

# Coarticulation is largely planned

713

D. H. Whalen

*Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, U.S.A.*

*Received 11th August 1989, and in revised form 13th November 1989*

---

Is coarticulation a reflection of planning an utterance or the automatic effect of producing speech? This question was examined by requiring speakers to begin reading nonsense / $\alpha$ CV/ or / $\alpha$ bVC $\alpha$ / strings (where C was /b/ or /p/ and V was /i/ or /u/) aloud before seeing the entire utterance. Those segments known before articulation began exerted normal anticipatory coarticulatory influences, while those seen after utterance onset usually did not. Perseverative coarticulation was, for the most part, present in both conditions, showing either that it is indeed a result of the mechanical nature of the speech organs or that the amount of planning necessary can be carried out in the time used for the first syllables. There was one anticipatory effect which began to emerge (even though unplanned) just before the stop closure, supporting an articulatory overlap explanation for some of the coarticulation. From the greater magnitude and extent of the coarticulatory influence of segments known in advance, we can infer that, for the most part, coarticulation is planned.

---

## Introduction

Coarticulation, the influence of one phoneme on another, has received a great deal of attention in the phonetics literature (e.g., Kent & Minifie, 1977; Liberman & Studdert-Kennedy, 1978; Fowler, 1980, 1984; Keating, 1985; Liberman & Mattingly, 1985). This attention is well deserved, since the widespread and complex set of influences among phonemes results in the lack of simple acoustic invariance in the speech signal which has made the field of phonetics so challenging. But while coarticulation clearly influences the output of the speech system, does it do so as part of the motor program or as a consequence of executing that program? The work of Sternberg and his colleagues (Sternberg, Monsell, Knoll & Wright, 1978; Sternberg, Wright, Knoll & Monsell, 1980; Monsell, 1986) has provided evidence that the motor program for speech is encapsulated so that the complexity of the program (at least within a stress foot) does not affect its initiation time in speeded productions. While the existence of these programs seems well-established by the experiments done to date, the details of the program are only partially evident. In particular, it is unclear whether the program consists of a sequence of fully-specified phoneme-sized commands, or of a series of gestures which can be executed with more simultaneity. If the former is the case, then the effects of coarticulation will appear to be the output of explicit contextualization rules. If the latter, then coarticulation could be seen as a direct consequence of the output of speech, where

activations for the various gestures produces the contextualization by means of an overlapping of their realizations.

The question of planning has taken on a slightly different guise in the phonetic literature, namely, whether the modifications of coarticulation are due to feature-spreading (as in Henke, 1966, and Keating, 1985, 1988) or by temporal concurrence, such as "shingling" of phones (Lieberman, Cooper, Shankweiler & Studdert-Kennedy, 1967, p. 441), overlap of prominence curves for phones (Fowler & Smith, 1986) or the phasing of gestures (in the articulatory phonology of Browman & Goldstein, 1985, 1986). The feature-spreading approach inherently postulates that coarticulatory effects will be equally large throughout the production of the affected phoneme (cf. Manuel 1987, pp. 18-19), although both Henke and Keating have introduced interpolating mechanisms which allow for an increase in the effects as the influencing phoneme is approached. In addition, Keating proposes that some phones will be unspecified for a feature, allowing for continuous change in its production. In general, however, features are claimed to spread as far back as they can, before reaching a contradictory specification. Conversely, the overlap models inherently predict severe limits on the duration over which coarticulation can occur. Fowler (1981, p. 38), for example, asserts that vowels should affect other stressed vowels only at the onset or offset (though certain components, such as lip rounding, might not be so limited). Measurements by Magen (1989), however, show vowel-to-vowel coarticulation throughout both stressed and weakly stressed vowels, even though the affecting vowel was separated from the affected one by a syllable containing the vowel schwa. Thus some modification of overlap models is already called for, and the present experiments will add to the evidence needed to make those modifications.

A special case of overlap is the frame model of Bell-Berti & Harris (1981, 1982). In this model, coarticulatory overlap is dependent on such properties of speech as stress and rate of articulation, but is independent of the number and composition of preceding segments. Thus lip-rounding activity will precede vowel onset by the same amount of time whether it is preceded by two or many segments, and whether these segments are intrinsically long or short. In their model, then, what is planned is not coarticulation *per se* but simply the onset of the following vowel, with the motor commands for the gesture affecting the signal more or less depending on the competing signals of the preceding gestures. Similarly, articulatory phonology postulates that the shape of gestures is insensitive to context, and that contextual variation is the result of the blending of the various levels of activation. In this sense, even though coarticulation can be seen in the signal, it has not been explicitly planned.

In all these theories, the input to the system is assumed to be complete, that is, all the phonemes of the utterance are known to the speaker. Yet if a speaker were to begin speaking before the whole utterance was known, the feature-spreading and overlap models make different predictions. By feature-spreading, if one phoneme is started before its context is known, then there should be no features to spread, thus coarticulatory influences should not appear. Even at the end of the production of an initial phoneme, there should be no influence, since the interpolation of the feature-spreading models is still based on features, and the initial segment does not have a feature value to interpolate from. In overlap models, if the phoneme is started before the context is known, the overlap should still appear since the

dynamic curves *are* the definition of the phonemes. Even in these theories, however, there is a mechanism by which the overlap would not appear: some articulatory gestures can be “hidden” (Browman & Goldstein, 1989), that is, the articulation can be present yet there is no acoustic evidence of it. In the best case, though, the two types of theories could be distinguished on the basis of productions which were begun on the basis of incomplete information.

The present experiments looked at coarticulatory effects in simple bisyllables (abí, abú, apí, apú) and trisyllables (əbíba, əbúba, əbípa, əbúpa) which were initiated before part of the utterance was known. This was accomplished by displaying on a computer screen part of the utterance (e.g., A\_U) and then filling in the missing part as soon as phonation began. In this way, various coarticulatory effects could be examined in conditions where they could be planned versus those where any effect would have to be made in mid-production. The conditions differed by the segment that was missing in the initial display. This could be either the consonant (B or P) or the final or middle vowel (I or U). Four coarticulatory effects were measured. The one associated with the vowel was the effect of shifting  $F_2$  slightly toward the  $F_2$  of a following vowel (Öhman, 1966). Three effects were associated with the voicing of the consonant. The offset frequency of a vowel's  $F_1$  before a voiced stop is lower than it is before a voiceless (Wolf, 1978; Lisker, 1986). Vowel duration is longer and closure duration shorter for a voiced stop relative to a voiceless one, with the two parts of the effect usually cancelling each other out so that overall duration is unaffected (Sharf, 1962). And, in the one effect which is perseverative for both the bi- and trisyllabic utterances, the  $F_0$  after release of a voiceless stop is high relative to a voiced stop (e.g. Hombert, Ohala & Ewan, 1979). In all these cases, it is possible to compare the two types of uncertainty, i.e., whether the vowel or the consonant was unknown at utterance onset, to see whether the coarticulatory effects emerge or not.

## 1. Experiment 1

The first experiment examined the four coarticulatory effects in nonsense bisyllabic utterances.

### 1.1. Method

#### 1.1.1. Subjects

The subjects were the author (DHW), a male colleague (LMG), and a female colleague (DMM). All were native speakers of American English.

#### 1.1.2. Stimuli

Each utterance consisted of a bisyllabic nonsense word. The first syllable was the low-back vowel /a/. The second syllable began with either /b/ or /p/ and ended with either /i/ or /u/. Stress was to be placed on the second syllable. Since the subjects were all familiar with the International Phonetic Alphabet, these were represented on the computer screen as ABI, ABU, API and APU.

#### 1.1.3. Equipment and procedure

Subjects produced the utterances in response to a visual display on a computer screen (Atari 800). In the first condition, only the two vowels appeared on the

screen initially, with a blank between them (e.g., "A\_U"). When this appeared, the subject could prepare for as long as he or she wished before initiating a response. Once the subject began phonating the /a/, a Schmidt trigger connected to the microphone output sent a signal to the computer. Once the signal was received, the consonant immediately appeared in the previously blank position, and the entire letter string remained on the screen for 1 s. The subject was instructed to incorporate that consonant into the ongoing production as rapidly and as smoothly as possible. In the second condition, the first two letters appeared on the screen (e.g., "AB\_"). Here again, the subject's task was to prepare to utter the bisyllable, and then begin. As soon as he or she did, the missing vowel appeared. That vowel was then incorporated as quickly and as naturally as possible into the utterance.

In each condition, 20 repetitions of each utterance type (e.g., /abi/) were randomized. If the subject made a mistake on a particular token, there was no repetition of that item. Instead, that token was excluded from the analysis. Errors were judged by the author and a colleague independently. Subject responses were recorded on audio tape on an Otari MX-5050 reel-to-reel recorder at  $7\frac{1}{2}$ /s. They were then low-pass filtered at 5 kHz and digitized through the Haskins PCM system at 10 kHz with 12 bit resolution for computer analysis.

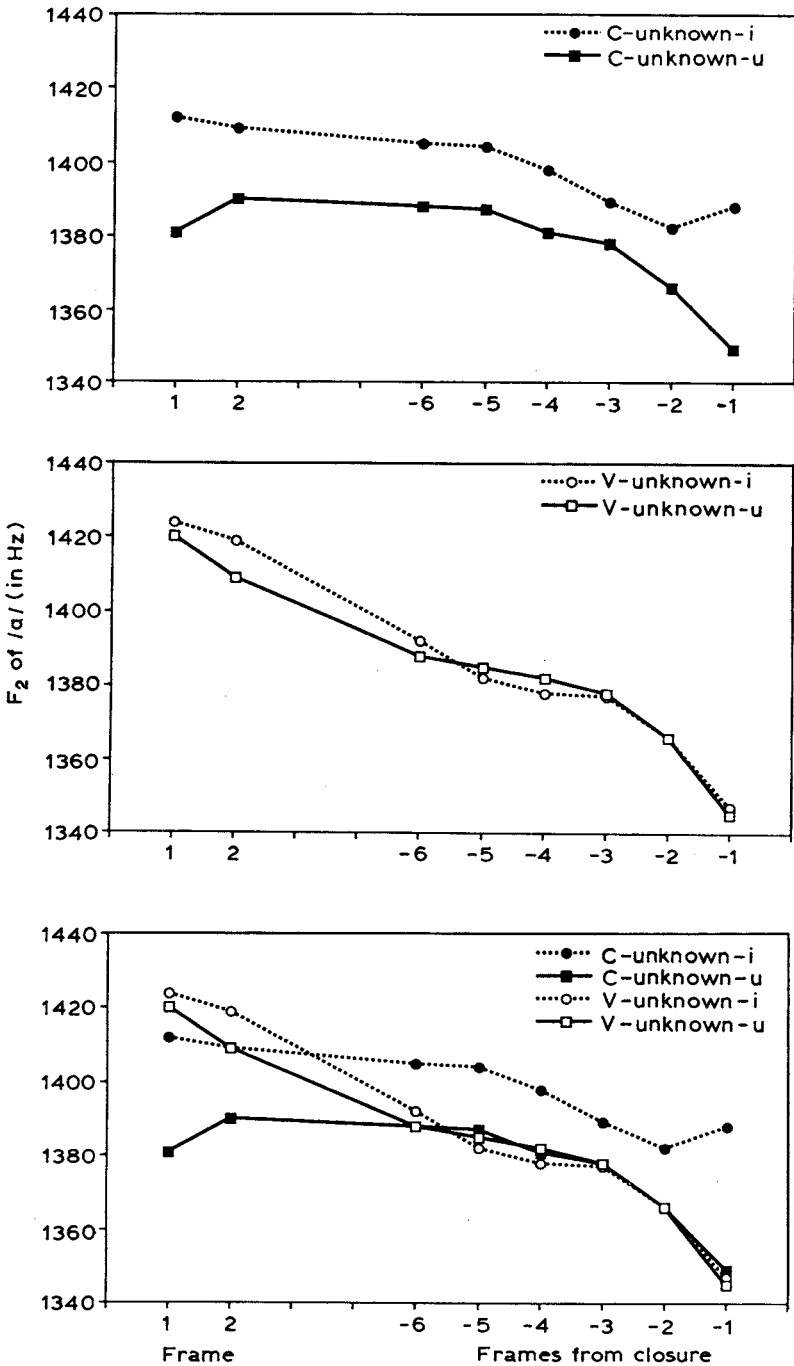
All duration measurements were made interactively with the wave-form editing program WENDY. All formant measurements were based on formants extracted from the LPC parameters generated by the ILS software package. These were further corrected by hand for those instances where the program had missed or added a formant. Missing values were excluded from the analysis.  $F_0$  measurements were made by delimiting individual pitch periods in the waveform, again with WENDY.

## 1.2. Results

Error rates (which included mispronunciations of either the vowel or the consonant) for the three subjects for the consonant-unknown and vowel-unknown conditions were: 3.75% and 8.75% (DHW); 15.0% and 5.0% (LMG); and 0% for both for DMM. In the consonant-unknown condition, DHW and DMM lost an additional 5.0% and 3.75% of the data due to technical malfunctions. Across the two subjects who made errors, about half (12 of 26) were on the item that was missing, while the others were on the consonant or variable vowel which had been visible from the beginning. Thus the procedure as a whole, not the nature of the unknown segment, seems to be responsible for the pattern of errors.

The results for the vowel-to-vowel effect on  $F_2$  are shown in Fig. 1. The first two data points for each function represent the first two LPC analysis frames (10 ms apart) at the beginning of the vowel. The next six points represent the frames just before the closure for the stop. In a few cases where the vowel was short, the same data appeared in the points before closure and the initial two.

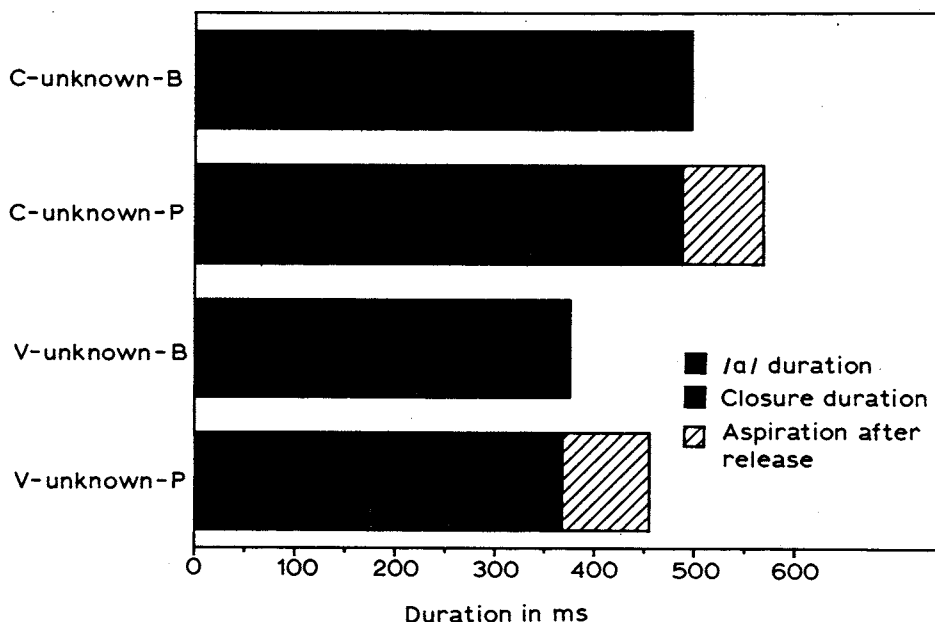
In the consonant-unknown conditions, the following vowel is known. And, as can be seen in the separation of the functions, that vowel exerts an influence on the  $F_2$  of /a/, in much the same way as originally found by Öhman (1966). By contrast, when the vowel is known, it affects neither the beginning nor the last vestiges of the /a/ before closure (i.e., the functions are hardly separable). Separate analyses of variance were performed on the beginning frames and the final frames in each of the



**Figure 1.** Effect of a following vowel on the F<sub>2</sub> of initial /a/, Experiment 1. In the consonant-unknown condition, the vowel is known at utterance initiation time. In the vowel-unknown condition, it is not. The first two points are the analysis frames (10 ms per frame) at the beginning of the vowel. The last points are the frames just before closure.

two conditions, with the factors of Speaker, Following Vowel, and Token (which allowed for missing values). The effect of Following Vowel was significant in the consonant-unknown condition for both the beginning frames ( $F(1, 205) = 8.12, p < 0.01$ ) and the final frames ( $F(1, 205) = 4.35, p < 0.05$ ). In the vowel-unknown condition, neither location showed a significant difference ( $F(1, 217) < 1.0$ , ns, for both). While speaker was a significant main effect in each analysis, indicating that the three subjects had different vowel spaces and/or vocal tract shapes, it entered into only one significant interaction. In the first frames of the consonant-unknown condition, the speaker by vowel interaction was significant ( $F(1, 205) = 5.45, p < 0.01$ ). While DHW and LMG showed the pattern evident in the first two points of Fig. 1, DMM had a non-significant difference in the opposite direction. This is at odds with the fact that she behaved like the other two subjects later in the vowel. Thus the effect of a following vowel on  $F_2$  of /a/ was clearly present in the consonant-unknown condition (for two subjects) and completely absent from the vowel-unknown condition.

The effect of consonant voicing on vowel duration was equally straightforward (Fig. 2). In the vowel-unknown condition, where the consonant was known, the usual lengthening of the vowel and shortening of the closure for the voiced stop relative to the voiceless was found. In the consonant-unknown condition, by contrast, there was no difference between the vowel duration (including the voiced transitions) before the two stops. In addition, there was no difference in the closure durations. This was in spite of the fact that the vowel duration was nearly twice as long as it was in the vowel-unknown condition. Separate analyses of variance were



**Figure 2.** Effect of the voicing of a following consonant on the duration of initial /a/ and the stop closure, Experiment 1. In the consonant-unknown condition (top two bars), the consonant is not known at utterance initiation time. In the vowel-unknown condition (lower two bars), it is.

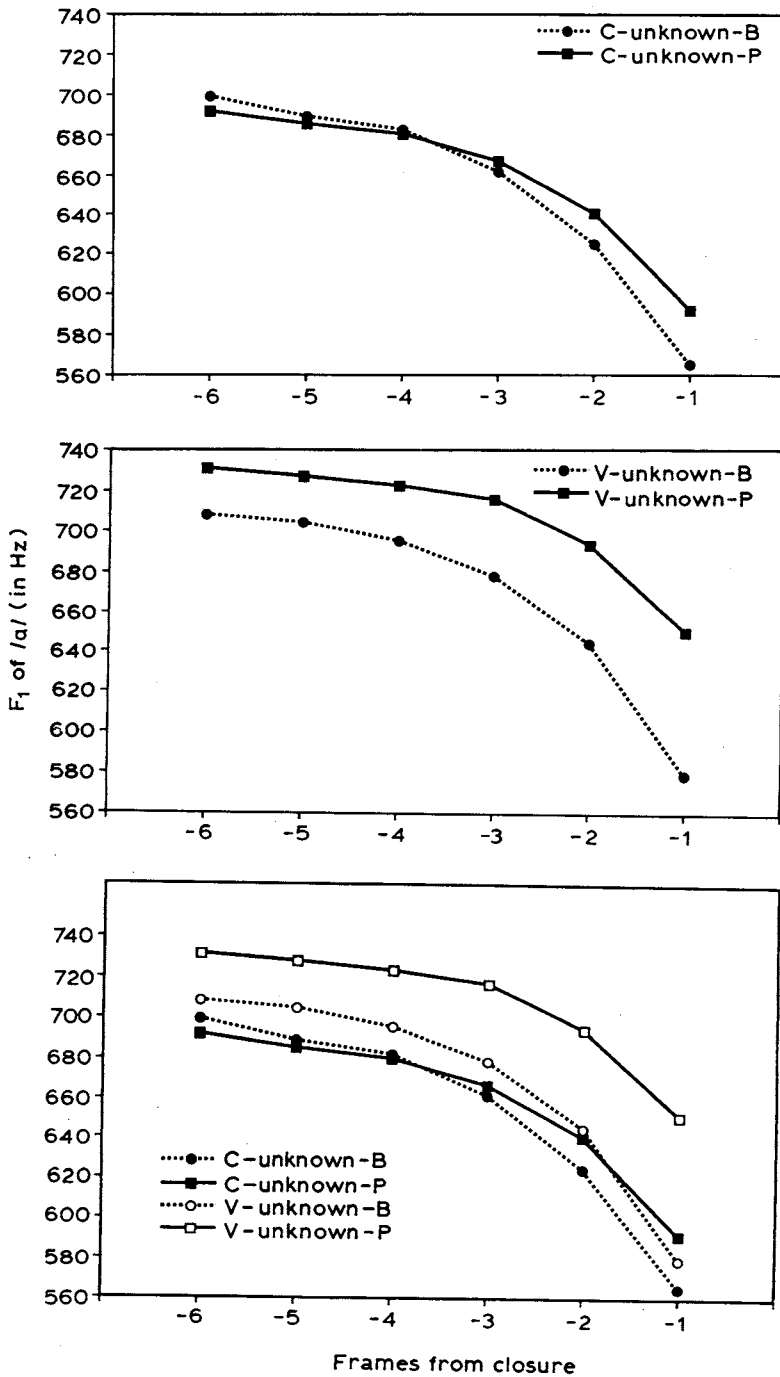
performed on vowel duration, closure duration, and total duration (vowel plus closure) for each of the two conditions, with the factors of Subject, Voicing and Token. Total duration did not differ with the voicing of the consonant in either condition ( $F(1, 205) < 1.0$  and  $F(1, 217) = 2.47$ , ns, for the consonant-unknown and vowel-unknown conditions respectively). Both vowel duration ( $F(1, 217) = 70.33$ ,  $p < 0.001$ ) and closure duration ( $F(1, 217) = 41.27$ ,  $p < 0.001$ ) differed in the vowel-unknown condition, but neither vowel duration nor closure duration differed in the consonant-unknown ( $F(1, 205) = 1.97$  and  $1.64$ , ns). Speaker was a highly significant factor, due in large part to the extreme durations in DMM's productions. Her closure durations were more like those of geminate stops than of typical American English stops. Nonetheless, none of the interactions of the voicing factor with speaker were significant, indicating that even DMM's extremely conservative production style still allowed the duration effect to emerge.

The effect of the stop's voicing on the  $F_1$  of the preceding vowel was strong when the consonant was known and virtually absent otherwise. In Fig. 3, the last six frames before closure for the stop show that the  $F_1$  of /a/ reaches a lower frequency before the voiced stop than before the voiceless when the consonant was known ( $F(1, 217) = 101.40$ ,  $p < 0.001$ ). (This effect does not appear to be due to the different durations of the vowels. Even if we shift the  $F_1$  from the voiced stop context by the difference in duration of the /a/, there is still a sizeable effect of voicing on  $F_1$ .) For the consonant-unknown condition in Fig. 3, there is no difference for all the points taken together ( $F(1, 205) = 1.30$ , ns). The apparent divergence in the last two frames before closure in the consonant-unknown condition, however, does reach significance ( $F(1, 205) = 4.17$ ,  $p < 0.05$  and  $10.86$ ,  $p < 0.01$ , for the next-to-last and last frames respectively). So, unlike the vowel effect on  $F_2$ , the voicing effect on  $F_1$  does show up at the very last instant.

First acoustic evidence (in the formant transitions) of the production of the missing segment appears about 300–320 ms from utterance onset in the consonant-unknown condition. The first solid evidence in the vowel-unknown condition appears at the release of the stop, around 360 ms, but the actual onset could have been up to 140 ms earlier, since the closures are on average, that long. It is possible that the time to resolve the identity of the missing segment was the same in both cases since both involved a choice between two phonemes.

In an earlier examination of consonant voicing on  $F_1$ , Summers (1987) found that  $F_1$  was affected throughout the syllable's duration. Analysis of the  $F_1$  in the first frames of the present stimuli showed no effect of consonant voicing in either condition.

The effect of the voicing of the stop on the following vowel's  $F_0$  was measured for each pitch period in the first 50 ms, and then one pitch period in the middle of the vowel. The "mid-vowel" pitch period was the one which included the point 100 ms after the onset of voicing for the voiceless stops; for the voiced stops, it was the pitch period which included the point 100 ms past the average VOT for the voiceless stops for that subject for that condition. This point was chosen to be somewhat later than Ohde's (1984) "target" pitch period, so that any effects would have to be somewhat stronger than the ones he found. In terms of absolute duration, this "mid-vowel" pitch period occurs somewhat earlier than the mid-point, but it was always after the most intense part of the vowel. For the statistical analysis, an equal number of pitch period values from each condition was selected, with a different



**Figure 3.** Effect of the voicing of a following consonant on the  $F_1$  of initial /a/, Experiment 1. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is.



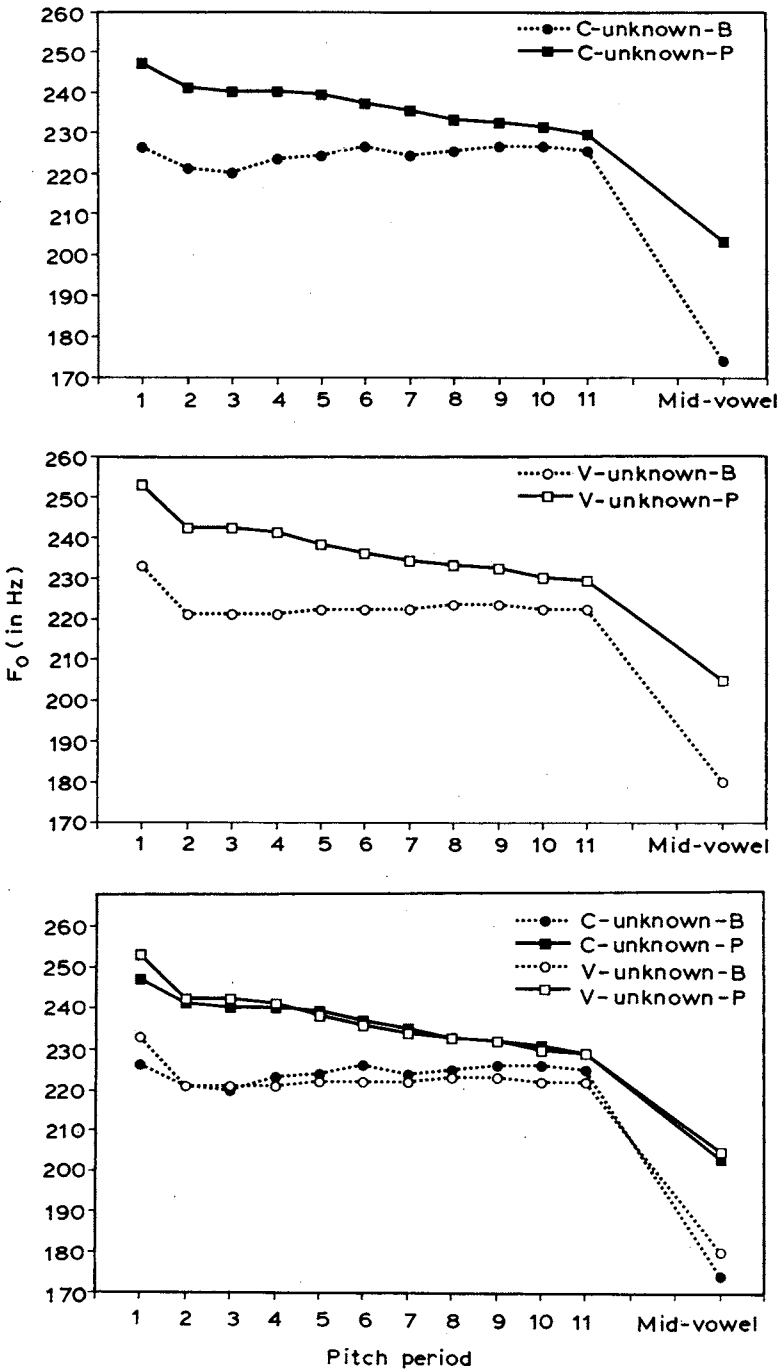
number for each speaker. The goal was to analyze the first few pitch periods in such a way that all the utterances would have an equal number of measurements, even though this meant leaving out some of the  $F_0$  pattern in the first 50 ms. The number of periods chosen obviously varied with the  $F_0$  itself, so that there were more data points for the female, fewer for the higher pitched male (DHW) and least for the lower pitched male.

As can be seen in Figs 4, 5 and 6, the voiceless consonant elicited a higher  $F_0$  than the voiced in both conditions. For DMM (Fig. 4), there was no main effect of condition either on the first pitch periods or on the mid-vowel ( $F(1, 149) < 1.0$  for each), while the effect of consonant was strong ( $F(1, 149) = 117.71$  and  $743.91$ ,  $p < 0.001$ , for early and mid-vowel respectively). There was an interaction of condition and consonant in the mid-vowel measurement; the effect was present in both, but slightly larger in the consonant-unknown condition. For LMG (Fig. 5), there was a main effect of condition for the early pitch periods ( $F(1, 134) = 84.94$ ,  $p < 0.001$ ), though it did not interact with consonant. The consonant also strongly affected the initial pitch periods ( $F(1, 134) = 166.32$ ,  $p < 0.001$ ). For the pitch period in the middle of the vowel, there was no effect of condition ( $F(1, 134) = 1.24$ , ns), but a strong effect for consonant ( $F(1, 134) = 61.85$ ,  $p < 0.001$ ) and an interaction between the two ( $F(1, 134) = 4.56$ ,  $p < 0.05$ ). Separate analyses of the two conditions revealed that the pitch period in the middle of the vowel differed with consonant in both conditions, but with greater magnitude in the consonant-unknown condition. For DHW (Fig. 6), there was an effect of condition and consonant for both the early and the midpoint pitch periods ( $F(1, 138) = 6.95$ ,  $p < 0.01$  and  $F(1, 138) = 372.96$ ,  $p < 0.001$  for the early,  $F(1, 138) = 12.22$ ,  $p < 0.001$  and  $F(1, 138) = 478.21$ ,  $p < 0.001$  for the midpoint). The interaction was significant only in the midvowel pitch period. Although the  $F_0$  was somewhat higher in the vowel-unknown condition, the early effect of the consonant was equally large in both conditions. In the middle of the vowel, the effect of the consonant's voicing was somewhat smaller in the consonant-unknown condition than in the vowel-unknown condition.

Thus, whether the consonant had been planned for in advance or not, the typical  $F_0$  pattern merged both at the beginning of the vowel and well after both the aspiration and release of the consonant. Some of the difference in the mid-vowel  $F_0$  might be due to the slight difference in duration of the final vowels. The syllables with voiceless stops tended to be about 35 ms longer than those with voiced stops. This means that the measurements for the voiced syllables were somewhat further along in a declining intonation pattern. However, it does not appear that the duration differences are large enough to account for the whole  $F_0$  difference. However, the difference in the size of the effects between conditions may be artifactual rather than real: the duration difference is larger in the vowel-unknown (45 ms) than the consonant-unknown (25 ms) condition. On the whole, then, this perseverative effect does not show a sensitivity to planning.

### 1.3. Discussion

For the three anticipatory effects of coarticulation, there is strong evidence that those effects which are not planned for do not appear, even when there would seem to be ample time for them to be incorporated into the on-going production. This is



**Figure 4.** Effect of voicing of a preceding stop on the  $F_0$  of final /i/ or /u/ for speaker DMM, Experiment 1. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is. Each pitch period was measured individually, so the higher  $F_0$ s are of shorter duration.

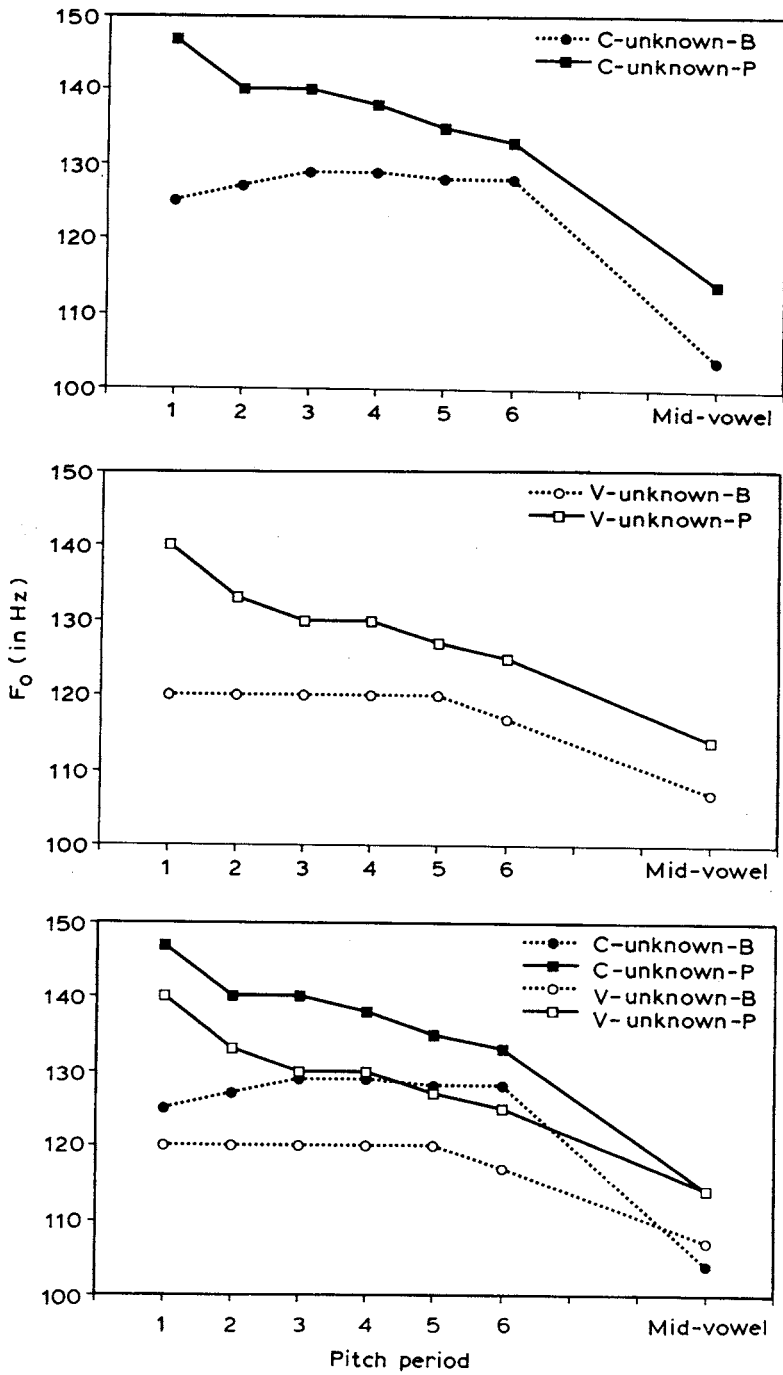
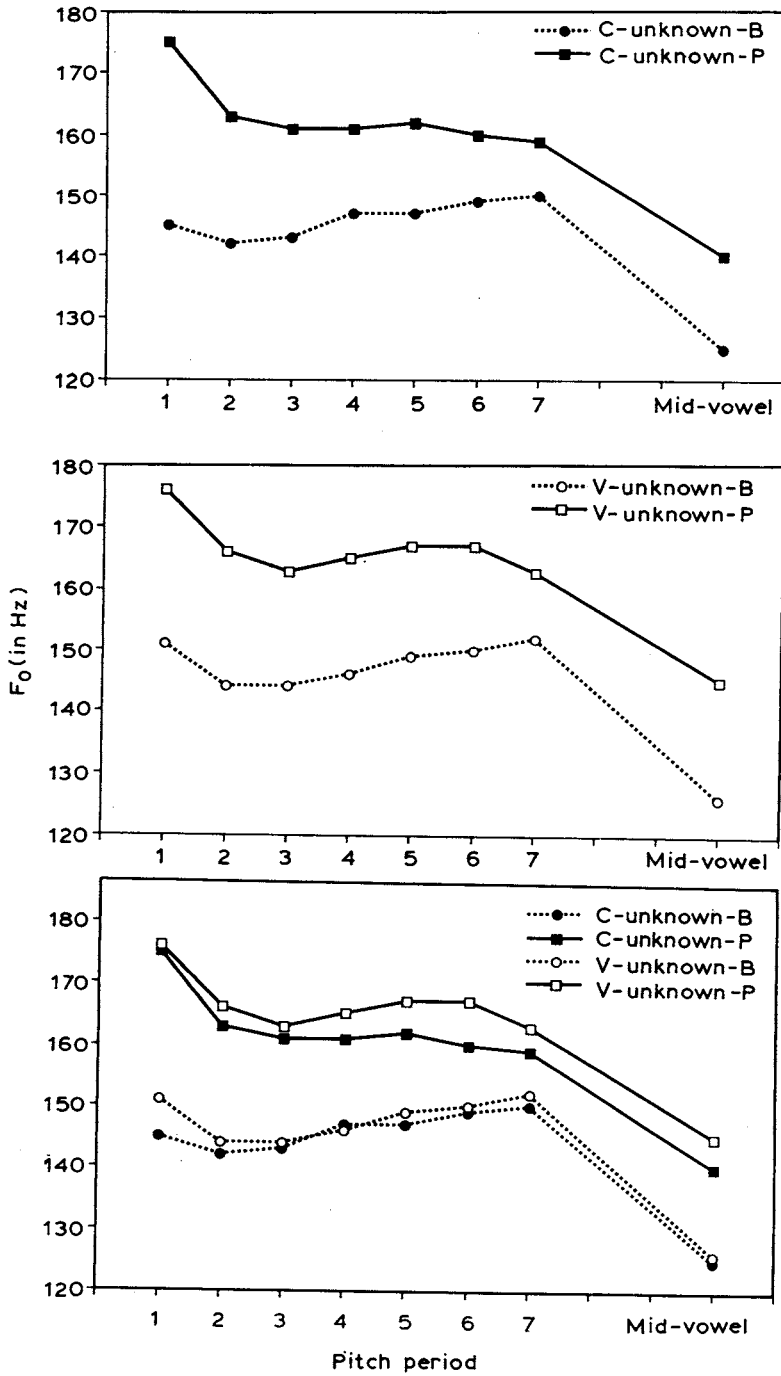


Figure 5. Effect of the voicing of a preceding stop on the  $F_0$  of final /i/ or /u/ for speaker LMG, experiment 1. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is. Each pitch period was measured individually, so the higher  $F_0$ s are of shorter duration.



**Figure 6.** Effect of the voicing of a preceding stop on the F<sub>0</sub> of final /i/ or /u/ for speaker DHW, Experiment 1. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is. Each pitch period was measured individually, so the higher F<sub>0</sub>s are of shorter duration.

true both of the frequency domain effects and the time domain one. The one perseverative coarticulatory effect did not show a similar tendency. Even though the task for the subjects was unnatural, those segments which were known at the time the utterances were initiated exerted their normal influences. This indicates that the productions were natural to the extent they could be.

The absolute duration for the time between the presentation of the missing letter and the onset of an acoustic effect was in the range defined by a variety of tasks from other researchers. These include a value of 270 ms for fast shadowing (Marslen-Wilson, 1973) and for initiating a learned sequence (Sternberg *et al.*, 1980), and 360 ms for naming one of two letters from print (Cattell, 1886, p. 527; Dick, 1971). It is appropriately longer than the 195 ms (Izdebski & Shipp, 1978) or 260 ms (Watson & Alfonso, 1983) speech initiation time for producing a known, isolated vowel. In another modality, initiation of a series of finger taps has been found to range from 220 to 280 ms, depending on the number of taps (Canic & Franks, 1989). Initiation of incompletely specified finger tapping sequences has been found to be closer to 380 ms (Semjen & Garcia-Colera, 1986, Experiment 3; Garcia-Colera & Semjen, 1988).

The absence of the effect of the following vowel in the  $F_2$  of the /a/ in the vowel-unknown condition has an expected component and an unexpected one. At the beginning of the /a/, the speakers would have had to be clairvoyant as well as coarticulatory, since the influencing segment had not even been presented. By the end, however, we might have expected some overlap to occur. The fact that none does occur seems to favor the feature spreading accounts outlined earlier. This conclusion is blunted by the relatively long closure durations obtained. All three subjects had closure durations in excess of the 100 ms window of coarticulation proposed by the frame model (Bell-Berti & Harris, 1981). Two of the subjects had essentially the same closure duration in both conditions (136 *vs.* 138 ms for LMG and 241 *vs.* 247 for DMM for the consonant-unknown *vs.* vowel-unknown conditions respectively). The third subject (DHW) had a longer closure (136 ms) for the vowel-unknown condition than for the consonant-unknown (102 ms). It is possible that the added duration was introduced just because the following vowel had not been properly planned for. The slow rise of the influence of the second vowel may in fact be a necessary aspect of vowel production, according to some overlap models. In that case, the early part of the vowel might be easier to incorporate into the utterance during the closure, where there would be no acoustic consequences of the vowel onset (cf. the "hidden" gestures of Browman & Goldstein, 1989). Whether this is the case or not, there is no obvious reason within a feature-spreading model for there to be a difference in the closure durations between the two conditions.

To summarize the  $F_2$  implications, the lack of an effect at the end of the /a/ is consistent with the prediction of feature-spreading models. But the closure durations are longer than normal in this experiment and particularly so for the vowel-unknown condition, leaving open the possibility that the essential part of the vowel's overlap was hidden by the closure itself. For one subject, the closure was longer in the vowel-unknown condition than in the consonant-unknown, possibly because it is easier to "hide" that part of the vowel than to incorporate coarticulation that had not been planned for.

The effect of the consonant voicing on the duration of the preceding vowel was

present in the vowel-unknown condition, but completely absent from the consonant-unknown condition. This seems odd when we consider how much longer the first vowel is in the consonant-unknown condition relative to the vowel-unknown condition. It would not be unreasonable to think that this increased duration would allow the speech apparatus time to introduce the durational differences. Yet, if that were the case, we would still have expected the vowel duration effect to show up in the second half of the vowel, and thus perhaps be of the same magnitude but reduced proportion compared with the vowel-unknown condition. Even if the duration effect decreased in magnitude, we still might have expected to see the nearly universal vowel duration effect, which though smaller, shows up in most languages (Chen, 1970). The lack of an effect on closure durations is also indicative of a need for this difference to be planned for. Given the differences in closing gesture velocity between voiced and voiceless stops (Summers, 1987), we might have expected that the closure duration difference was a necessary consequence of producing the two kinds of stops. Such is not the case. Here again, if the coarticulatory effects are not planned for, they do not show up.

The difference in the behavior of  $F_1$  between the two conditions shows a planning effect as well. In the vowel-unknown condition, where the consonant is known, there is a clear effect on the  $F_1$  offset due to the voicing of the stop. When the consonant is not known, this effect disappears almost entirely. Only in the last 20 ms before closure do we see an effect, while the other condition shows an effect at least 60 ms back. The conservative behavior of DMM, who had a period of voicelessness in each of her closures regardless of the voicing of the stop, might have contributed to this lack of a voicing effect. Yet the effect was just the same for the other two subjects, who often maintained voicing throughout the closure for voiced stops. Thus the effect on offset frequency is not due directly to the successful cessation of voicing (cf. Fant, Ishizaka, Lindqvist & Sundberg, 1972). Nor is it a by-product of the duration difference between the two stops, since there was no duration difference in the consonant-unknown condition. Rather, there are differences in the shape of the closing trajectory which are present when they are anticipated and all but absent when they must be incorporated mid-stream.

The voicing of the consonant did not affect  $F_1$  throughout the vowel as it had in Summers (1987). Summers's stimuli were quite different, being CVCs in sentential context, which made for shorter durations. So the lack of an effect at the beginning of the vowels in the present experiment might be due simply to the relatively long durations of those vowels. Another possibility is that Summers'  $F_1$  effect might, in fact, reflect coarticulation between the two consonants, thus affecting the vowel only indirectly. To decide this question, we would need to examine stimuli which are like Summers's in having more typical durations than the ones used here, but which are vowel-initial or perhaps /h/ initial.

For the most part, the perseverative coarticulatory effect (the  $F_0$  pattern due to the voicing of the stop) does not show an effect of initial planning. All three subjects have a higher  $F_0$  for the voiceless stop than for the voiced at the onset of voicing after release, and at a point well beyond the release of the stop. This might indicate that the planning that such an effect requires can be done in the time it takes to complete the first vowel and the stop closure, or it might indicate that both the early and late effects of voicing on the vowel's  $F_0$  are an automatic consequence of involving the cricothyroid muscle in the devoicing gesture (as argued by Löfqvist,

Baer, McGarr & Story, 1989). Since the differences in the present case extend well beyond the approximate upper limit on relaxation time of the cricothyroid (about 120 ms; Löfqvist *et al.*, 1989, p. 1320), some further explanation is called for. Whether that further explanation is another automatic one (such as laryngeal height) or one that calls for planning remains to be seen.

While we have evidence that subjects can complete the partially-specified utterance they have launched, we do not see the restitution of anticipatory coarticulation effects in that utterance. The perseverative effect, in contrast, shows very little difference between the two conditions. This outcome is compatible with feature-spreading accounts of coarticulation, though there are considerations that make articulatory-overlap accounts acceptable as well. The feature-spreading accounts predict that even interpolation will be absent, since there is no feature to interpolate to. However, such accounts have no explanation for the increased closure duration made by DHW in the vowel-unknown condition. The overlap accounts might attribute this added duration to an act of "hiding" the gradual increase of the influence of the second vowel. This would allow the formant structure of the first vowel to proceed as planned, with the second vowel's onset being accomplished in the silent closure. More disturbing for the overlap accounts is the virtual lack of an  $F_1$  effect for the consonant-unknown condition. It would seem that success in making a stop would include the modification of this parameter, due most likely to a difference in the speed of the gesture. This difference is almost absent, however, indicating that even overlap accounts must allow either for planning of coarticulation or for different but linguistically equivalent gestures for the stops.

## 2. Experiment 2

The second experiment was designed to examine two perseverative effects rather than just one, and to put the location of the missing phoneme somewhat farther along in the utterance. If recovery of planning can occur within the span of a vowel and a stop closure, adding an extra syllable should restore almost all of the coarticulatory effects. To this end, the test stimuli contained an initial syllable consisting only of schwa, followed by /biba/, /bipa/, /buba/, or /bupa/. The missing segment was either the first full vowel (/i/ or /u/) or the second consonant (/b/ or /p/). Stress was to be placed on the first full vowel. The influence of the vowel on the  $F_2$  of adjacent vowels could be examined both in the schwa and the /a/. With the /a/, the effect is no longer an anticipatory one, as in the first experiment, but a perseverative one.  $F_2$  vowel effects in the schwa should follow the pattern found for /a/ in the first experiment.

### 2.1. Method

#### 2.1.1. Subjects

The subjects were the two male subjects of the first experiment and a different female colleague (CLS). The conservative strategy adopted by the earlier female subject seemed likely to further submerge the differences in the planning conditions, so she was replaced. All were native speakers of American English.

### 2.1.2. Stimuli

Each utterance consisted of a trisyllabic nonsense word. The first syllable was the mid-central unstressed vowel /ə/. The second syllable began with /b/, with either /i/ or /u/ as the vowel. The third syllable began with either /b/ or /p/ and ended with the low-back vowel /ɑ/. These were represented on the computer screen as UHBIBA, UHBIPA, UHBUBA, and UHBUPA. Stress was to be placed on the middle syllable.

### 2.1.3. Equipment and procedure

Subjects produced the utterances in response to a visual display on a computer screen. In the first condition, the position for the /i/ or /u/ vowel appeared initially as a blank (e.g., "UHB\_PA"). When this display appeared, the subject could take as long as he or she wished to prepare before initiating a response. Once the subject began phonating the /ə/, the missing letter (I or U) immediately appeared in the previously blank position. The subject was instructed to incorporate that vowel into the ongoing production as rapidly and as smoothly as possible. In the second condition, the position for the consonant appeared initially as a blank (e.g., "UHBU\_A"), and the rest proceeded as in the other condition.

Twenty repetitions of each utterance type were presented in random order. Errors were again excluded from the analysis. The utterances were input into the computer as in the previous experiment and measured as before, with duration and  $F_0$  measured by hand from the waveform and  $F_1$  and  $F_2$  taken from the LPC analysis.

## 2.2. Results

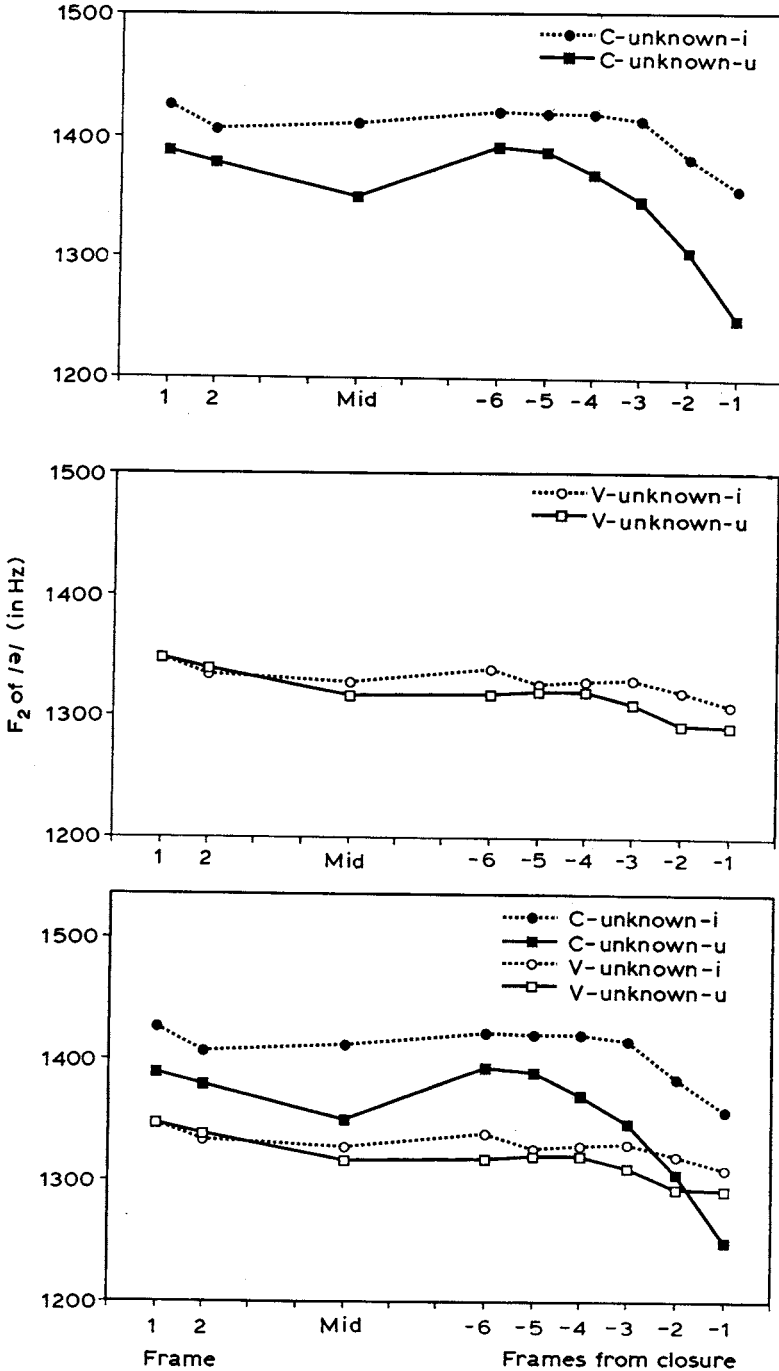
The same coarticulatory effects which were examined in the first experiment were looked at here. The main differences were that the  $F_2$  effect of the /i/ and /u/ vowels could be seen both in anticipatory position on the schwa and in perseverative position on the /ɑ/. The duration and  $F_1$  effects associated with the voicing of the medial /b/ or /p/ now appeared on the /i/ and /u/ vowels rather than the /ɑ/.

Error rates (which included mispronunciations of either the vowel or the consonant) for the three subjects for the consonant-unknown and vowel-unknown conditions were: 11.3% and 17.5% (CLS); 7.5% and 20.0% (LMG); and 7.5% for both for DHW. All errors in the consonant-unknown condition were consonant mistakes, while vowel mistakes accounted for only 42% of the errors in the vowel-unknown condition. This pattern differs somewhat from that in the first experiment, where neither condition had an effect on the type of error.

Acoustic consequences of the missing phonemes for the consonant-unknown condition occurred at about the same point as they did in the first experiment, at around 320 ms. For the vowel-unknown condition, evidence appeared earlier, at around 230 ms.

Figure 7 shows the effect of the middle vowel, /i/, or /u/, on the  $F_2$  of the initial schwa, measured at the beginning, the midpoint, and just before closure.  $F_2$  of schwa is affected throughout its duration when the vowel is known at initiation time, but is affected only at the end when the vowel is not known. The effect of vowel identity on  $F_2$  was strong ( $F(1, 398) = 10.37, 48.01, 60.05, p < 0.01$  for the beginning, middle and end respectively), though it did differ by condition ( $F(1, 398) = 12.48, 20.97, 13.60, p < 0.001$  for the interaction of condition and





**Figure 7.** Effect of an upcoming vowel on the  $F_2$  of initiation /ə/, Experiment 2. In the consonant-unknown condition, the vowel is known at utterance initiation time. In the vowel-unknown condition, it is not. The first two points are the analysis frames at the beginning of the vowel. The next point is the frame at the midpoint of the /ə/ as defined by its duration. The last points are the frames just before closure.

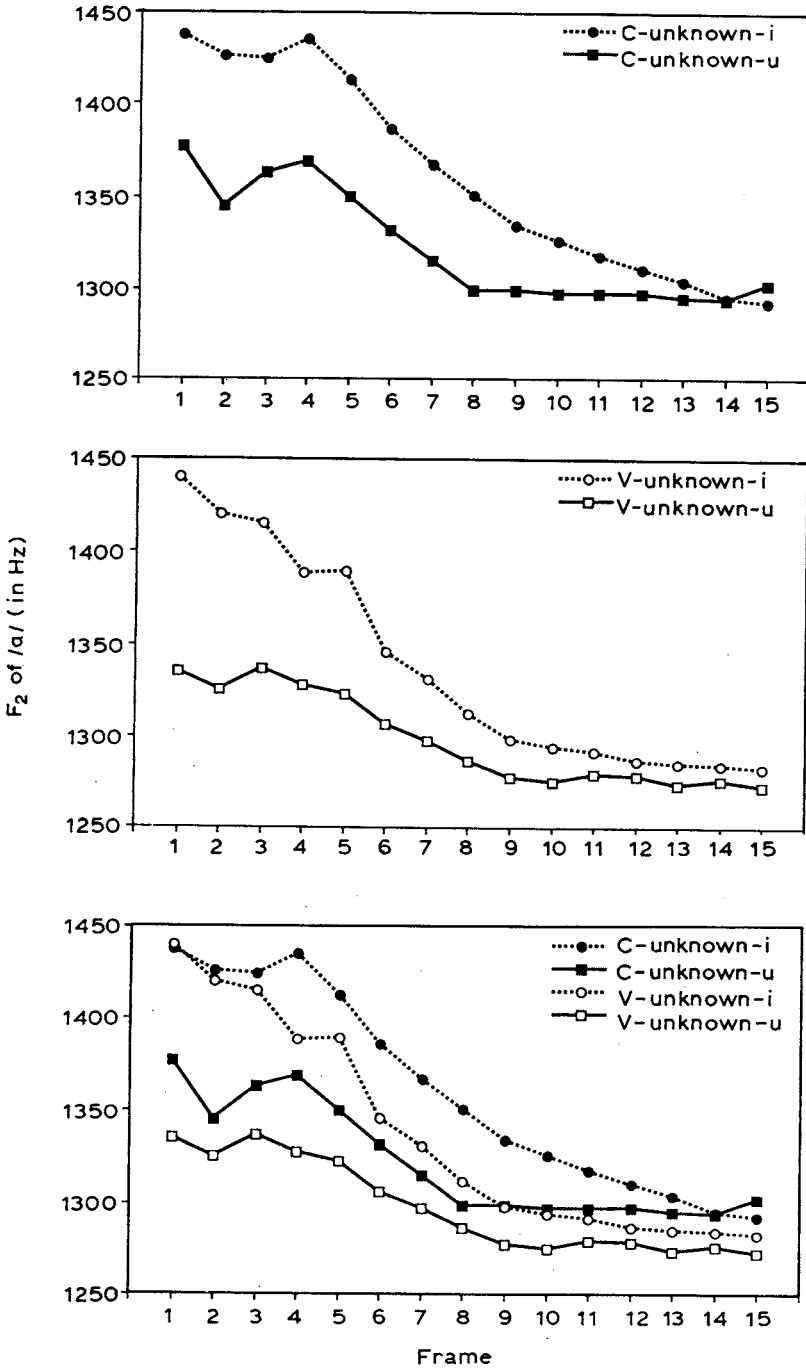
vowel for beginning, middle and end respectively). In the vowel-unknown condition, the beginning ( $F(1, 192) < 1.0$ ) and midpoint ( $F(1, 192) = 2.65$ , ns) were unaffected by the following vowel, while the end was ( $F(1, 192) = 4.93$ ,  $p < 0.05$ ). Thus, in contrast to the first experiment, the final portion of the vowel was affected even in the vowel-unknown condition, though to a far smaller degree than in the consonant-unknown condition. Note also that the amount of vowel-to-vowel coarticulation increases (from 29 Hz to 197 Hz) as closure is approached in the consonant-unknown condition, but varies unsystematically within a small range (7–28 Hz) for the vowel-unknown condition.

It is of interest that the closure for the /b/ between the schwa and the middle vowel was not abnormally long for one of the subjects (DHW) in this condition, although his closures had been unusually long in the first experiment in this condition. In fact, for DHW, the closures had essentially the same duration in both conditions. For the other two subjects, this /b/ closure was longer in the vowel-unknown condition than in the consonant-unknown. Given that one explanation of the lack of an effect of the following vowel on the  $F_2$  in Experiment 1 was that the closure was effectively "hiding" the vowel gesture, a more direct test of this possibility here seemed in order. The correlation between closure duration and the size of the effect on  $F_2$  of the following vowel was measured for each subject for each condition. To obtain the size of the effect, the mean for the other vowel was subtracted from each  $F_2$  value. More specifically, for the vowels in the /i/ context, the mean value for each frame of the schwa transition in /əbuba/ was subtracted from the corresponding frame of the /əbiba/ transitions. Similarly, the mean for /əbupa/ was subtracted from the value of the /əbipa/ tokens. For /u/, the effect was in the other direction, so the values of frames of the /əbuba/ transitions were themselves subtracted from the mean value of the /əbiba/ frames, and the /əbupa/ from the /əbipa/. In this way, effects in the predicted direction were always positive, while negative numbers reflected the opposite of the predictions. Average values for the last three frames before closure were then correlated with the closure duration of the first stop, /b/. (For both the  $F_2$  differences and the durations, values outside two standard deviations from the mean were excluded.) A negative correlation would show that the shorter the closure is, the greater the effect of the following vowel.

As shown in the correlation values of Table 1, two of the three subjects showed a significant negative correlation in the last three analysis frames of the schwa in the vowel-unknown condition. No significant correlations occurred in the consonant-

TABLE I. Correlations between the size of the effect of a following  $F_2$  (averaged over the last three analysis frames of the schwa vowel) with the closure duration for the stop

Subject	Condition	Pearson's $r$	$N$	$p <$	Closure	SD
CLS	V unknown	-0.285	61	0.05	140.9	47.2
	C unknown	0.074	65	ns	69.3	11.5
LMG	V unknown	-0.336	59	0.01	89.5	25.5
	C unknown	-0.023	72	ns	82.9	11.1
DHW	V unknown	-0.190	67	(0.15)	61.8	14.3
	C unknown	0.103	70	ns	64.6	8.1

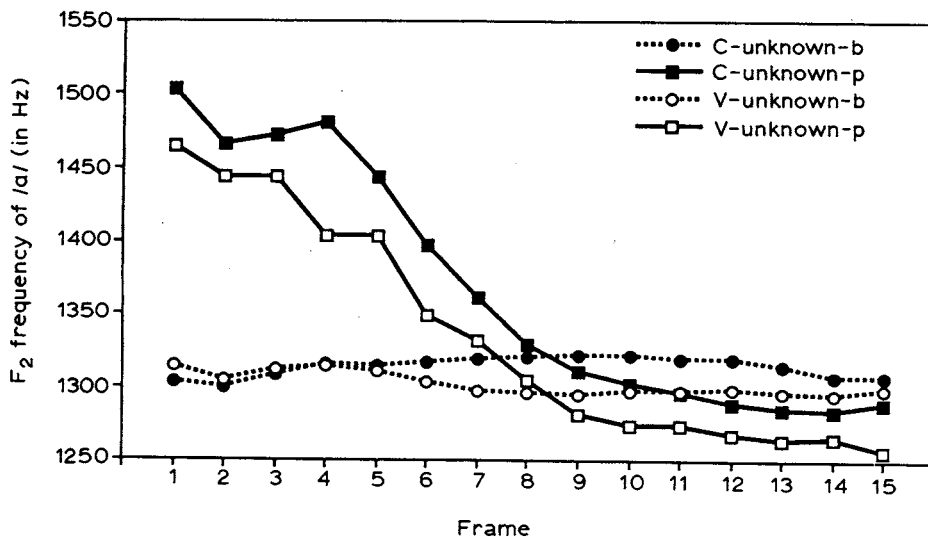


**Figure 8.** Effect of a preceding vowel on the F<sub>2</sub> of final /a/, Experiment 2. In the consonant-unknown condition, the vowel is known at utterance initiation time. In the vowel-unknown condition, it is not. Each frame represents 10 ms of speech.

unknown condition, but this may be due to the small standard deviations of the closures (see column 5 of Table 1), leaving little room for a correlation to appear. The subject who had a correlation that just missed being significant (DHW) had the only significant vowel-to-vowel effect in individual analyses. (These individual analyses were performed in conjunction with the correlation analysis, and were done in spite of the fact that there was no interaction between subject and vowel effect in the main analysis.) Here again, the standard deviation of the closure durations was small. Thus it appears that when the vowel had not been planned for, it can affect the end of the preceding vowel as long as it is not "hidden" by the stop closure.

As can be seen in Fig. 8, the effect of the preceding vowel on the  $F_2$  of /a/ exists well into the vowel, and equally in the two conditions. An analysis of variance for the first 15 analysis frames with the grouping factors of speaker, condition, consonant, and vowel was performed. A large number of frames was used so that at least half of the comparisons for /p/ would involve voiced formants. (The mean VOT for /p/ was 61 ms, so the first six frames would contain no voiced formants, and aspiration shifts the frequency of the formants; see Repp and Lin, 1987.) The effect of the preceding vowel was quite strong ( $F(1, 397) = 98.07, p < 0.001$ ). There was no interaction of vowel and condition, with or without other factors. Thus the effect of preceding vowel on  $F_2$  was present whether that vowel had been known at initiation time or not, and to an equivalent degree in both conditions.

One more effect of  $F_2$  which has been postulated to be coarticulatory by some (e.g. Fant, 1973), namely, the higher  $F_2$  for aspirated stops compared with unaspirated, can be examined here as well. For both conditions,  $F_2$  was about 200 Hz higher for /p/ than for /b/ at the consonant release (Fig. 9), an effect which matches the results of Repp and Lin (1987) almost exactly. This supports their conclusion that the effect is an acoustic one, dependent only on the effects of



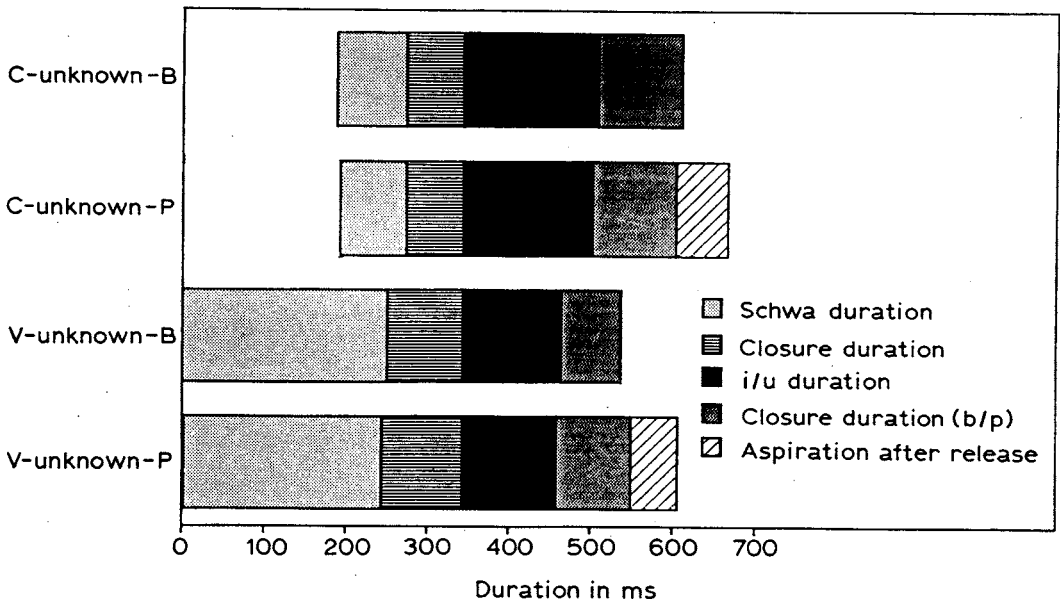
**Figure 9.** Effect of the voicing of a preceding consonant on the  $F_2$  of final /a/, Experiment 2. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is. Each frame represents 10 ms of speech.

coupling due to the open glottis. Unlike their results, however, there was not a complete convergence of the  $F_2$  once the vowel steady-state was reached. Instead, the steady-state  $F_2$  after /p/ was actually about 35 Hz lower than that after /b/. Perhaps this is an over-compensation of some sort, but it does not vary with condition, so it will not be further explored.

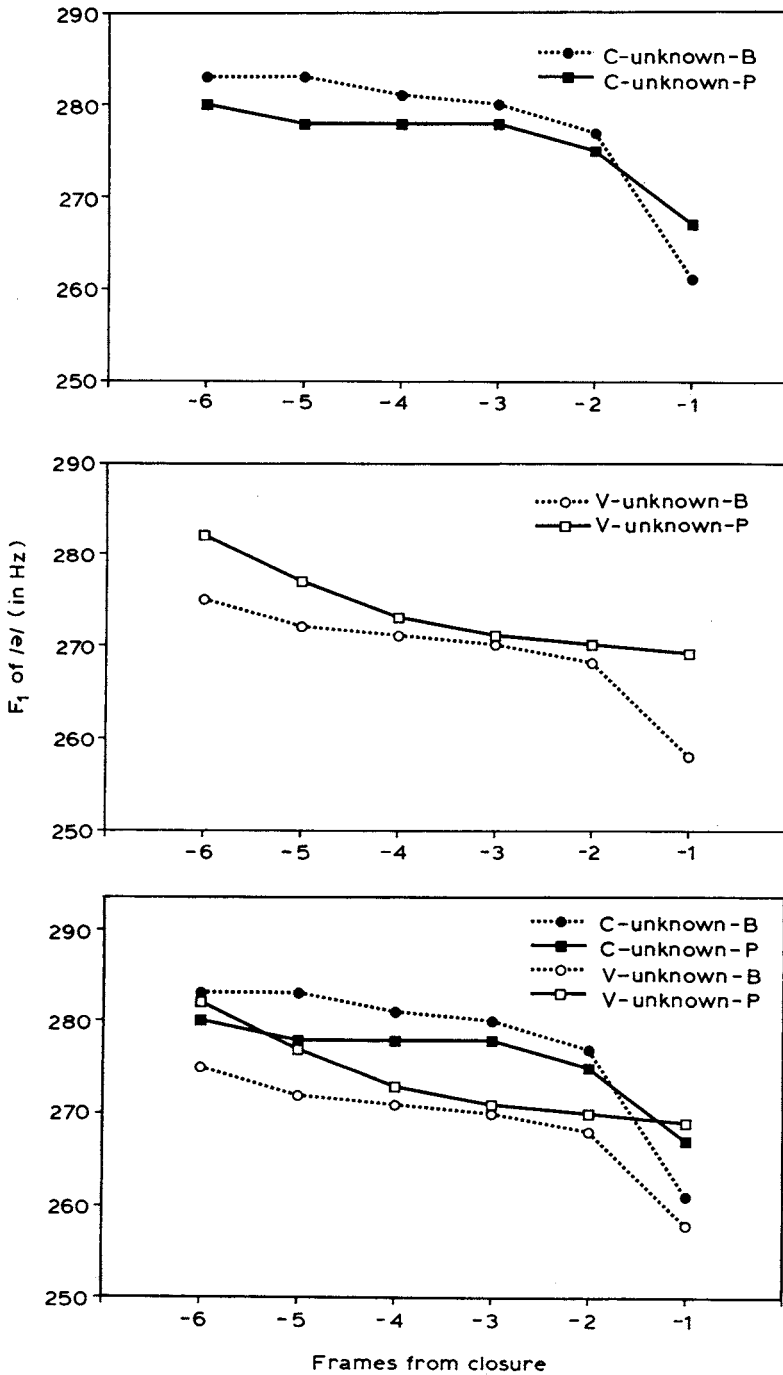
The effect of the consonant voicing on vowel and closure duration can be seen in Fig. 10. The bars in this figure have been aligned at the onset of the middle vowel, thus showing the differences in the duration of the first syllable yet allowing the voicing effect to be seen more easily. In the consonant-unknown condition, voicing affects neither the vowel duration (164 ms before /b/, 159 ms before /p/,  $F(1, 205) < 1.0$ , ns) nor the closure duration (103 ms for /b/, 102 ms for /p/,  $F(1, 205) < 1.0$ , ns). When the consonant is known (vowel-unknown condition), by contrast, both the vowel duration (120 ms before /b/, 114 ms before /p/,  $F(1, 192) = 8.41$ ,  $p < 0.01$ ) and the closure duration (74 ms for /b/, 91 ms for /p/,  $F(1, 192) = 35.79$ ,  $p < 0.001$ ) are influenced.

The  $F_1$  values for the offset of the /i/ or /u/ vowel are displayed in Fig. 11. They show an extremely small trend toward a lowering due to voicing of the stop. Statistical analysis shows that this is not significant either as a main effect of voicing ( $F(1, 398) < 1.0$ , ns) or as an interaction with condition ( $F(1, 398) < 1.0$ , ns). For the final frame by itself, there is a marginal effect ( $F(1, 398) = 3.49$ ,  $p < 0.10$ ). In short,  $F_1$  was unaffected by the voicing of the stop whether it was the consonant which was unknown or the vowel.

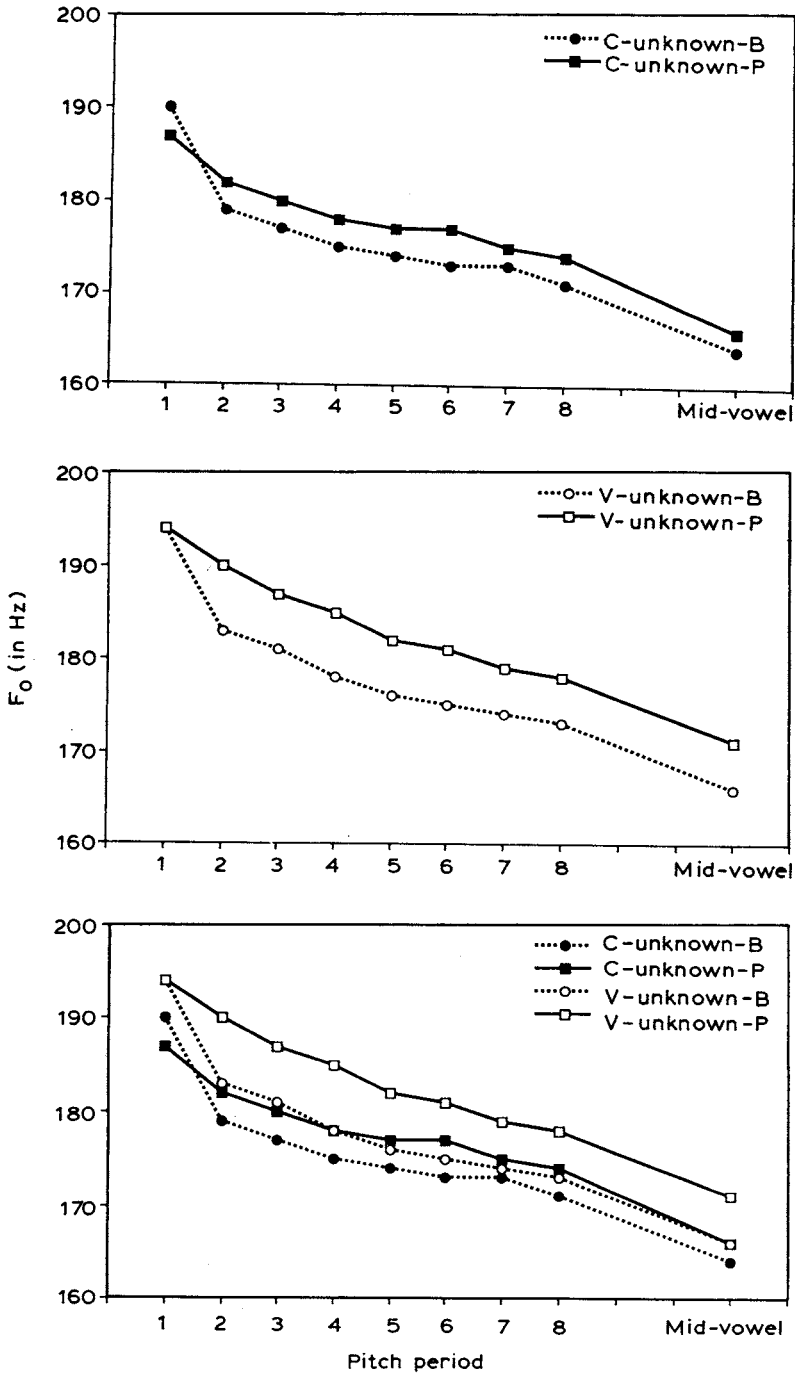
Figures 12 to 14 show the effect of voicing on the  $F_0$  after release of the consonant, again for each subject separately, to overcome the difference in the



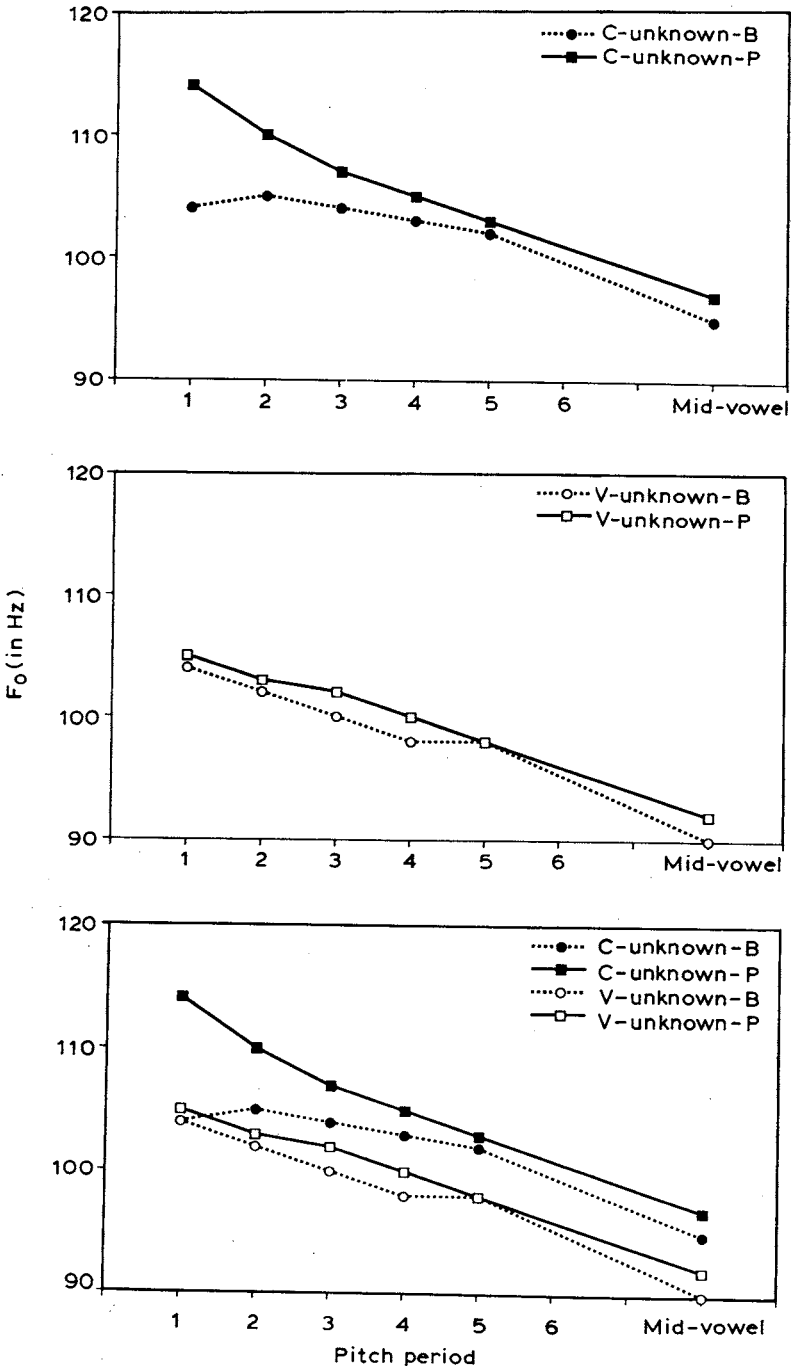
**Figure 10.** Effect of the voicing of an upcoming consonant on the duration of initial /a/ and the stop closure, Experiment 2. In the consonant-unknown condition (top two bars), the consonant is not known at utterance initiation time. In the vowel-unknown condition (lower two bars), it is.



**Figure 11.** Effect of the voicing of an upcoming consonant on the  $F_1$  of medial /i/ or /u/, Experiment 2. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is.

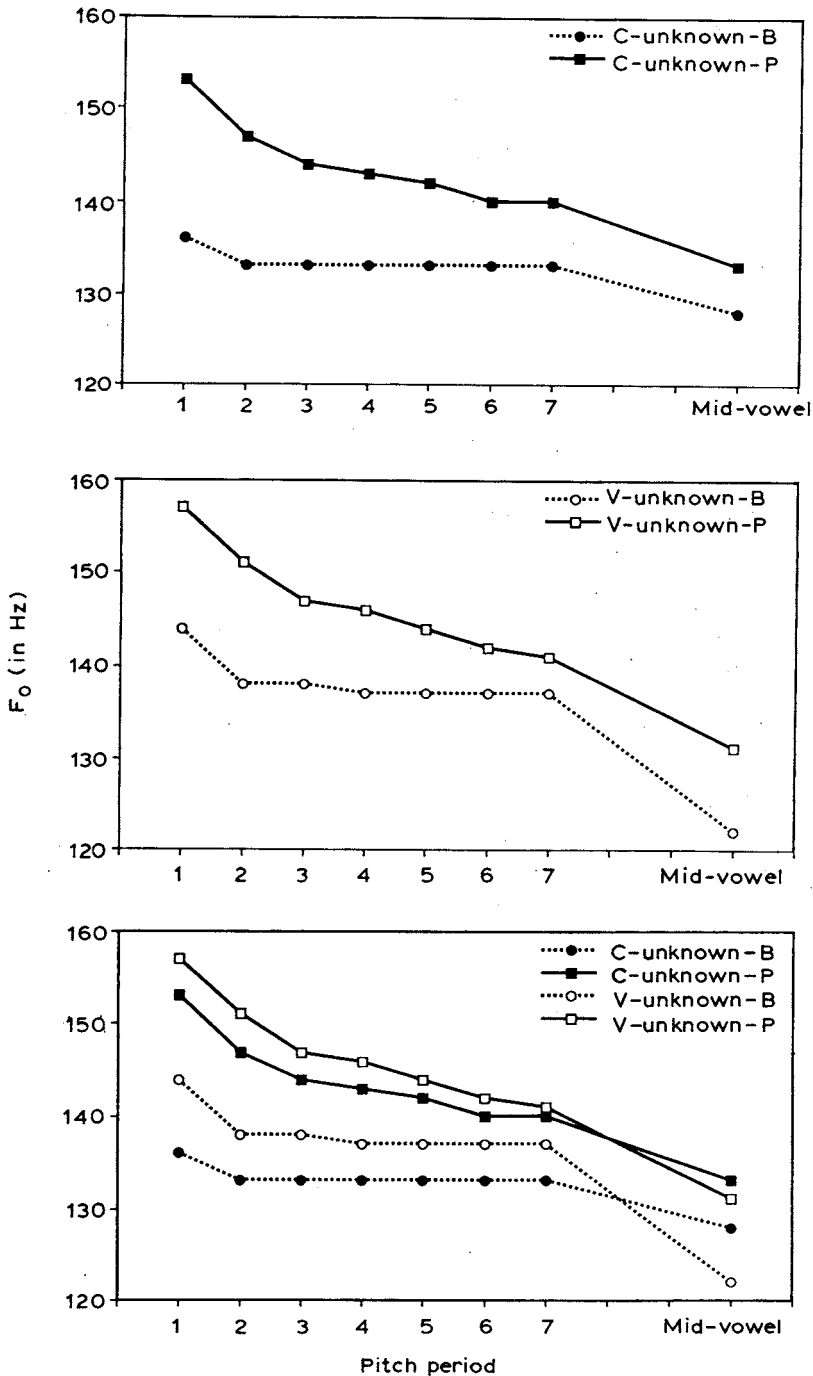


**Figure 12.** Effect of the voicing of a preceding stop on the  $F_0$  of final /a/ for speaker CLS, Experiment 2. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is. Each pitch period was measured individually, so the higher  $F_0$ s are of shorter duration.



**Figure 13.** Effect of the voicing of a preceding stop on the  $F_0$  of final /a/ for speaker LMG, Experiment 2. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is. Each pitch period was measured individually, so the higher  $F_0$ s are of shorter duration.





**Figure 14.** Effect of the voicing of a preceding stop on the  $F_0$  of final /a/ for speaker DHW, Experiment 2. In the consonant-unknown condition, the consonant is not known at utterance initiation time. In the vowel-unknown condition, it is. Each pitch period was measured individually, so the higher  $F_0$ s are of shorter duration.

number of data points. For speaker CLS (Fig. 12), there are main effects of condition and voicing ( $F(1, 128) = 5.85$  and  $6.21$ ,  $p < 0.05$  respectively) with no interaction ( $F(1, 128) = 1.78$ , ns). For speaker LMG (Fig. 13), there is an effect of voicing for both the initial portion in the consonant-unknown condition ( $F(1, 69) = 24.36$ ,  $p < 0.001$ ), and in the middle ( $F(1, 69) = 15.26$ ,  $p < 0.001$ ). There was no effect at the beginning in the vowel-unknown condition ( $F(1, 57) = 1.04$ , ns) though there was in the middle ( $F(1, 57) = 7.51$ ,  $p < 0.01$ ). Though there is a difference between conditions for this subject, it is not a difference that could be predicted by the nature of the missing phoneme, since the consonant effect on  $F_0$  failed to appear in the condition where the vowel was unknown. For speaker DHW (Fig. 14), the  $F_0$  results were similar to those of Experiment 1. The pitch periods immediately after release differed in the expected direction for both the consonant-unknown condition ( $F(1, 70) = 44.93$ ,  $p < 0.001$ ) and the vowel-unknown condition ( $F(1, 70) = 137.64$ ,  $p < 0.001$ ). The same held true of the mid-vowel pitch periods ( $F(1, 170) = 4.91$ ,  $p < 0.05$ , for the consonant-unknown condition, and  $F(1, 70) = 57.09$ ,  $p < 0.001$ , for the vowel-unknown condition).

### 2.3. Discussion

Of the three anticipatory effects of coarticulation measured here, two (the  $F_2$  effect of the following vowel on schwa and the duration difference due to voicing of the stop) showed the predicted difference between the conditions, while the third (the effect of voicing on  $F_1$ ) failed to show up in either condition. This latter failure seems to be due to the intrinsically low  $F_1$  of the vowels involved. All three of the perseverative effects (the effect of the preceding vowel on the  $F_2$  of /a/, and the effects of voicing on the  $F_2$  and  $F_0$  of /a/) showed no difference between the conditions. The conclusions from the first experiment receive further support, that coarticulation is largely planned, and that some of this planning can occur during the course of an utterance. The effect of the *i/u* vowel on the  $F_2$  of the preceding schwa does show some partial restoration under favorable conditions in the vowel-unknown condition, providing support for an overlap model.

The vowel-to-vowel coarticulation showed up on initial schwa in much the same way that it appeared on the /a/ in the first experiment, with one interesting difference: just before closure, the vowel effect began to appear even in the vowel-unknown condition. Further, there was a correlation for two of the subjects between the duration of the closure separating the schwa from the /i/ or /u/ and the size of the coarticulatory effect. When the closure was small, the effect was large, showing that articulatory overlap was having acoustic consequences even though the vowel had not been planned for before utterance onset. Such evidence is consistent with the notion that the articulatory onset of vowels precedes the typical acoustic landmarks by a considerable amount (Fowler, 1979, 1980). The largest stretch of this vowel, however, shows no coarticulation.

The perseverative effects of the middle vowel on the final /a/ were present in both conditions, and equally strongly in both. This shows either that perseverative coarticulation is basically mechanical (cf. Recasens, 1984, 1987), or that planning can be restored in the course of the utterance. Differences in magnitude of the effect do not help us decide the question in this instance, since two different vowels were involved. The anticipatory effect was much larger, but schwa is more amenable to

this kind of coarticulation than /a/ is. There is no way within this paradigm to sort out these two possible explanations for the incorporation of perseverative effects. However, the presence of  $F_0$  coarticulation for stop voicing late in the vowel argues that at least some perseverative effects are due to planning rather than to mechanical considerations.

The effects of stop voicing on vowel and closure duration mirrored those of the first experiment. If the consonant was not known at initiation time, the duration effects did not appear. This is in spite of the fact that the syllable which would have exhibited the duration effects was now the second syllable rather than the first one. In this instance, perhaps the duration of the first syllable was not long enough to allow the coarticulatory planning. Or, perhaps its structure was inappropriate. (That is, if planning occurs at the level of the stress foot and if the unstressed syllable was structurally part of the same stress foot as the following stressed syllable, the lack of recovery may have been due to the unavailability of the structure rather than to a lack of time.)

The lack of the effect of voicing on  $F_1$  is probably artifactual. High vowels (in this case, /i/ and /u/) have relatively low frequency  $F_1$ s. Other researchers have found that this obscured the coarticulatory effect (Hillenbrand, Ingrisano, Smith & Flege, 1984). Thus this portion of the experiment seems to reflect a lack of an appropriate measure rather than a lack of coarticulation.

The behavior of  $F_0$  after the release of the stops indicates that the subjects were introducing a typical devoicing gesture (Löfqvist *et al.*, 1989). Though smaller in magnitude than in the first experiment, the effect was consistently present. (The difference in magnitude probably reflects the shifting of primary stress from the syllable containing the stop (Experiment 1) to the preceding syllable (Experiment 2).) The effect in the midpoint of the vowel, however, does not seem likely to be due to the cricothyroid activity for the voicelessness *per se*, since the duration of the effect is so long. Some other automatic effect (such as laryngeal height) may be responsible. If not, then we may have evidence that planning has been recovered over the course of the utterance, rather than an indication of a difference between anticipatory and perseverative coarticulation.

### 3. General discussion and conclusion

Coarticulation, though presumed to be due to the constraints of producing speech in real time, is largely a result of planning an utterance rather than an automatic consequence of successfully producing that utterance. This was shown by forcing subjects to begin uttering a nonsense string before the entirety of the string was known. The unknown element was either a vowel or a consonant, and, for the most part, anticipatory coarticulatory effects associated with the unknown segment did not appear even though effects associated with the known segment did. Subjects were, therefore, successfully producing a normal utterance to the extent possible, and inserting a new phoneme as the experiment required. Perseverative effects, however, were mostly present in both conditions. This indicates either that perseverative effects are indeed different from anticipatory ones in being mechanical rather than planned, or that planning for these effects can be accomplished after utterance onset. The relatively long duration of the  $F_0$  effect due to voicing argues for a planning explanation. In the second experiment, the appearance of vowel-to-

vowel coarticulation in just those cases where a stop closure was short enough for articulatory overlap to have an acoustic effect indicates that some coarticulation is indeed automatic. If coarticulation consisted only of such effects, however, its description would be relatively simple. The effects that cover a much larger stretch of time, the ones which need to be planned, are the ones that make coarticulation such a challenging problem.

Under the best of circumstances, these results would allow us to test the differing predictions of feature-spreading and articulatory overlap models. Unfortunately, there is little agreement in phonology on which features are specified for particular values and thus which features might spread. In addition, feature-spreading accounts have been modified in recent publications, while overlap models in general have not. For example, Keating (1985) proposes interpolation over unspecified segments as an adjunct to feature spreading. Thus the best we can do is to elaborate some of the implications for each class of theories of the results for the present study. The descriptions of feature-spreading accounts, though, will usually be of modified versions. As we will see, there will remain difficulties for the feature-spreading accounts, and necessary amendments for the overlap accounts.

If vowel-to-vowel coarticulation is based on phonological features, then the degree of such coarticulation should be predictable from the features used by a language. With a complex vowel system using many features, we could expect that there would be less coarticulation than with a simple vowel system which leaves many features unspecified. In fact, that is what we find. When a language has a small vowel inventory, it seems to allow greater coarticulation, as shown by Magen (1984), Manuel & Krakow (1984) and, Manuel (1987). These authors propose explanations which do not depend on features, however. Manuel (1987), for example argues that speakers are sensitive to the acoustic differentiation of the vowels of their language, so that small inventories allow greater latitude for subphonemic variations. That is, she predicts that vowel-to-vowel coarticulation will be more globally related to a vowel's articulation rather than to the phonological features, but with output constraints that force it into a particular part of the vowel space.

The aspect of the current study which lends most support to the feature-spreading accounts is that some of the coarticulation was evenly distributed throughout the vowel affected. For example, the effect of the following vowel on the  $F_2$  of /a/ in Experiment 1 (Fig. 1) is of the same magnitude at the beginning as it is up until the last frame before closure. If this coarticulation is due to the spread of a phonological feature, such consistency is to be expected, although the "mechanical" effects at the end of the vowel can also be expected. Certainly some dynamic overlap models will have to be modified to account for such effects, but the types of interpolation used in feature-spreading models must also be constrained if they are to allow for relatively steady-state effects. However, the type and size of coarticulation must also be predicted by the features used if the feature-spreading models are to be fully supported. The effects reported here seem too small to be due to the implementation of a feature such as [+round]. In the example here, there is some evidence that the effect in the consonant-unknown condition is due to the /i/ context, since the /u/ context and the lack of an effect are similar (see Fig. 1). If so, then the choice of the feature to account for the coarticulation is difficult. It is surely not height, which would show up on  $F_1$ , not on  $F_2$ , and which is already contrarily specified for /a/. The effects on  $F_2$  will be due to front/back and/or rounding. Yet /a/ cannot change

its specification for "back" without becoming a different vowel, and it is already "-round" (distinctively so for subjects that distinguish /a/ and /ɔ/). The remaining features available in a feature-based vowel system, labial and ATR, do not seem to do the trick. Even if they do, the size of the effect obtained (in this case, around 25 Hz) seems much too small to be attributed to the change of a feature. It may indeed be important, as Keating (1988) proposes, to allow feature matrices to exit the phonology with some features unspecified, but relying on the inventory of phonological features does not seem to give enough detail to explain coarticulation. The same can be said of any feature-spreading account which does not have more (or at least more continuous) phonetic features than phonological ones.

Theories of coarticulatory overlap have different difficulties to handle. The most obvious, both from the present work and that of Magen (1989), is that overlap must extend beyond the stressed vowel (indeed, beyond the stress foot in Magen's data). This overlap must be extensive in time but not in magnitude. Since the size of the effect does not vary through the majority of the vowel, extending the onset of the tail into the vowel by a simple interpolation algorithm will not serve here any better than it does with feature-spreading accounts. An extension of articulatory phonology (Browman & Goldstein, 1985, 1986) might allow for the modification of one vowel's input parameters as well as articulatory overlap with the same influencing vowel. This two-component approach (i.e., both gestural modification and gestural overlap are allowed) fits the present results (especially those of Experiment 2) quite well, but it remains to be seen whether the increase in the power of the model is warranted. An alternative can be put in terms of prominence curves (Fowler and Smith, 1986), which do not correspond directly to physiological measurements or to gestures (since prominence is associated with phonemes, not gestures) but are rather theoretical constructs which are intended to express, non-mathematically, the structure of utterances. The range of phenomena addressed by prominence curves would be enlarged by the postulation of extended "tails", which would allow the overlap of vowels at a reduced level for long periods. While it may be possible to make articulatory phonology and the prominence curve account generate the same results, two aspects of coarticulation suggest that the "tail" approach might be more natural. First, having the tail be a part of the prominence curve makes it natural to assume that the characteristics of the tail depend on the characteristics of the main prominence. With parameter modification, such a relationship can be specified but need not be. Second, if the tail is indeed connected to the main prominence, then the connection should be maintained for the tail to exist. If, for example, a consonantal gesture that was incompatible with the vowel's tail preceded the vowel, it would be reasonable to think that the entire tail would be suppressed. This would begin to account for the apparent lack of coarticulation across consonants in Russian (Keating, 1988).

One result from each experiment supports the gestural overlap approach. In Experiment 1, one subject (DHW) had longer closure durations in the vowel-unknown condition, perhaps in response to the greater ease of hiding the onset of the vowel gesture in the silent stop closure than superimposing it on the end of the already launched preceding vowel gesture. In Experiment 2, the closure durations of the first stop were short enough that correlations could be made between the size of the vowel-to-vowel coarticulation and the closure duration. In the vowel-unknown conditions, there was indeed a negative correlation, indicating that when the closure

was short, the coarticulatory effect was big, just as the overlap models predict. Since the overlap models were originally designed to show that the vowel gesture necessarily precedes the point usually measured as "vowel onset" (Fowler, 1979), these results are promising. Even though modifications need to be made, it indeed seems that the a failure to produce the "tails" (or equivalent) would still allow the phoneme to be recognized correctly, while the effects near the stop closure are unavoidable because they represent the true onset of vowel activity.

The results for the perseverative coarticulatory effects are somewhat ambiguous. Since they are, for the most part, equally strong in both conditions, it is either the case that they are successfully planned for during the preceding vowel and stop closure, or that they do in fact represent physical constraints as claimed by such researchers as Gay (1977) and Recasens (1987). The  $F_0$  difference late in the vowel is the one most difficult to attribute to mechanical effects. The mechanism most often cited to explain the  $F_0$  difference, the residual effect of the devoicing activity on the laryngeal control of  $F_0$ , does not seem to extend long enough to account for the effects found here and elsewhere (Ohde, 1984; Löfqvist *et al.*, 1989). While other more automatic components may yet be found, it currently seems more likely that even in the perseverative direction, planning is at work. Even the effects attributable to more sluggish articulators may call for a planning explanation. The stimuli used by Magen (1989) and by Recasens (in prep.) (which included a stressless schwa between two stressed vowels), for example, have more distance (in time and segments) between the influencing segment and the influenced than could be encompassed by mechanical effects. Still, they find vowel-to-vowel effects. To that extent, we can lean toward the planning explanation. That is, while there are acknowledged differences in the magnitude and behavior of anticipatory and perseverative effects, we have not exhausted the range of alternative explanations. One promising explanation is that a given amount of coarticulation might affect preceding and following contexts differently because the contexts are themselves inherently different. Recent unpublished work by Fowler & Vatikiotis-Bateson showing that amplitude of the speech signal and indeed the speech gestures regularly declines over the course of an utterance (much as  $F_0$  does) sets up an interesting possibility: if the strength of a vowel's coarticulation is dependent on the vowel's own strength, then vowels will have a larger effect on following segments than on preceding ones. That is, the strength of the coarticulation would be the same, but the strength of the affected phoneme would be less, resulting in greater evidence of coarticulation. There are, of course, other differences to be explained, but this example shows that the explanations need not depend on different mechanisms for anticipatory and perseverative coarticulation.

The lack of the duration effects of consonant voicing in the consonant-unknown condition may not be surprising given that the effect in English is thought to be phonologized. We should not expect phonological effects to appear by virtue of articulation *per se*. However, the difference in speed of closure, as seen in the  $F_1$  transition differences, is almost absent as well. Both the normal duration and  $F_1$  effects seem to be "enhancing" the voicing distinction, as Stevens & Keyser (1989) might say. Yet here again, the two effects do not seem to be easily attributed to specific, binary values of phonological features, as these authors require. As an alternative, the structures and relationships proposed by the task-dynamic approach to speech production (e.g. Saltzman, 1986; Saltzman & Kelso, 1987) may account

for the success of producing the voicing distinction even in the absence of many of its normal characteristics. In that approach, the definition of a linguistically significant gesture is not the variety of articulatory movements which go into the utterance, but rather the goal which a large family of movements can attain. Thus if some portion of an articulator's trajectory is disturbed, other articulatory maneuvers can make up the difference. Defining speech as a set of coordinative structures, then, seems to allow the continuous variation and responsiveness to a changing environment needed for describing such facts as those presented here (cf. Saltzman & Munhall, 1989).

The magnitude and extent of coarticulatory effects seem to exceed the requirements of the speech articulators for accommodation of the competing demands made by successive phonemes. Acceptable, distinctive utterances can be accomplished without many of the coarticulatory effects found consistently in speech. It seems, then, that coarticulation is geared toward making speaking easier and/or enhancing the effectiveness of the speech stream for the listener. When phonemes influence one another, the cohesion of the sound is presumably increased. We have evidence in reaction times that this information is taken into account even when it does not affect overt phonemic decisions (Martin & Bunnell, 1982; Whalen, 1984). Since these reaction time studies found effects even with single tokens of utterances, it should be the case that even the vowel-to-vowel effects examined here should be usable, despite the large number of repetitions required to measure the effect. Fowler & Smith (1986), in fact, found that vowel-to-vowel coarticulation does affect perception, though they examined the effects on schwa. Acoustic measurements seem less sensitive than a human listener, so a listener's ability to detect the presence of this kind of coarticulation should not be ruled out simply because we cannot measure it successfully on single tokens. The stretch of the utterance over which coarticulation operates is of great interest, since it tells us how much of an utterance should normally cohere. Vowels with primary stress do not seem to be the primary source of coarticulatory spread (Magen, 1989), despite their initial appeal as the locus for cohesion. It is possible that lexical stress (even on the neologistic utterances used here) is governed by the constraints that apply to the segments, and that more discourse-oriented features such as focus are responsible for differences in coarticulatory span. Similarly, we should not rule out the possibility that individual differences in amount and organization of coarticulation will be detectable, perhaps contributing to the ease or difficulty some people have in making themselves understood. Knowing the bounds of such coarticulation would also further inform the improvement of speech synthesis by rule.

Without being explicit about the mechanisms of planning, it is still possible to find evidence for planning or its lack. In the present experiments, there is ample evidence of a great diminution of anticipatory coarticulation when a talker begins an utterance before knowing the influencing segment. Since normal coarticulatory effects were consistently found for other segments in those same conditions, we can be confident that the utterances were not simply unnatural throughout. Rather, talkers seem to plan for as much coarticulation as they can. If perseverative coarticulation is planned (as there is some evidence it is), then this planning can be accomplished before a syllable starts, in the course of a preceding vowel and stop closure. This is further evidence that talkers structure their utterances to include coarticulation, presumably because it serves a communicative function.

This research was supported by NSF Grant BNS-8111470 and NICHD Grant HD-01994 to Haskins Laboratories. Part of this research was presented at the 28th annual meeting of the Psychonomic Society, Seattle, WA, November, 1987. Invaluable assistance in data collection and analysis was provided by Diana Matson, Mark Pitt, and Yi Xu. Comments on earlier drafts by Caroline Smith, Carol A. Fowler, Vincent L. Gracco, Sharon Y. Manuel, Daniel Recasens, Elliot L. Saltzman, Anders Löfqvist, Alvin M. Liberman, Catherine P. Browman, Louis M. Goldstein, Margaret Hall Dunn, Rena A. Krakow, Suzanne Boyce, Kevin G. Munhall, Patrice Speeter Beddor, Mary Beckman and an anonymous reviewer were most helpful.

### References

- Bell-Berti, F. & Harris, K. S. (1981) A temporal model of speech production, *Phonetica*, **38**, 9–20.
- Bell-Berti, F. & Harris, K. S. (1982) Temporal patterns of coarticulation: lip rounding, *Journal of the Acoustical Society of America*, **71**, 449–454.
- Browman, C. P. & Goldstein, L. (1985) Dynamic modeling of phonetic structure. In *Phonetic linguistics* (V. A. Fromkin, editor), pp. 35–53. New York: Academic Press.
- Browman, C. P. & Goldstein, L. (1986) Towards an articulatory phonology, *Phonology Yearbook*, **3**, 219–252.
- Browman, C. P. & Goldstein, L. (1989) 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.
- Canic, M. J. & Franks, I. M. (1989) Response preparation and latency in patterns of tapping movements, *Human Movement Science*, **8**, 123–139.
- Cattell, J. M. (1886) The time taken up by cerebral operations, *Mind*, **11**, 220–242, 377–392, 524–538.
- Chen, M. (1970) Vowel length variation as a function of the voicing of the consonant environment, *Phonetica*, **22**, 129–159.
- Dick, A. O. (1971) Processing time for naming and categorization of letters and numbers, *Perception & Psychophysics*, **9**, 350–352.
- Fant, G. (1973) Stops in CV-syllables. In *Speech sounds and features* (G. Fant, editor), pp. 110–139. Cambridge, MA: MIT Press.
- Fant, G., Ishizaka, K., Lindqvist, J. & Sundberg, J. (1972) Subglottal formants, *Speech Transmission Laboratory Quarterly Progress and Status Report (Stockholm: Speech transmissions Laboratory, KTH)*, **1**, 1–12.
- Fowler, C. A. (1979) “Perceptual centers” in speech production and perception, *Perception & Psychophysics*, **25**, 375–388.
- Fowler, C. A. (1980) Coarticulation and theories of extrinsic timing control, *Journal of Phonetics*, **8**, 113–133.
- Fowler, C. A. (1981) A relationship between coarticulation and compensatory shortening, *Phonetica*, **38**, 35–50.
- Fowler, C. A. (1984) Segmentation of coarticulated speech in perception, *Perception & Psychophysics*, **36**, 359–368.
- Fowler, C. A. & Smith, M. (1986) Speech perception as “vector analysis”: An approach to the problems of segmentation and invariance. In *Invariance and variability of speech processes* (J. S. Perkell & D. H. Klatt, editors), pp. 123–139. Hillsdale, NJ: Laurence Erlbaum Associates.
- Garcia-Colera, A. & Semjen, A. (1988) Distributed planning of movement sequences, *Journal of Motor Behavior*, **30**, 341–367.
- Gay, T. J. (1977) Articulatory movements in VCV sequences, *Journal of the Acoustical Society of America*, **62**, 183–193.
- Henke, W. L. (1966) Dynamic articulatory model of speech production using computer simulation. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Hillenbrand, J., Ingrisano, D. R., Smith, B. L. & Flege, J. E. (1984) Perception of the voiced-voiceless contrast in syllable-final stops, *Journal of the Acoustical Society of America*, **76**, 18–26.
- Hombert, J. M., Ohala, J. & Ewan, W. (1979) Phonetic explanation for the development of tones, *Language*, **55**, 37–58.
- Izdebski, K. & Shipp, T. (1978) Minimal reaction times for phonatory initiation, *Journal of Speech and Hearing Research*, **21**, 638–651.
- Keating, P. A. (1985) CV Phonology, experimental phonetics, and coarticulation, *UCLA Working Papers in Phonetics*, **62**, 1–13.
- Keating, P. A. (1988) Underspecification in phonetics, *Phonology*, **5**, 275–292.
- Kent, R. & Minifie, F. (1977) Coarticulation in recent speech production models. *Journal of Phonetics*, **5**, 115–133.



- Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967) Perception of the speech code, *Psychological Review*, **74**, 431-461.
- Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised, *Cognition*, **21**, 1-36.
- Liberman, A. M. & Studdert-Kennedy, M. (1978) Phonetic perception. In *Handbook of sensory physiology*, Vol. VIII, *Perception* (R. Held, H. Leibowitz, & H. L. Teuber, editors), pp. 143-178. New York: Springer-Verlag.
- Lisker, L. (1986) "Voicing" in English: a catalogue of acoustic features signalling /b/ versus /p/ in trochees, *Language and Speech*, **29**, 3-11.
- Löfqvist, A., Baer, T., McGarr, N. & Story, R. S. (1989) The cricothyroid muscle in voicing control, *Journal of the Acoustical Society of America*, **85**, 1314-1321.
- Magen, H. (1984) Vowel-to-vowel coarticulation in English and Japanese, *Journal of the Acoustical Society of America*, **75**, S41.
- Magen, H. (1989) An acoustic study of vowel-to-vowel coarticulation in English. Unpublished Ph.D. dissertation, Yale University.
- Manuel, S. Y. (1987) Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages. Unpublished Ph.D. dissertation, Yale University.
- Manuel, S. Y. & Krakow, R. A. (1984) Universal and language particular aspects of vowel-to-vowel coarticulation, *Haskins Laboratories Status Report on Speech Research*, **SR77/78**, 69-78.
- Marslen-Wilson, W. D. (1973) Linguistic structure and speech shadowing at very short latencies, *Nature*, **244**, 522-523.
- Martin, J. G. & Bunnell, H. T. (1982) Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences, *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 473-488.
- Monsell, S. (1986) Programming of complex sequences: evidence from the timing of rapid speech and other productions. In *Generation and modulation of action patterns* (H. Heuer & C. Fromm, editors), pp. 72-86. Berlin: Springer.
- Ohde, R. N. (1984) Fundamental frequency as an acoustic correlate of stop consonant voicing, *Journal of the Acoustical Society of America*, **75**, 224-230.
- Öhman, S. E. G. (1966) Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America*, **39**, 151-168.
- Recasens, D. (1984) Vowel-to-vowel coarticulation in Catalan VCV sequences, *Journal of the Acoustical Society of America*, **76**, 1624-1635.
- Recasens, D. (1987) An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences, *Journal of Phonetics*, **15**, 299-312.
- Repp, B. H. & Lin, H.-B. (1987) Difference in second-format transitions between aspirated and unaspirated stop consonants preceding [a], *Language and Speech*, **30**, 115-129.
- Saltzman, E. L. (1986) Task dynamic coordination of the speech articulators: a preliminary model. In *Generation and modulation of action patterns* (H. Heuer & C. Fromm, editors), pp. 129-144. Berlin: Springer.
- Saltzman, E. L. & Kelso, J. A. S. (1987) Skilled actions: a task dynamic approach, *Psychological Review*, **94**, 84-106.
- Saltzman, E. L. & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production, *Ecological Psychology*, **1**, 333-382.
- Semjen, A. & Garcia-Colera, A. (1986) Planning and timing of finger-tapping sequences with a stressed element, *Journal of Motor Behavior*, **18**, 287-322.
- Sharf, D. J. (1962) Duration of post-stress inter-vocalic stops and preceding vowels, *Language and Speech*, **5**, 26-30.
- Sternberg, S., Monsell, S., Knoll, R. L. & Wright, C. E. (1978) The latency and duration of rapid movement sequences: comparisons of speech and typewriting. In *Information processing in motor control and learning* (G. E. Stelmach, editor), pp. 469-505. New York: Academic Press.
- Sternberg, S., Wright, C. E., Knoll, R. L. & Monsell, S. (1980) Motor programs in rapid speech: additional evidence. In *The perception and production of fluent speech* (R. A. Cole, Editor). Hillsdale, NJ; LEA.
- Stevens, K. N. & Keyser, S. J. (1989) Primary features and their enhancement in consonants, *Language*, **65**, 81-106.
- Summers, W. V. (1987) Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses, *Journal of the Acoustical Society of America*, **82**, 847-863.
- Watson, B. C. & Alfonso, P. J. (1983) Foreperiod and stuttering effects on acoustic laryngeal reaction time, *Journal of Fluency Disorders*, **8**, 183-205.
- Whalen, D. H. (1984) Subcategorical phonetic mismatches slow phonetic judgments, *Perception & Psychophysics*, **35**, 49-64.
- Wolf, C. G. (1978) Voicing cues in English final stops, *Journal of Phonetics*, **6**, 299-309.