, 148–158.
- Gay, T. & Hirose, H. (1973) Effect of speaking rate on labial consonant-vowel articulation, *Journal of Phonetics*, **2**, 47–63.
- Goldman-Eisler, F. (1956) The determinants of the rate of speech output and their mutual relations, *Journal of Psychomatic Research*, **1**, 137–143.
- Grosjean, F. & Deschamps, A. (1975) Analyse contrastive des variables temporelles de français spontané. II. Comparaison du français oral dans la description avec l'anglais (description) et avec le français (interview radiophonique), *Phonetica*, **28**, 191–226.
- Hamlet, S. L. & Stone, M. (1978) Compensatory alveolar consonant production induced by wearing a dental prosthesis, *Journal of Phonetics*, **6**, 227–2481.
- Hardcastle, W. J. (1972) The use of electropalatography in phonetic research, *Phonetica*, **25**, 197–215.
- Hardcastle, W. J. (1985) Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences, *Speech Communication*, **4**, 247–263.
- Hardcastle, W. J., Gibbon, F. & Nicolaidis, K. (1991) EPG data reduction methods and their implications for studies of lingual coarticulation, *Journal of Phonetics*, **19**(3), 251–266.
- Hardcastle, W. J. & Roach, P. (1979) An instrumental investigation of coarticulation in stop consonant sequences. In *Current issues in the phonetic sciences* (H. Hollien & P. Hollien, editors), pp. 531–540. Amsterdam: John Benjamins.
- Hughes, O. M. & Abbs, J. H. (1976) Labial-mandibular coordination in the production of speech: implications for the operation of motor equivalence, *Phonetica*, **33**, 199–221.
- Johnson, K., Ladefoged, P. & Lindau, M. (1993) Individual differences in vowel production, *Journal of the Acoustical Society of America*, **94**(2), 701–714.
- Kent, R. D. (1970) *A cinefluorographic-spectrographic investigation of the component gestures in lingual articulation*. Unpublished Ph.D. Dissertation, University of Iowa.
- Kent, R. D. & Moll, K. L. (1972) Cinefluorographic analyses of selected lingual consonants, *Journal of Speech and Hearing Research*, **15**, 453–473.
- Kohler, K. J. (1992) Gestural reorganization in connected speech: a functional viewpoint on 'Articulatory Phonology', *Phonetica*, **49**, 205–211.
- Kozhevnikov, V. & Chistovich, L. (1965) *Speech: articulation and perception* (Translated and distributed by Joint Publications Research Service, Washington D.C., N. JPRS30, 543, Moscow-Leningrad. Originally published in Russian, 1965., Trans.). Washington DC: Joint Publications Research Service No. 30.
- Ladefoged, P., DeClerk, J., Lindau, M. & Papçun, G. (1972) An auditory-motor theory of speech production, *UCLA Working Papers in Phonetics*, **22**, 48–75.

- Lindblom, B. (1963) Spectrographic study of vowel reduction, *Journal of the Acoustical Society of America*, 35(11), 1773–1781.
- Lindblom, B. (1964) Articulatory activity in vowels, *STL-QPSR*, 2, 1–5.
- Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H&H theory. In *Speech production and speech modeling* (W. J. Hardcastle & A. Marchal, editors), pp. 403–39. Dordrecht: Kluwer.
- Löfqvist, A. & Yoshioka, H. (1981) Laryngeal activity in Icelandic obstruent production, *Nordic Journal of Linguistics*, 4, 1–18. (Also published in *Haskins Laboratories: Status Report on Speech Research* SR-63/64, 1980, 275–292).
- Miller, J. L., Grosjean, F. & Lomanto, C. (1984) Articulation rate and its variability in spontaneous speech: a reanalysis and some implications, *Phonetica*, 41, 415–225.
- Munhall, K. G. & Löfqvist, A. (1992) Gestural aggregation in speech: laryngeal gestures, *Journal of Phonetics*, 20, 93–110.
- Ostry, D. J. & Munhall, K. G. (1985) Control of rate and duration of speech movements, *Journal of the Acoustical Society of America*, 77(2), 640–648.
- Saltzman, E. & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production, *Ecological Psychology*, 1, 333–382.
- Shaiman, S., Adams, S. G. & Kimelman, M. D. Z. (1995) Timing relationships of the upper lip and jaw across changes in speaking rate, *Journal of Phonetics*, 23, 119–128.
- Zsiga, E. C. (1994) Acoustic evidence for gestural overlap in consonant sequences, *Journal of Phonetics*, 22, 121–140.

Appendix: Artificial palates with defined regions

These pictures show the artificial acrylic palates worn by each speaker. The pictures are to scale, and the front of the mouth is oriented to the top of the pictures. Some impression of the depth of the speakers' palatal vaults can be gained by noting the darkness in this area, with increased darkness being associated with increased depth. The electrodes are shown by small circles. The electrodes that were excluded from both regions are shown by X's with white centers. (See text for the criteria for region definition). The heavy dark horizontal line marks the division of the remaining electrodes into front and back region groups.

