DETERMINING THE EXTENT OF COARTICULATION: EFFECTS OF EXPERIMENTAL DESIGN

Carol E. Golfer,† Fredericksa Bell-Berti,‡ and Katherine S. Harris†

Abstract. Substantial differences in the reports of the extent of anticipatory coarticulation have made the task of deciding among unifying models of the process difficult. Two conceptually distinct groups of theories of coarticulation have emerged, one positing the migration of articulatory features to preceding segments and the other positing the temporal cohesiveness of the components of segmental articulations. In studies of anticipatory lip rounding, a possible source of the differences reported in its extent prior to a rounded vowel is that the alveolar consonants commonly employed in these studies are presumed to be unspecified with regard to lip configuration. Thus, the presence of EMG activity and/or protrusive lip movement during these consonants has been presumed to indicate vocally conditioned lip activity. However, if this activity is directly related to the production of the consonant(s), then the interpretation of these results is problematic unless the experimental design allows for the differentiation of consonantal and vocalic effects. We offer here both data suggesting the need for such considerations and a paradigm that takes these considerations into account.

Introduction

The phenomena of anticipatory coarticulation have generally been presumed to reflect underlying aspects of speech motor control (e.g., Kozhevnikov & Chistovich, 1961; MacNeill, 1970). However, substantial differences in reports of the extent of anticipatory coarticulation make difficult the task of providing one model to account for these data. Two types of conceptually distinct theories of anticipatory coarticulation exist, both of which attempt to explain the apparently nondiscrete nature of speech output despite a presumed discrete input. According to one type of theory, upcoming phones are scanned for salient features, which then migrate to as many antecedent phones as are neutral for, or in no way antagonistic to, the migrating feature (e.g., Bahlhoff & Noll, 1968; Henke, 1966; Kozhevnikov & Chistovich, 1961; Summan & Westbury, 1981). Thus, given some number of consonants unspecified for lip configuration immediately preceding a rounded vowel, these models predict that rounding will vary in its onset in direct proportion to the number and/or

#A version of the paper was presented at the 103rd Meeting of the Acoustical Society of America. Chicago, Ill, May 1982.
†Also The Graduate Center, The City University of New York.
‡Also St. John's University.
Acknowledgment. This work was supported by NINCDS grant NS-13617 and BRS grant RR-05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-82/83 (1985)]
duration of preceding segments. For example, Benguerel and Cowan (1974) reported that upper lip protrusion (in anticipation of a rounded vowel) begins as early as the first consonant in clusters of as many as six consonants. However, the second type of theory proposes that the observed co-occurrence of components of proximate segments results, not from feature migration, but from the overlapping of articulatory components of those segments (e.g., Bell-Verti & Harris, 1981, 1982; Fowler, 1980). Thus, in the absence of conflicting demands, the onsets of different components of the articulation of a given phone will bear a stable temporal relationship to each other. For example, Engstrøm (1981) reported that lip protrusion activity for the rounded vowel /u/ occurs at a relatively fixed time before the onset of voicing for that vowel, regardless of the number of preceding consonants.

Despite their conceptual differences, however, a basic premise, having its roots in traditional linear generative phonology, is common to these models: namely, that a phone is neutral (i.e., unspecified) for a particular feature when that feature is not essential to its realization (Chomsky & Halle, 1968, pp. 402-403). Consequently, when activity associated with a given feature occurs during a segment that is "neutral" for that feature, that activity must be associated with another segment, and the time at which this activity begins is then assumed to reflect the extent of anticipatory coarticulation. In fact, however, it may be that feature descriptions are incomplete. For example, as Benguerel and Cowan (1974) have noted, American English /r/ is commonly produced with lip protrusion, although this protrusion often goes unmentioned in articulatory descriptions of /r/.

Upon closer consideration, it would appear that many of the differences in the existing literature might be reconciled, and thus allow the development of a single explanation for them, were these assumptions reconsidered. The work presented here is part of a study designed to account for the conflicting results of previous studies, and therefore to test the predictions of the different models of anticipatory coarticulation.

Methods

The alveolar consonants /t/ and /s/, whose articulation would be presumed to be neutral for lip constriction, were combined to form nine sequences designed to vary both in the number of consonants and in overall sequence durations. The vowels in these utterances were /i/ and /u/, where V₁ was always /i/, while V₂ was either /i/ or /u/. Thus, there were two vowel conditions, the /IC1/ and /IC2/ conditions, each occurring with the nine different consonant string combinations, for a total of eighteen utterance types (Table 1). The sequences were made by combining "words," and were presented to the subjects in orthographic writing. The subjects were instructed to speak at a comfortable rate, in a conversational manner, without undue attention to marking word boundaries. Thus, the subjects could, and did, differ in the ways in which they executed a given sequence (for example, /joʊnd toʊl/ was often realized as the sequence [ˈlaʃtoʊl]). Two native speakers of American English, produced between fifteen and twenty repetitions of each of the eighteen VC₁V₂, spoken within the carrier phrase "It's a ________ again."

Surface electromyographic (EMG) recordings (Allen, Lubker, & Harrison, 1972) of orbicularis oris inferior (O01), right and left, were made simultaneously with lip movement recording. Lip movements were tracked with an optoelectrical tracking system (Capstan Co. Model 400 Optical Tracking System).
that sensed the position, in both the x and y planes, of an infrared light-emitting diode (LED) positioned on the lower lip. All data were simultaneously recorded on a 14-channel FM tape.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Number of Consonants</th>
<th>Consonant String Duration (in milliseconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TB</td>
</tr>
<tr>
<td>i$tu</td>
<td>1</td>
<td>75</td>
</tr>
<tr>
<td>i$su</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>i$stu</td>
<td>2</td>
<td>245</td>
</tr>
<tr>
<td>i$#tu</td>
<td>2</td>
<td>230</td>
</tr>
<tr>
<td>i#su</td>
<td>2</td>
<td>300</td>
</tr>
<tr>
<td>i#stu</td>
<td>3</td>
<td>305</td>
</tr>
<tr>
<td>i##tu</td>
<td>3</td>
<td>280</td>
</tr>
<tr>
<td>i##su</td>
<td>3</td>
<td>385</td>
</tr>
<tr>
<td>i#stu</td>
<td>4</td>
<td>360</td>
</tr>
<tr>
<td>i@ti</td>
<td>1</td>
<td>83</td>
</tr>
<tr>
<td>i@si</td>
<td>1</td>
<td>227</td>
</tr>
<tr>
<td>i@#ti</td>
<td>2</td>
<td>240</td>
</tr>
<tr>
<td>i@#si</td>
<td>2</td>
<td>335</td>
</tr>
<tr>
<td>i@#stl</td>
<td>3</td>
<td>331</td>
</tr>
<tr>
<td>i@#stl</td>
<td>3</td>
<td>284</td>
</tr>
<tr>
<td>i#stl</td>
<td>3</td>
<td>391</td>
</tr>
<tr>
<td>i#stl</td>
<td>4</td>
<td>392</td>
</tr>
</tbody>
</table>

The EMG signals were rectified, and both the EMG and movement data were integrated and then digitized using a PDP 11/45 computer. The durations of the consonant strings were measured for each token of each utterance type, using a PCM waveform-editing program. The beginning of the consonant string was defined as the point at which either the friction appeared in the waveform (in consonant strings beginning with /t/), or the higher formants disappeared from the waveform (indicating the onset of closure in consonant strings beginning with /t/). The point in the acoustic signal corresponding to the release of the consonant occlusion immediately preceding V, was identified as the end of the consonant string and served as the acoustic reference, or line-up, point for subsequent ensemble averaging. Thus, when V, was preceded by /t/,
the line-up point was the burst; when \( V \) was preceded by /u/, the line-up point was the end of friction before the second vowel.

The beginning of OOI activity associated with the /IC\(_u\)/ sequences was determined by identifying the time at which the EMG activity increased to a level equivalent to the baseline plus five percent of the difference between the baseline and the peak EMG levels. The beginning of the related movement was determined by identifying the onset of anteriorly-directed lip movement.

**Results**

Some representative EMG data are shown for each subject (Figure 1a). The EMG signals in each panel represent the ensemble average OOI EMG activity of an /IC\(_u\)/ utterance, with consonant string length (i.e., both the number of segments and the durations of the sequences) differing across panels. The onset of EMG activity occurs earlier as consonant string duration increases, so that it would appear that there has been a migration of lip rounding back to the beginning of the consonant string. In fact, when the onset of OOI EMG activity for each of the nine /IC\(_u\)/ utterances is plotted against the respective consonant string duration (Figure 1b), it seems that, for both subjects, these onsets bear an obvious relationship to consonant string duration. That is, they occur earlier as string duration increases, with correlation coefficients of r=.98 and .91 for TB and CH, respectively.

Although these results might be interpreted as evidence that lip rounding has spread to the beginning of the "neutral" consonant string, we believe that it is imperative to determine whether all of the EMG activity is actually vowel-related or, alternatively, if it reflects consonantal lip gestures. In other words, if the OOI activity during the consonant string is vowel-related, we would not expect to find such activity during the same consonant string when it is followed by an unrounded vowel. We therefore examined OOI activity for the minimally contrastive /IC\(_u\)/ utterances, samples of which are shown in Figure 2a. It is clear that, even within this unrounded vowel environment, there is a significant amount of orbicularis oris activity during the consonant string articulation. In fact, if we treat these /IC\(_u\)/ data as we did those for the /IC\(_u\)/ utterances, identifying the onset of EMG activity for each utterance and plotting these times against consonant string durations (Figure 2b), the resulting scatter plots are strikingly similar to those for the /IC\(_u\)/ utterance set (Figure 1b). That is, OOI activity begins earlier as consonant string duration increases. (Subject CH produced only eight of the nine /IC\(_u\)/ utterances.) Obviously, then, this EMG activity cannot reflect the onset of vowel-related lip rounding (i.e., the migration of the vowel feature) since the relationship between consonant string duration and the onset of OOI activity is observed in both rounded and unrounded vowel environments. Indeed, correlation coefficients are as high or higher for these /IC\(_u\)/ utterances (r=.98 and .99 for TB and CH, respectively) than they are for their rounded counterparts.

It is obvious, then, that the progressively earlier EMG activity must reflect consonant-related events. This is more apparent when the EMG curves for the minimally contrastive /IC\(_u\)/ and /IC\(_u\)/ utterances are superimposed (Figure 3). The two signals diverge in the vicinity of the acoustic onset of \( V \), with a second peak of activity evident when \( V \) is /u/, while EMG activity is suppressed when \( V \) is /i/. However, because the EMG signal never returns to a baseline level prior to /u/, the onset of the /u/-related
Figure 1. Upper panels (a): Ensemble-average EMG data for subjects TB (left) and CH (right) recorded from orbicularis oris inferior (OOI) for three /I Canyon/ utterances. Lower panels (b): EMG onset time (ms before line-up point) vs. consonant string duration for /I Canyon/ utterances. Time 0 represents the release of the consonant occlusion, determined from the acoustic waveform. The arrows indicate the average of the acoustic onsets of the consonant strings.
Figure 2. Upper panels (2a): Ensemble-average EMG data for subjects TB (left) and CH (right) recorded from orbicularis oris inferior (OOI) for three /t/ utterances. Lower panels (2b): EMG onset time (ms before larynx-up point) vs. consonant string duration for /t/ utterances. Time 0 represents the release of the consonant occlusion, determined from the acoustic waveform. The arrows indicate the average of the acoustic onsets of the consonant strings.
Figure 3. Ensemble-average EMG data for the two subjects, recorded from orbicularis oris inferior (OOI) for three minimally contrastive pairs of /i/, /u/ utterances.

Figure 4. Statistically determined point of separation ("EMG separation onset") between minimally contrastive pairs of /i/, /u/ utterances vs. the average duration of the consonant sequences of the /i/, /u/ utterances of each pair.
EMG activity was determined statistically as the time at which the difference (in microvolts) following the divergence of the two signals reached significance (p<.05).

The statistically determined onsets of rounded vowel activity are plotted as a function of consonant string duration for the nine minimal pairs for subject TB, and for eight minimal pairs for subject CH (Figure 4). In contrast to the consonant-related EMG activity (see Figures 1 and 2), these onsets bear no obvious relation to the durations of the consonant strings. Rather, with the exception of the /i#t/ utterance, they occur within a fairly restricted range, bearing a stronger relationship to the onset of the rounded vowel than to the onset of the consonant string.

The EMG data thus show the following: First, for these two subjects, some lip activity appears to be inherent in the production of alveolar consonants. Second, the onset of EMG activity for /u/ appears to be related to the acoustic onset of that vowel, and not to the compatibility of the vowel and consonant articulations. Finally, even when there is lip activity for adjacent consonants and vowels, they appear to be organized as independent gestures, as the separate peaks of OOI activity for the /i#t/ utterances suggest.

Figure 5 shows movement data for both subjects, for the same /i#t/ utterances whose EMG data are presented above (Figure 1a). For TB, the data show a substantial forward lip movement in the vicinity of the acoustic onset of the consonant string, a position that is then sustained through V1. However, while there is a less obvious separation between the consonant and vowel gestures in the movement than in the EMG records, there are troughs in the movement traces for all but the shortest utterance. For subject CH, the anterior lip movement associated with the rounded vowel is more clearly separated from the anterior movement occurring earlier in the utterance.

When the movement traces for the /i#t/ and /i#t/ utterances are superimposed (Figure 6), the pattern is the same as that for the EMG records. That is, regardless of the identity of V1, the curves are nearly identical through the consonant string, diverging in the vicinity of the onset of the second vowel. However, because of hardware limitations at the time of recording, the baselines for these data are not always aligned; for this reason we were unable to determine statistically the times at which each minimally contrasting pair differed, as we had done for the EMG data. Furthermore, when the temporal relationships between the consonant-related EMG and the earliest anteriorly directed movements are examined, there are clearly differences for the two subjects. For subject TB, the earlier onset of OOI activity is associated with consonant-related forward lip movement. That is, there is an appropriate contraction time interval between the EMG and corresponding movement (Figure 7a). For subject CH, however, the earlier OOI activity is not associated with any significant anterior lip movement for the consonant string (Figure 7b). Rather, this movement is associated with the first vowel.

We are therefore faced with the question of what the consonant-related EMG activity means in terms of movement for subject CH. Figure 8 shows OOI activity for the three representative /i#t/ utterances, along with both the corresponding horizontal and vertical movement traces. It can be seen that, while the EMG and horizontal lip movements are poorly correlated in the vicinity of the consonant string, there is a good temporal correlation (i.e.,
Figure 5. Antero-posterior lip position as a function of time for the two subjects for three /tʌ/ utterances. The arrows indicate the average of the acoustic onsets of the consonant strings.

Figure 6. Antero-posterior lip position for both subjects as a function of time for three minimally contrastive pairs of /tʌ/ utterances.
Figure 7. Ensemble-average EMG and lip position data as a function of time for both subjects for three /ɪɾ/ utterances.

Figure 8. Ensemble-average EMG and lip position data for subject CH for three /ɪɾ/ utterances. The thin line represents OOI-R data, the thick line anterior lip position data (lip X), and the dashed line vertical lip position data (lip Y).
termination interval) between the consonant-related EMG and vertical lip movement. Thus, for this subject, the same muscle appears to be contributing to both vertical movement (in the production of the consonant string) and horizontal movement (in the production of the vowel), differences in orbicularis oris function that have been noted previously (cf. O'Dwyer, Quinn, Guitton, Andrews, & Neilson, 1981).

Discussion

The data offered here suggest that there are a number of reasons for the difficulty in reconciling the differences between sets of previously reported data on the extent of anticipatory coarticulation. One of these reasons resides in the unproven assumption that, in a speech sound's articulation has not been described as including a particular gesture, then, first, that gesture has little, if any, consequence for the production of the sound and, second, that speech sound is "unspecified" for that gesture/feature. However, phoneticians have long known that the description of the articulation of speech sounds is incomplete (cf. Pike, 1943, p. 152); our data clearly indicate that, for some speakers at least, some alveolar consonants traditionally assumed to have no intrinsic lip gestures do in fact have such gestures as part of their natural production. Thus, the assumption that these consonants are neutral with regard to lip configuration is untenable.

These data also provide evidence of the complexity of the electromyographic and kinematic data collected for studying coarticulation processes. First, it is impossible to separate active protrusion gestures from passive relaxation of lips that have been retracted, except by observing the activity of the muscles responsible for those protrusion gestures. Second, the EMG data may more closely reflect the underlying segmental structure of speech than do kinematic data. For example, while we see no trough in the movement traces of the /iʃu/ utterance for subject TB, there are clearly separate peaks of OOI activity for both the consonant and vowel segments, suggesting the segmental nature of the underlying articulatory organization.

In addition to providing insights into the causes of some of the apparent discrepancies resulting from problems in experimental design, we would also suggest that another source of conflict in attempts to develop a single model of anticipatory phenomena stems from presupposing that the timing of the onset of rounding is an entirely anticipatory phenomenon. It is notable that in both this study and our earlier work (Bell-Berti & Harris, 1981), the onset of vowel-related lip rounding is closer to the acoustic onset of the rounded vowel for sequences of the form /iʃu/ than for any other sequence. This result might seem to provide some limited support for the feature migration hypothesis, if this sequence were compared with only one longer sequence (see, e.g., Sussman & Westbury, '81). However, we believe that an equally plausible explanation is that the result reflects the suppression of lip rounding until the first vowel can be completed without distortion. That is, the onset of rounding may be constrained by the carryover effects of a preceding (unrounded) vowel. Thus, in a sequence like /iʃu/, where the vowel-to-vowel interval is fairly short, the rounding onset might be delayed relative to other sequences where the consonantal sequence occupies a longer time slot. In fact, Sussman and Westbury's (1981) observation of systematic differences in the onset of lip rounding as a function of the identity of the preceding unrounded vowel may be interpreted as evidence of the same carryover effect.
Summary

These data were part of a study designed to account for conflicting results of previously reported studies by suggesting that at least some of the apparent discrepancy arises from experimental design. Because our two subjects produced alveolar consonants with significant orbicularis oris activity in both rounded and unrounded vowel environments, we were able to establish that those gestures that were variable in their onsets on both the EMG and movement levels were clearly tied to something that was acoustically variable as well—namely, the onsets of consonant strings of differing durations. We also observed separate consonant and vowel-related activity, as in the EMG recordings of the /i/-/u/ utterances, where there were almost always distinct peaks for each. Furthermore, our EMG data may be interpreted as reflecting a stable onset of lip rounding independent of consonant string duration, except for the case of the shortest consonant string. And, while the tendency has been to view all of these phenomena as reflecting only anticipatory coarticulation, we believe it more likely that they represent the combined effect of carryover and anticipatory processes.

References


Footnotes

1We have limited ourselves here primarily to a consideration of anticipatory phenomena. This limitation was imposed because most theoretical discussions have focused on anticipatory coarticulation.

2The literature in this area contains two different indices to consonant string length: the number of consonant segments (e.g., Danillof & Moll, 1968; Lubker & Gay, 1980) and the duration of the consonant sequence (e.g., Bell-Berti & Harris, 1974, 1982; Engstrand, 1981). Although these two measures are related, the relationship is not isomorphic (see, for example, Table 1).

3Subject TB is a speaker of educated Greater Metropolitan New York City English. Subject CH is a speaker of educated Central Florida English.

4This result is compatible with results of other studies using subjects known to produce the alveolar consonants /s/ and /t/ without lip rounding (cf. Bell-Berti & Harris, 1982; Engstrand, 1981), although these studies clearly still subscribe to the possibility that alveolar consonants have inherently neutral lip specifications.

5The observation of "troughs" in EMG and movement records is not new (cf. Bell-Berti & Harris, '87; Engstrand, 1983; Gay, 1977). The fact that a trough is absent in movement records when the intervocalic consonant is short may not reflect differences in gestural organization, but, rather, biomechanical constraints that could influence the response characteristics of the lips. That is, with movement being rather slow relative to EMG activity, it is hardly surprising that the lips do not have time to protrude, retract, and protrude again for the rounded vowel during the 75 ms /t/ closure.

6We would note, however, that there was no consistent pattern of DC offsets between the /1C1/ and /1C2/ utterances, suggesting that these differences were independent of vowel rounding.