Coarticulatory stability in American English /r/

Suzanne Boyce and Carol Y. Espy-Wilson
Electrical, Computer, and Systems Engineering Department, Boston University, 44 Cummington Street, Boston, Massachusetts 02215 and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 15 May 1995; accepted for publication 4 February 1997)

A number of different researchers have reported a substantial degree of variability in how American English /r/ coarticulates with neighboring segments. Acoustic and articulatory data were used to investigate this variability for speakers of ''rhotic'' American English dialects. Three issues were addressed: (1) the degree to which the $F3$ trajectory is affected by segmental context and stress, (2) to what extent the data support a ''coproduction'' versus a ''spreading'' model of coarticulation, and (3) the degree to which the major acoustic manifestation of American English /r/—the time course of $F3$—reflects tongue movement for /r/. The $F3$ formant trajectory durations were measured by automatic procedure and compared for nonsense words of the form /'waCrav/ and /wa'Crav/, where C indicates a labial, alveolar, or velar consonant. These durations were compared to $F3$ trajectory durations in /'warav/ and /wa'rav/. In addition, formant values in initial syllables of words with and without /r/ were examined for effects of intervening consonant contexts. Results indicated similar $F3$ trajectory durations across the different consonant contexts, and to a lesser degree across stress, suggesting that coarticulation of /r/ can be achieved by overlap of a stable /r/-related articulatory trajectory with movements for neighboring sounds. This interpretation, and the concordance of $F3$ time course with tongue movement for /r/, was supported by direct measures of tongue movement for one subject. © *1997 Acoustical Society of America.*
[S0001-4966(97)02106-1]

PACS numbers: 43.70.Bk, 43.70.Fq [AL]

INTRODUCTION

In the standard ''rhotic'' dialects of American English (where /r/ is pronounced in all allowable contexts), /r/ has been described as coarticulating with adjacent segments in a number of interesting ways. The best known of these effects involve vowels. For instance, vowels next to consonantal /r/ show coarticulatory effects known as ''/r/-coloring'' (Ladefoged, 1982; Giergerich, 1992; Bronstein, 1967). However, coarticulatory effects on neighboring consonants have also been described (Olive *et al.*, 1993; Shoup and Pfeifer, 1976; Zue, 1985). For speech recognition systems, this variability can result in the misclassification of nearby vowels and/or consonants as /r/ (Espy-Wilson, 1994).

This paper is concerned with acoustic and articulatory aspects of the way consonantal /r/ interacts with adjoining consonant and vowel segments in ''rhotic'' varieties of American English. Because /r/ as produced by American English speakers appears to involve several articulators acting in concert and shows wide variability in articulatory configuration between speakers, we concentrate on analysis of consistency in its acoustic signature. The results we describe are important for phonological descriptions of American English, and for the design of speech recognition systems, as well as for models of motor control in normal and disordered speech.

A. Acoustics of /r/

The most salient feature of American English /r/, whether consonantal or vocalic, is its low $F3$, which can range between 1100 and 2000 Hz but which is normally in the region of 1600 Hz for both men and women (Espy-Wilson, 1992, 1994; Nolan, 1983; Lehiste and Peterson, 1961; Lehiste, 1962; Zue, 1985; but cf. Hagiwara, 1995 for an examination of male-female differences). For other segments of American English, the typical $F3$ range occurs between 2100 and 3000 Hz (Peterson and Barney, 1952; Shoup and Pfeifer, 1976). Typically, the $F3$ transition between surrounding segments and /r/ shows a marked trajectory of movement beginning at 2000 Hz. When /r/ is surrounded by sonorant segments, a complete $F3$ trajectory representing movement toward and away from the articulatory configuration for /r/ can be seen. For all types of /r/ this trajectory resembles an inverted parabola. In general, it is reasonable to assume that the time course of frequency change in $F3$ below 2000 Hz reflects the time course of articulatory movement specific to /r/. In other words, a parabolic $F3$ trajectory below 2000 Hz reflects /r/-related movement.[1]

In a study of semivowels, Espy-Wilson (1992, 1994) found that when lowering to 2000 Hz or below was used as a criterion for /r/ identification in a speech recognition system, segments adjacent to /r/ were routinely misidentified as /r/ proper. This result reflected the fact that $F3$ values were frequently lowest on these adjacent segments, while $F3$ values during the segment transcribed as /r/ were somewhat higher. A typical example is illustrated in Fig. 1, which shows a spectrogram of the word ''everyday'' spoken by a native American English speaker. The results of a formant-tracking program (Espy-Wilson, 1987) have been superimposed on the spectrogram. Vertical lines in the phonetic transcription at top show the boundaries of /r/ and neighboring segments as assigned by a standard acoustic segmentation

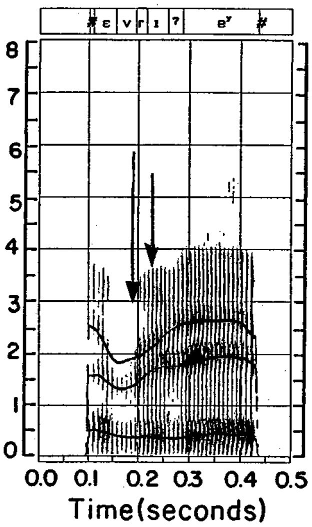

FIG. 1. Spectrogram of the word ''everyday'' with formant tracks overlaid and phonetic transcription at top. Arrows indicate boundaries assigned to the /r/ by the segmentation procedure. The lowest point of $F3$ occurs outside of the boundaries assigned to /r/ by segmentation procedure.

procedure (Seneff and Zue, 1988).[2] Note that this word shows a full, parabolic $F3$ trajectory typical of /r/. Two other facts stand out: (1) the lowest point of $F3$ (which we noted above probably corresponds to the most extreme /r/-related movement) occurs during the preceding fricative, and (2) inside the boundaries assigned to /r/ proper we see the rising portion of the acoustic /r/ trajectory (corresponding to offset from the articulatory extremum). Thus ''variability'' in the instantiation of /r/ for this word appears to involve variability in the way the articulatory movement (and associated acoustic parabolic trajectory) is placed with respect to segmental boundaries. Additionally, the $F3$ trajectory appears continuously through articulation of the labial obstruent.

B. Models of coarticulation

Classically, coarticulation is defined as an assimilation in the articulation of one segment, a ''target'' segment, as a result of a neighboring ''home'' segment. It is said to occur when the effects of one segment show up during production of another segment. Physically, coarticulation may be manifested as a change in dynamic characteristics of movement (shape/displacement/duration of the articulatory movement) as well as change in placement within the vocal tract. Because articulatory postures are attained dynamically, through movements whose trajectory exhibits a defined onset, extremum, and offset, trajectory duration may increase as a result of a longer onset, a plateau of movement at the extremum, or a longer offset. Shape may change as a result of durational change in any of these components, or because of change in the displacement of the movement.

Current theories of anticipatory coarticulation, i.e., coarticulation between a target and a following home segment, explain these effects in one of two ways. In one approach, known as the ''feature-spreading'' or ''spreading'' account, the underlying articulatory plan (including trajectory of movement, placement of movement in the vocal tract, etc.) for producing the target segment has been altered from the form it would take if it were bordered by different neighboring segments; in other words, articulatory plan varies by segmental context (Daniloff and Moll, 1968; Hammarberg, 1976; Kent and Minifie, 1977; Keating, 1988; also see Perkell and Matthies, 1992, among others). Articulatorily, this conception has two critical assumptions: (1) coarticulatory effects occur because articulatory postures associated with the ''home'' segment are achieved over an extended period of time (longer than required for the home segment *per se*), and (2) the degree of coarticulatory change will vary according to the difficulty of sustaining simultaneously the ''home'' and ''target'' articulatory postures. In particular, it is assumed that if the ''home'' and ''target'' specifications are easy to reconcile, the two segments will be coarticulated for a longer period of time, while if the ''home'' and ''target'' are difficult to reconcile (an articulator directed to be in two places at once, for instance), there will be less coarticulation. For example, lip retraction required for the vowel /i/ is considered to conflict with coarticulatory spreading of lip rounding (Hammarberg, 1976; Perkell and Matthies, 1992). In contrast, anticipatory spreading of coarticulation is expected to be at a maximum when adjacent segments involve different articulators. For instance, anticipation of tongue movement for vowels such as /i/ or /ɑ/ is considered to be maximal when the preceding consonant is a labial, since the tongue is theoretically free to move (Harris and Bell-Berti, 1984). Moreover, unrounded consonants such as /s/ and /t/ are often assumed to be potentially compatible with rounding coarticulation (cf. Perkell and Matthies, 1992; Boyce *et al.*, 1990 for overview). Predictions of coarticulatory effect are less clear when adjacent segments require movement by the same articulator in similar but not identical directions or to similar but not identical positions in the vocal tract. Investigators looking at cases such as the interaction of tongue dorsum movement for /u/ or /i/ and adjacent /k/, have concluded that the (observed) articulatory trajectories of both tend to be affected, according to constraints on individual segments (Recasens, 1985). A schematic illustration of one version of the spreading model, showing the contrast between movement for isolated segments versus segments in context, is illustrated in Fig. 2(a) and (b).

A different view, known as the ''coproduction'' approach, is that much of what we call coarticulation can be explained, not by changing the segmental articulatory plan, but as the result of overlap and consequent ''blending'' between (unaltered) articulatory plans for adjacent segments; i.e., specified articulatory trajectories for adjacent target and home segments combine to produce a movement trajectory that is intermediate between them (Munhall and Lofqvist, 1992; Gracco and Lofqvist, 1994). Arguments using articulatory and acoustic data to support this view have been advanced by Harris and Bell-Berti (1984), Gelfer *et al.* (1989), Boyce *et al.* (1990), Browman and Goldstein (1986), Browman and Goldstein (1990), Bell-Berti and Krakow (1991), Fowler (1993), and Bell-Berti *et al.* (1995) among others. This model is schematized in Fig. 2(c). Some critical assumptions of this view include (1) the underlying motor programming for articulatory movement to and from the extremum articulatory configuration remains relatively stable

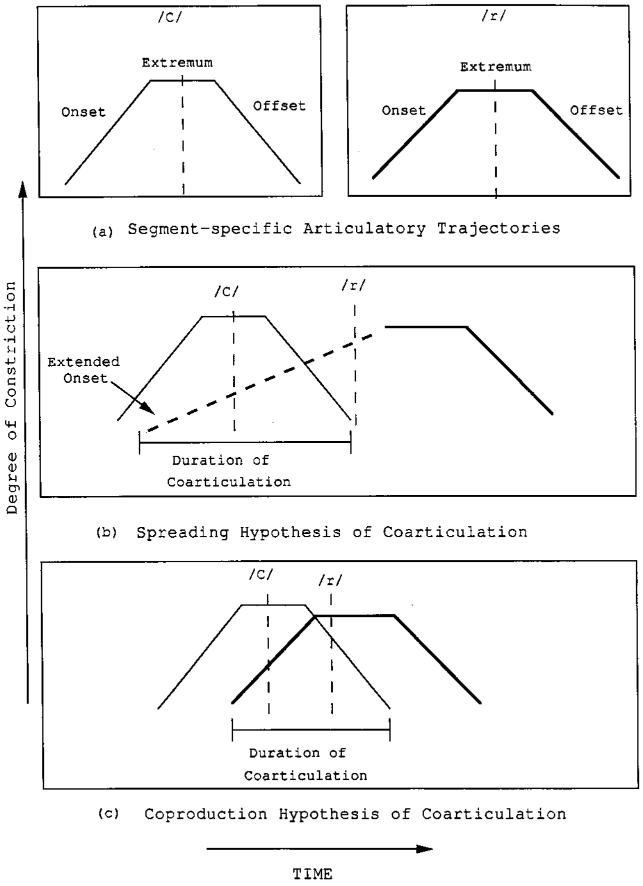
/C/
Extremum
Onset
Offset
/r/
Extremum
Onset
Offset
(a) Segment-specific Articulatory Trajectories
Degree of Constriction
/C/
/r/
Extended Onset
Duration of Coarticulation
(b) Spreading Hypothesis of Coarticulation
/C/
/r/
Duration of Coarticulation
(c) Coproduction Hypothesis of Coarticulation
TIME

FIG. 2. A schematization of two viewpoints of coarticulation: (a) individual gestures for a consonant and /r/, (b) coarticulation based on one version of the spreading model (Perkell and Matthies, 1992), (c) coarticulation based on the coproduction model.

across segmental contexts, and (2) underlying trajectories must be deduced from observed trajectories by judicious examination of observed trajectories across different segmental contexts. An important aspect of this view is that the speech motor system prefers to maintain segmental plans, and stable trajectories, when possible. Proponents of the coproduction viewpoint have also suggested that accommodation to the particular requirements of segmental context may be accomplished by displacement, or ''sliding'' of articulatory movement trajectories away from their home segment; in other words, the extremum of the articulatory posture, and thus its spatiotemporally stable onset, extremum and offset, can be shifted in time (Browman and Goldstein, 1986, 1990). For instance, difficult interactions between specifications on adjacent segments may be mediated by changing the spacing of associated articulatory movements in time. Changes in speech rate, stress, and syllable position, etc. may be accomplished either by changes in the segmental articulatory plan (Gracco and Lofqvist, 1994) or by ''sliding'' (Browman and Goldstein, 1986, 1990).

In general, data suggesting changes in articulatory trajectories due to context (and not attributable to blending) would constitute support for the ''spreading'' approach, while data suggesting stable trajectories would constitute support for the ''coproduction'' approach. For /r/, it is reasonable to assume that changes in the acoustic trajectory specific to /r/ will reflect changes in the articulatory instantiation of /r/. Thus, the spreading model predicts that we will see changes in the duration and/or shape of acoustic /r/ trajectories across segmental context. In contrast, the ''coproduction'' model predicts stability in the /r/-related $F3$ trajectory.

This issue is particularly important for /r/, because the use of ''retroflex,'' ''bunched,'' and ''mixed'' versions of /r/ (which use different combinations of tongue tip and tongue dorsum to make constrictions along the palate) varies in a nonobvious way among the population (Delattre, 1967; Delattre and Freeman, 1968; Bernthal and Bankson, 1993; Lindau, 1985; Hagiwara, 1995; Westbury *et al.*, 1995; Narayanan *et al.*, 1997). Further, complex interactions between pharyngeal constriction, labial constriction, and different types of tongue constrictions during /r/ makes coarticulatory conflicts hard to predict. Given these caveats, however, we can make the following generalizations. (1) For all /r/'s, we might expect coarticulation to occur most freely (and in the spreading model, trajectory duration/shape to change the most) when adjacent segments do not involve the tongue at all, i.e., labial and glottal consonants. (2) In addition, we might expect that the /r/ variant used by the subject would affect the way /r/ coarticulates with surrounding segments. For instance, a subject who uses his tongue tip primarily to make an oral cavity constriction for /r/ in a vocalic context might show different contextual effects on the /r/ trajectory when neighboring segments involve the tongue tip versus the tongue dorsum. A similar argument can be made for subjects who use primarily the tongue dorsum during /r/. Speakers may also respond to the difficulty of sequencing /r/ with alveolar or velar consonants by alternating between /r/ variants according to context (Espy-Wilson and Boyce, 1994). Each of these possibilities (and others not mentioned) suggests different scenarios, depending on the particular characteristics of the /r/ variant used, and the particular articulatory interactions involved. Thus although much remains unknown about how /r/ coarticulates with surrounding phones, it seems reasonable to assume that (1) for purposes of the spreading hypothesis, labial contexts provide fewer challenges to coarticulation than velar or alveolar contexts, and (2) if context has any effect, we might expect this to emerge in a comparison of the shape and duration of the $F3$ trajectory for /r/ across vocalic, labial, velar, and alveolar contexts.

I. GENERAL METHODOLOGY

Data for this study include acoustic signal data recorded from seven speakers (three female and four male) and articulatory tongue movement data recorded from one of the male speakers. Articulatory data were used to confirm methodology and conclusions from acoustic data. Methodology and results specific to the acoustic data are described in experiment 1. Methodology and results specific to articulatory data are described in experiment 2. Methodology shared between experiments is described below.

Seven speakers produced five repetitions of experimental nonsense words /wɑvrɑv/, /wɑbrɑv/, /wɑgrɑv/, /wɑdrɑv/, and five repetitions of the control nonsense words /wɑrɑv/, /wɑwɑv/, /wɑvɑv/, /wɑbɑv/, /wɑgɑv/, and /wɑdɑv/. Each nonsense word was produced in two stress conditions: with stress on the first syllable (initial stress) and with stress on the second syllable (final stress); e.g., /'wɑrɑv/, /wɑ'rɑv/. As a control for nonsense word effects, three speakers produced five repetitions each of a smaller set of real words structured to resemble a representative sample of the nonsense words, e.g., ''Africa,'' ''begrime,'' and ''barometer'' plus cases of word-initial and word-final /r/ such as ''rob'' and ''bar.'' All words were embedded in the carrier phrase ''Say ——— for me.''

The subjects produced the experimental stimuli in the same order five times with reference to a handheld paper list. For all subjects except RD, acoustic signals were digitized at 16 kHz on a SUN workstation via the ESPS/WAVES signal processing software. For subject RD, the acoustic signal was digitized at 10 kHz on a DEC workstation via the MITSYN signal processing software. For analysis, the signals were transported to a SUN workstation and subjected to signal processing using ESPS/WAVES.

Subjects were speakers of fully rhotic versions of standard American English from Missouri, western Massachusetts, upper New York state, western Pennsylvania, Michigan, Philadelphia, and Washington state. Speakers were instructed to produce words at a self-selected comfortable and consistent rate, in a natural manner, and were given a short practice session. Of the seven subjects, four (three females and one male) were phonetically sophisticated, three (all males) were not. At the time of recording, speakers HSS, BS, and MS had some notion of the purpose of the study; the four male speakers HD and RD had none.[3] The experimental nonsense words were designed to include cases with labial, alveolar, and velar consonants before /r/. The control words were included to allow analysis of the formant trajectories characteristic of these consonants as well as those of the labial most like /r/ (/w/) and of /r/ itself. Additionally, the comparison between /g/ and /w/ provided a rough indication of the extent of $F3$ lowering attributable to rounding. Segments following /r/ in the experimental nonsense words were the same across words, allowing consistent comparison in the raising portion of the $F3$ trajectories. The experimental words /wɑbrɑv/ and /wɑvrɑv/ were expected to present the most favorable conditions for coarticulation of /r/; that is, we expected that if ''spreading'' of /r/ articulation occurs, these words would show longer $F3$ trajectories (and presumably longer articulatory trajectories) than those for words with singleton /r/. If articulatory movements for a segment are spatiotemporally stable, as predicted by the coproduction model, then we expected trajectories to be the same as those for words with singleton /r/. The words /wɑgrɑv/ and /wɑdrɑv/, because /g/ and /d/ involve the tongue, were expected to represent more difficult coarticulatory challenges. The spreading model predicts that /r/ trajectories in such words would be shorter than in control words such as /wɑrɑv/, or words with labial consonants such as /wɑvrɑv/.

II. EXPERIMENT 1: ACOUSTIC INVESTIGATIONS

Acoustic data were used to determine (a) whether stress and consonantal context affect $F3$ trajectory durations, and (b) if changes in stress, context, and/or trajectory duration may affect the shape of $F3$ trajectories.

A. Methodology

All subjects except RD were recorded in a sound-treated room using a Sennheiser directional microphone and a high quality Yamaha audio cassette tape recorder. Subject RD was recorded in a quiet hard-walled room using a high-quality SONY directional electret condenser microphone. The acoustic signal for RD was recorded digitally on line using the MITSYN signal processing software.

Formant tracks were computed for all the utterances using the ESPS/WAVES formant tracker and a 10-ms frame rate. Alignment between formant tracks and spectrograms was handled automatically as part of the WAVES program. For the purpose of analyzing $F3$ trajectory duration and shape, the formant tracks were edited by the two authors working together to eliminate noisy or erroneous data points as described below. To cut down on editing, for each word the three tokens best analyzed by the WAVES formant tracker were chosen (in some instances, more tokens were included). Files containing $F3$ values from edited formant tracks were transferred to a Macintosh IIsi computer and analyzed using standard graphics and statistics programs.

All editing was done by visual reference to spectrograms for each token with results of the formant tracker superimposed, and power spectra where appropriate. Figure 3 shows illustrative spectrograms with superimposed formant tracks for tokens /wɑrɑv/, /wɑvrɑv/, /wɑdrɑv/, and /wɑgrɑv/ produced by speaker JM. Several steps were involved in editing the formant tracks. First, $F3$ tracks during the word-initial /w/ were deleted. The criterion for the start of the following vowel (V1) was the beginning of strong energy in $F1$. Second, $F3$ tracks during the word-final /v/ were deleted. The criterion for the end of the second vowel (V2) was the end of strong energy in $F1$. If the formant tracks appeared continuous and unambiguous, as in panel (a) of Fig. 3, no further editing was done. If the $F3$ tracks during the intervocalic obstruents were noisy, as in panel (c) of Fig. 3, the frequency values were deleted while maintaining the correct spacing in time between retained values. The criterion for deletion of frequency values was $F3$ spectral amplitude 30 dB or more below the amplitude at the lower frequency spectral peak. Parts (b) of the two panels of Fig. 4 show examples of edited formant tracks with and without deletion of noisy values. Note that because the initial and final consonants are eliminated from the edited version, the edited versions are shorter than the spectrographic version.

In some cases obstruents were produced with incomplete vocal tract closure, and the formant tracker was able to detect consistent and appropriate $F3$ values in at least some portion of the acoustically defined closure interval. These values were retained under any one of the following conditions: (a) there was little or no stop burst, (b) energy at low frequencies was present throughout the closure interval, and (c) the time course of the formant tracks was similar over the five repeti-

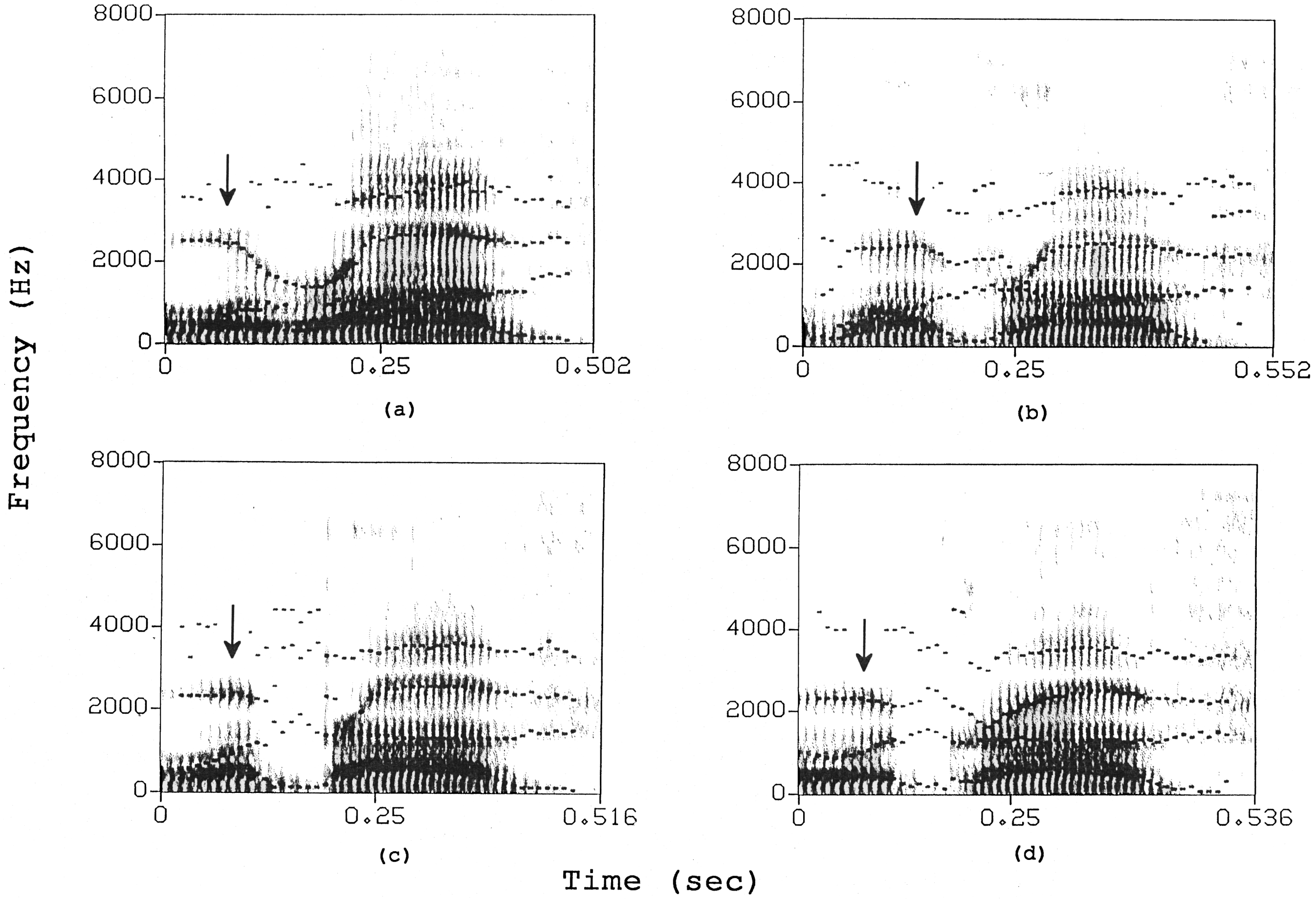

FIG. 3. Spectrograms of (a) /wɑ'rɑv/, (b) /wɑ'vrɑv/, (c) /wɑ'brɑv/, and (d) /wɑ'grɑv/ produced by speaker JM. Arrows point to syllabic peaks found during the first syllable by an automatic procedure.

tions produced by each speaker. A relatively unambiguous example of this type can be found in the left panel of Fig. 4. In some cases, the formant tracker incorrectly assigned values belonging to $F3$ as belonging to $F2$ or $F4$; these values were replaced by the correct values. Additionally, there were ambiguous cases in which the formant tracker identified energy simultaneously at two points in the spectrum which might plausibly reflect $F3$. Examples of these cases are shown in panels (b) and (d) of Fig. 3. (Generally, the two paths would be represented as belonging to $F2$ and $F3$, or $F3$ and $F4$, or some mixture of the two.) Almost invariably in these cases of ''double'' paths, one track resembled the pattern of $F3$ seen during the closure in control words /wɑdɑv/, /wɑgɑv/, /wɑbɑv/, /wɑvɑv/, while the visible portion of the other track resembled the pattern seen for /r/ in /wɑrɑv/. Our strategy for dealing with these cases is described below (see Sec. II A 1). In uncertain cases, formant values were determined to be valid or invalid by referring to formant patterns in the control utterances, including /wɑrɑv/.

1. "Double" F3 paths

Cases in the data with discernible ''double path'' resonances (i.e., there were two simultaneous resonances that might be called $F3$) occurred in all consonant contexts, for both stress conditions, and for all speakers. Parallel cases of ''double'' trajectories for representative tokens of /'wɑvrɑv/ (produced by speaker RD), and /wɑ'drɑv/ (produced by speaker JM), are shown in Figs. 5 and 6. These are contrasted with representative tokens of /'wɑrɑv/ or /wɑ'rɑv/, and control words /'wɑvɑv/ or /wɑ'dɑv/, as appropriate. In the /'wɑvrɑv/ case, energy was present throughout the fricative constriction, and formant tracks for both paths were relatively continuous. In the /wɑ'drɑv/ case, the formant tracks show evidence of both paths, but in a less continuous fashion.

The pattern shown by resonances in the range 1500–2500 Hz during the intervocalic intervals of /'wɑvrɑv/ (Fig. 5) is typical. At the end of the initial vowel (V1) two resonances appear that might be labeled $F3$; these are greatly attenuated during consonant constriction (depending on the degree of constriction) but may still be discerned in the signal. As can be seen, the ''lower path'' resonance trajectory in /'wɑvrɑv/ is quite similar to what we see for the $F3$ trajectory in the control word /'wɑrɑv/. However, if we follow the ''upper'' path resonance trajectory, two points stand out: (1) that the falling-rising portion of the trajectory is much shorter than that seen in /'wɑrɑv/ and occurs a considerable time after the end of V1, and (2) that immediately after the end of V1, and during the /v/ constriction, the ''upper path'' resonance trajectory resembles the $F3$ values tracked during the /v/ constriction in /'wɑvɑv/. The ''double'' resonance pattern of the /wɑ'drɑv/ token shown in Fig. 6 bears a simi-

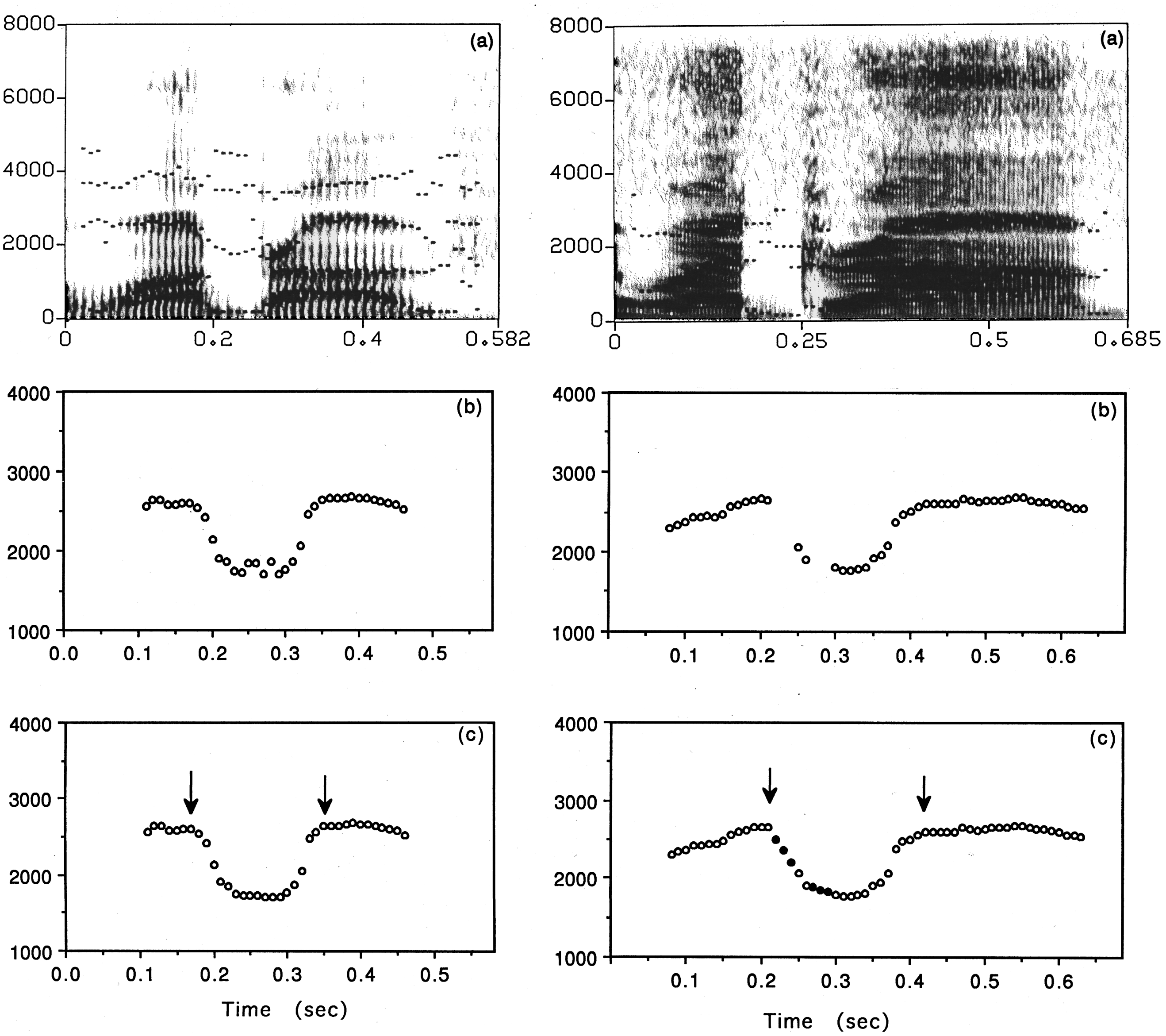

FIG. 4. (a) Spectrograms of /wɑ'brɑv/ produced by JM (left panel) and /wɑ'drɑv/ produced by HSS (right panel) with formant tracks overlaid. (b) Edited *F*3 tracks. (c) Smoothed and/or interpolated edited *F*3 tracks with automatically determined inflection points indicated with arrows.

lar relationship to the *F*3 trajectories of /wɑ'rɑv/ and /wɑ'dɑv/. In this case, however, while the initial lowering of *F*3 at the end of the vowel is evident, there are missing values during the /d/ closure. As with the /'wɑvrɑv/ case discussed above, the observable portion of the lower path aligns well with the *F*3 trajectory visible in /'wɑrɑv/, whereas the ''upper'' path trajectory resembles that of the /d/ in /'wɑdɑv/. The shorter duration of the fall–rise portion of the ''upper path'' trajectory can be attributed to constriction narrowing for the contextual consonant rather than the slower articulation of /r/. Altogether, it seems clear that the ''upper path'' trajectory in these situations reflects the influence of the obstruent preceding /r/, while the ''lower path'' trajectory reflects the influence of the /r/. (Presumably, the variation we see in whether ''double'' resonances can be discerned in the signal, and whether the ''upper path,'' or ''lower path'' resonance is stronger, can be attributed to normal token-to-token variation in the articulation of consonant and /r/ segments.) This reasoning was confirmed by articulatory data from RD (see Sec. III). Thus, the ''lower path'' *F*3 trajectory was the object of measurement in all cases.

2. Identifying trajectory end points

Because visual identification of trajectory beginning and end would be subject to experimenter bias, an automatic procedure was developed to identify trajectory beginning and end points for the /r/-related *F*3 trajectory of initial- and final-stress tokens of the /wɑrɑv/, /wɑbrɑv/, /wɑvrɑv/, /wɑdrɑv/, and /wɑgrɑv/ nonsense words. Typically, trajectories in these data show some gradual lowering and raising movement on the periphery prior to, and following, an identifiable ''bend'' associated with /r/. We defined trajectory edges as inflection points at these ''bends,'' and duration of the trajectory as the time between inflection points. Our program found these inflection points based on a combination of the first and second differences of the *F*3 trajectory. When

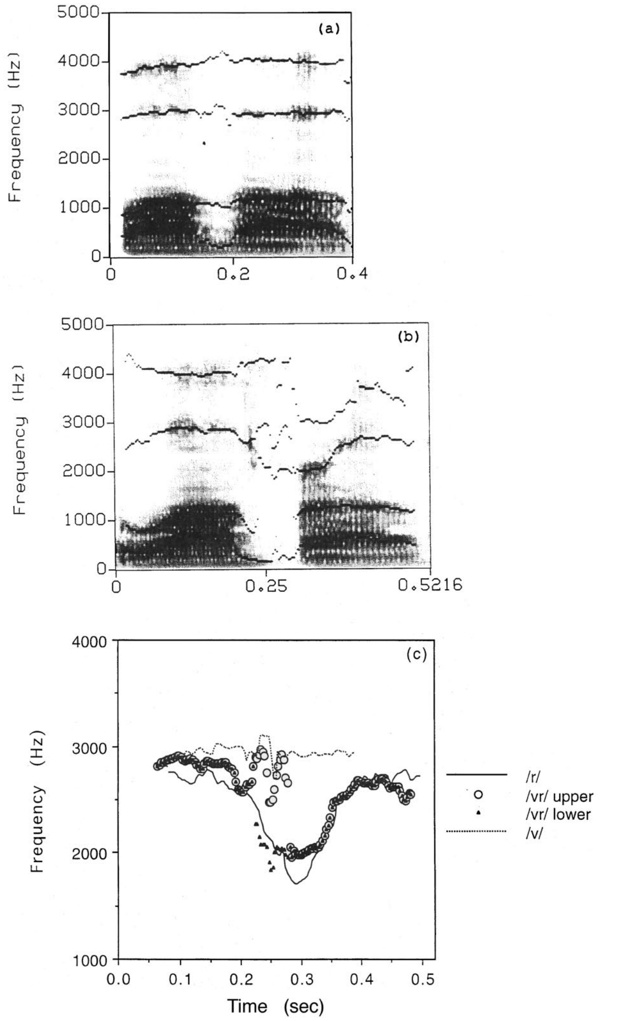

FIG. 5. (a) Spectrogram of /'wɑvɑv/ with formant tracks overlaid. (b) Spectrogram of /'wɑvrɑv/ with formant tracks overlaid that show two paths for $F3$ during the /v/. (c) Comparison of formant tracks taken from one token of /'wɑrɑv/ and from /'wɑvɑv/ and /'wɑvrɑv/. Data are from speaker RD. Note that part (c) is repeated in Fig. 9 which shows alignment with articulatory data.

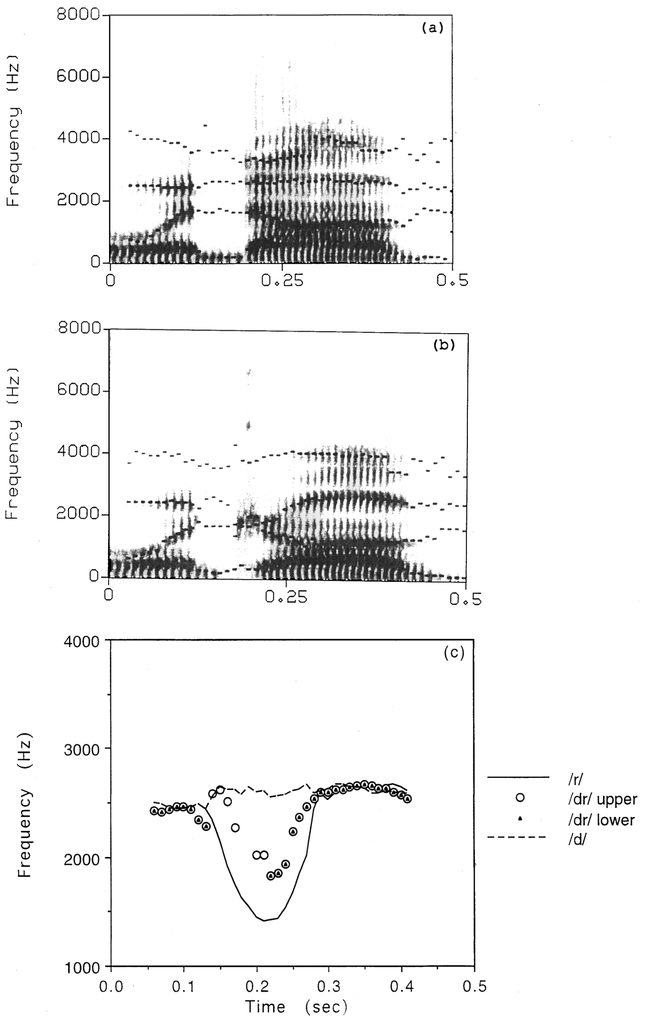

FIG. 6. (a) Spectrogram of /wɑ'dɑv/ with formants tracks overlaid. (b) Spectrogram of /wɑ'drɑv/ with formants tracks overlaid that show two paths for $F3$ during the /d/ closure. (c) Comparison of formant tracks taken from one token of /wɑ'rɑv/ [the spectrogram of this token is shown in Fig. 3(a)] and from /wɑ'dɑv/ and /wɑ'drɑv/. Data are from speaker JM.

two competing inflection points were found, the more peripheral was used. Minor local perturbations in the $F3$ tracks were smoothed by hand, and missing values (corresponding to ill-tracked or noisy values eliminated during editing) were filled in by a simple linear interpolation algorithm, producing a continuous trajectory. Parts (c) of Fig. 4 show examples of interpolated $F3$ tracks where the interpolated values are indicated by filled squares. Arrows show inflection points as found by this algorithm.

Variability due to the automatic procedure was of two types. First, because the automatic procedure was forbidden to assign trajectory beginning during the interpolated portion, and trajectory beginning was typically identified on the left edge of the interpolated region, the automatic procedure tended to find slightly longer trajectory durations for /Cr/ word tokens with noisy consonant closure intervals, in contrast to measurements for tokens with voicing through the consonant closure interval or for singleton /r/ words. Second, minor differences in trajectory slope on the right and left edges could affect the determination of inflection point. (Slope was in turn dependent on the $F3$ values for initial and final vowels, which were idiosyncratic to speaker as well as to degree of vowel reduction and stress.) An error of 1 sample point in locating trajectory end points corresponded to an error of 10 ms, due to the 10-ms frame rate for the formant tracker. Tokens where the automatic procedure missed a visually identifiable peripheral inflection point were adjusted by hand. We estimate error conservatively at ±20 ms for each trajectory end point. Thus two trajectories of equal duration might conceivably be measured as different by 40 ms. Figure 7 illustrates the typical situation found across tokens for all subjects in our study. Although this token of /wɑ'rɑv/ and two tokens of /wɑ'vrɑv/ produced by speaker HD have extremely similar $F3$ trajectories, and visual measurement would identify very similar trajectory beginning and ending points, sensitivity to minor differences in slope caused the automatic procedure to calculate the duration differences between inflection points (i.e., the trajectory durations) as 180 ms for the singleton /r/ token (inflection points indicated by open arrows) versus 220 and 230 ms for the two /vr/ tokens (inflection points indicated by filled ar-

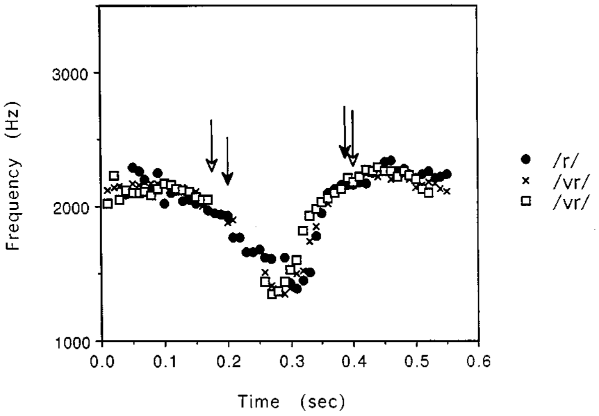

FIG. 7. Edited *F*3 tracks of one token of /wɑ'rɑv/ and two tokens of /wɑ'vrɑv/ produced by HD. Arrows show automatically assigned inflection points.

rows). The token of /wɑ'grɑv/ illustrated in Fig. 8 is another case in point; although the similarity between /wɑ'grɑv/ and other tokens is patent, a slightly more gradual slope along the right-hand edge caused the /wɑ'grɑv/ token trajectory duration to be computed as 250 ms. (Other tokens of /wɑ'grɑv/ for this subject showed similar trajectories but were measured at 210 and 190 ms.) Similarly, the durations of the remaining trajectories in Fig. 8 vary between 160 ms (in the case of /wɑ'rɑv/, inflection points indicated by filled arrows) and 200 ms (in the case of /wɑ'brɑv/, inflection points indicated by open arrows). Random pairings of tokens across the dataset for all speakers showed parallel patterns of variability. Thus, our measurement procedure appeared likely to overestimate the true variability of the dataset. Although this was not ideal, any findings of consistent behavior were unlikely to be artifactual in nature.

B. Qualitative and quantitative results

The feature-spreading account of coarticulation predicts that the *F*3 trajectory for /r/ will vary in trajectory shape and duration across contexts (although its visibility may be obscured at points by token-to-token variation and by the acoustic effects of neighboring sounds). In contrast, the data showed consistent evidence of trajectory similarity across the dataset. This similarity was maintained between tokens of the same word, across different consonant contexts, and to a large degree across stress.

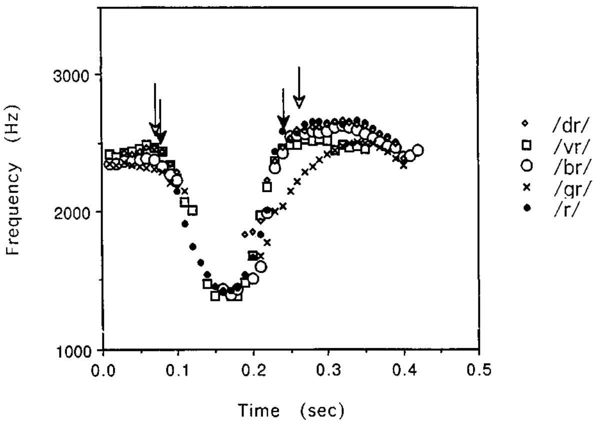

FIG. 8. Edited *F*3 tracks of one token of /wɑ'rɑv/, /wɑ'vrɑv/, /wɑ'brɑv/, /wɑ'drɑv/, and /wɑ'grɑv/ produced by JM (smoothing and interpolation are not shown). Spectrograms of these tokens are shown in Fig. 3 and part (b) of Fig. 6.

1. Shape of F3 trajectory

Figure 7 shows edited *F*3 tracks (before interpolation or smoothing) extracted from one token of /wɑ'rɑv/ and two tokens of /wɑ'vrɑv/ spoken by subject HD. All tokens are lined up at the beginning of V1, which was of approximately the same duration for each. It is clear that the *F*3 trajectories of the /wɑ'vrɑv/ tokens show extremely similar falling–rising shape and duration, as predicted. [To emphasize the similarity between these trajectories and that for /wɑ'rɑv/ (which does not include a consonant interval), the *F*3 trajectory for /r/ in /wɑ'rɑv/ was shifted to the right by 30 ms.] This token-to-token similarity in *F*3 trajectory shape was consistent across the dataset, although for some tokens the trajectories might begin earlier or later in time. A similar picture for tokens of different consonant-/r/ combinations and singleton /r/ can be seen in Fig. 8, which shows the edited *F*3 tracks taken from one token each of /wɑ'vrɑv/, /wɑ'brɑv/, /wɑ'grɑv/, /wɑ'drɑv/, and /wɑ'rɑv/ for the same subject. The tokens of /wɑ'vrɑv/, /wɑ'grɑv/, and /wɑ'brɑv/ are lined up at the beginning of V1. For all tokens except /wɑ'drɑv/, the duration of V1 and of the occlusion interval were approximately the same. (Here the *F*3 trajectory for /r/ in /wɑ'rɑv/ was shifted to the right by 70 ms.) Because for this token of /wɑ'drɑv/ the vowel and occlusion interval were slightly shorter than those for the other tokens, the /wɑ'drɑv/ trajectory was shifted to the right by 20 ms. Spectrograms (with formant tracks superimposed) of these words are shown in Figs. 3 and 6. The similarity in both shape and duration of the trajectories is striking. This similarity was repeated for each speaker's data across the dataset.

2. Duration of F3 trajectory

The consistency of trajectory durations was tested statistically as follows. The *F*3 duration values were entered into analyses of variance using the factors speaker (HD, RD, WJ, JM, SS, MS, BS), stress (initial or final), and context (/b/, /d/, /g/, /v/, or /r/). The hypotheses being considered were (1) whether *F*3 trajectory duration differs as a function of stress condition, and (2) whether *F*3 trajectory duration differs across contexts (speaker-to-speaker differences were expected). Because of correlations naturally existing across data from particular subjects, particular items, and particular stress patterns, we elected to treat each of these factors as a correlated variable in a repeated measures analysis of variance. Separate ''subject'' and ''item'' repeated measures analyses of variance were performed using, respectively, subject variability, consonant context variability, and stress variability as the error term. The subject analysis used context and stress as ''within,'' or ''repeated'' measures while speaker was a ''between'' or ''grouping'' factor. The items analysis used context and stress as ''between'' variables while subject variability was a ''within'' measure. Because for subject RD data from /br/ was not collected, the items analysis context factor included /vr/, /dr/, /gr/, and /r/. The dependent variable in all cases was duration of the $F3$ trajectory. For both sets of analyses, individual cells were represented for each analysis by means across tokens for a particular subject×context×stress combination. (Standard errors of tokens within all combinations exhibited a range of 0–25 ms, with a mean standard error of 9.96 ms.)[4]

Overall, context was not significant in either analysis (subject: df=4,23, $F=1.99$, $p>0.10$; item: df=3,3, $F=0.161$, $p>0.10$), suggesting that $F3$ trajectory duration was consistent regardless of whether /r/ was the single intervocalic consonant, or whether it followed /b/, /v/, /d/, or /g/. Stress was significant in both subject and item analyses (subject: df=1,6, $F=28.8$, $p<0.01$; item: df=1,3, $F=99.8$, $p<0.001$), indicating that measured $F3$ trajectory duration was different according to stress pattern. The interaction of stress×context, tested in the subject analysis, was not significant (subject: df=4, 23, $F=1.5$ $p>0.10$), suggesting that the effect of stress pattern was consistent across categories of items. There was a significant effect of speaker (item: df=6, 18, $F=23.8$, $p<0.001$) but interactions between context and speaker (item: df=18, 18, $F=1.1$, $p>0.10$) or stress and speaker were not significant (item: df=6, 18, $F=1.8$, $p>0.10$), suggesting that although speaker identity affected the duration of measured $F3$ trajectories, these effects were consistent across all other variables. The speaker effect reflected characteristically longer or shorter $F3$ trajectories for different speakers, presumably relating to intrinsic differences between subjects in terms of tongue musculature, mouth size, speech motor habits, etc. Subject-to-subject differences for singleton /r/ (/wɑrɑv/) words, for instance, ranged from speaker BS's 206 ms to speaker HD's 264 ms. The overall mean across all subjects and all contexts was 224 ms.

The stress effect appeared to be due to a tendency for the $F3$ trajectory edge detection algorithm to find a longer $F3$ trajectory (by approximately 30 ms) in words whose initial vowel was unstressed. Trajectory length was positively correlated, across speakers and tokens, with the degree to which an unstressed initial /ɑ/ vowel was reduced (to /ə/). The longest trajectory measurements were seen for speakers whose natural $F3$ in back and central vowels was relatively low. Because identification of the /r/ trajectory beginning was dependent on the degree of lowering from /ɑ/, it is not clear how much of the stress effect is attributable to expansion of the /r/-related $F3$ trajectory and how much to difficulty in automatic identification of a relatively nonsalient inflection point. The fact that the slope of $F3$ lowering and trajectory shape was extremely consistent for all speakers across stress suggests the latter (see Fig. 7).

As noted above, the context factor did not reach significance in either subject or item analyses, suggesting that trajectories for /r/ in labial, alveolar, and velar contexts were similar to those in singleton /r/ words. There were no significant interactions between speaker and context or between context and stress, indicating that the effects of context and stress were similar across subjects. Separate subject and item analyses excluding singleton /r/ words also showed no significant effect of context; that is, there were no significant differences in trajectory duration between consonant contexts /b/, /v/, /d/, or /g/. There was a (nonsignificant) trend in the data for measured singleton /r/ trajectories to be shorter than those for /Cr/ words by approximately 20 ms.[4] Because trajectories for /wɑrɑv/ words were visually similar to those with /Cr/ consonants (random groupings of trajectories appearing very like those pictured in Figs. 7 and 8), we attribute the slightly longer measured trajectory durations in /Cr/ words (versus singleton /r/ words) to the existence of the consonant closure interval and consequent measurement artifact due to interpolation (see Sec. II A above). (The number of tokens where $F3$ in the closure interval was sufficiently noisy to require interpolation, and the duration of interpolated regions was approximately the same for different consonant contexts pooled across speakers, although each speaker's pattern was different.)

3. Effects beyond trajectory edges

As Figs. 7 and 8 indicate, there is some lowering of $F3$ before trajectory beginning as identified by the automatic procedure. It is possible that such lowering is anticipatory in nature as predicted by the feature spreading model; i.e., the time at which lowering begins may expand and contract according to context. Anticipatory lowering of this type would be expected to differ according to the identity of the consonant before /r/; again, earlier lowering would be expected when the intervening consonant was labial, while less and/or later lowering would be expected when the intervening consonant was alveolar or velar. Alternatively, lowering before trajectory edge may be part of a stable articulatory complex of movement for /r/. To test this question, formant values for /Cr/ and control words with singleton consonants were compared. Formant values were measured at syllable peaks as determined by an automatic procedure that identified the leftmost energy maximum in a 640–2800 Hz band and averaged the frequency value at this point with the values of the preceding and following frames. This method identified reliable formant values in a region both close to trajectory edge and salient for vowel perceptual identity. Syllable peak time points are illustrated by arrows in Fig. 3.

If the lowering we see before trajectory edge is under way by initial syllable peak, we might expect that formant values for control words would be slightly higher than those for /Cr/ words. This will be true whether the lowering reflects part of a relatively consistent, stable /r/-related movement, or if it reflects spreading of the /r/-related movement into preceding segments. However, the feature spreading model predicts more lowering when the consonant context is labial than alveolar or velar. If the lowering does not take place during the syllable peak, but after it (i.e., if /r/-related lowering does not start until after the syllable peak, we expect formant values for control words and /Cr/ words to be the same. The coproduction model predicts that lowering may or may not occur during the initial syllable peak, depending on the placement of the stable /r/ trajectory relative to the rest of the word. The amount of any lowering found would also be dependent on placement of the trajectory, which might vary according to context. Thus, a finding of lowering by itself is compatible with both coarticulation models. However, a finding of consistency in the amount of lowering across different contexts is most compatible with the coproduction model.

The amount of lowering prior to initial syllable peak was tested statistically as follows. Formant values were entered into subject and item repeated measures analyses as described above for trajectory duration testing, using the factors stress (word-initial syllables were stressed or unstressed according to word stress condition), rhotic (control words versus words containing /Cr/ clusters), consonant (/b/, /d/, /g/, or /v/) and speaker (subject). Because of missing /b/ context data from subject RD, two items analyses were performed, one using only /d/, /g/, and /v/ contexts and one that repeated /v/ data for /b/ context in /b/ cells. The pattern of results was the same; only the latter analysis is reported below. Both analyses used means across tokens for subject ×context×stress×rhotic combinations. (Standard errors for tokens within cells ranged from 1–141 Hz, with a mean standard error of 37 Hz.[5]) Overall, the effects of stress and rhotic were significant in both the subject and items analysis; stress (subject: df=1, 5, $F=10.9$, $p<0.05$; item: df=1, 10, $F=111.8$, $p<0.001$), rhotic (subject: df=1, 5, $F=10.5$, $p<0.05$; item: df=1, 10, $F=31.9$, $p<0.001$). These results were due to overall higher $F3$ at the measurement point for stressed /ɑ/ vowels, probably due to reduction during unstressed vowels, and overall lower $F3$ at the measurement point for vowels in /Cr/ words, probably due to proximity to the $F3$ trajectory. The mean difference between $F3$ at syllable peak for /Cr/ versus control words was 53.4 Hz. The mean difference between $F3$ at syllable peak for initial stress versus final stress words was 103.9 Hz. Speaker was not a separate variable in the subject analysis (data entries being treated as correlated) but was significant in the items analysis (item=6,60, $F=16.6$, $p<0.001$). This result was expected given differences in vocal tract geometry for different speakers. The main effect of consonant was significant as well (subject: df=3,15, $F=4.3$, $p<0.05$; items: df=3, 10, $F=6.3$, $p<0.05$), indicating that proximity to different consonants affects $F3$ during the preceding vowel. Among interactions, only those between speaker and variables stress and rhotic were significant; stress×speaker (df=6,60, $F=5.1$, $p<0.01$); rhotic×speaker (df=6,60, $F=2.6$, $p<0.05$). (Interactions with speaker were assessed in the items analysis only, as speaker was not a main variable in the subject analysis.) The interaction consonant×speaker was not significant (df=18,60, $F=1.1$, $p>0.05$). The interaction of stress with speaker was expected, given that different speakers had different patterns of vowel reduction. The interaction of rhotic with speaker indicates different amounts of lowering (in Hz) for /r/ during the preceding vowel. This could be due to the fact that different speakers have different formant values for $F3$ during /ɑ/, and thus lower less or more for /r/ as a normal part of the /r/ trajectory. In addition, we know that speakers differ in the overall duration of their trajectories; if these trajectories have stable durations across context they are likely to begin at different points in the vowel (a plausible version of the coproduction model). Alternatively, it could reflect changes in the consonant×rhotic interaction across speakers. This would be a plausible variant of the feature spreading model. However, such an effect should be echoed in a significant consonant×rhotic interaction and/or a significant consonant ×rhotic×stress interaction. In this study, interaction effects between nonspeaker variables were assessed in the subject analysis; none were significant; stress × rhotic (subject: df=1,5, $F=0.23$, $p>0.05$), consonant×stress (subject: df=3,15, $F=0.18$, $p>0.05$) and consonant×rhotic (subject: df=3,15, $F=0.44$, $p>0.05$), stress×consonant ×rhotic (df=3,15, $F=0.5$, $p>0.05$) (Similar results were obtained in a 3-factor ANOVA treating subject and item identity as uncorrelated variables, with the exception that in this analysis no interactions were significant.) The lack of interaction effects with consonant suggests that early lowering, like the duration of the bulk of the trajectory itself, is not affected by consonant context.

III. EXPERIMENT 2: ARTICULATORY MOVEMENT FOR /r/

A. Methodology

To get some idea of correspondence between $F3$ trajectory and articulatory movement of the tongue, and to confirm the validity of the $F3$ trajectory duration measure, articulatory plus acoustic data were obtained from one speaker, RD, who produced a subset of the experimental corpus listed above. These articulatory data from RD were used to confirm: (a) that the acoustic time course of $F3$ represents the articulatory time course of (primary constriction) tongue movement for /r/, (b) that use of the ‘‘lower’’ rather than ‘‘upper’’ path for this subject, and, by analogy, other subjects, is the appropriate acoustic index of /r/, and (c) that articulatory trajectories show shape and duration consistency similar to that found for $F3$ trajectories. Because of the difficulty of visually or algorithmically separating tongue movement for /r/ from that for /g/ and /d/, only the comparison between singleton /r/ and /vr/ contexts is shown here.

The corpus included all nonsense words with the exception of /wɑbɑv/ and /wɑbrɑv/ as well as real words with parallel structure. Movement of two electromagnetic transducers placed on the tongue tip and tongue dorsum was recorded via an Electro-Magnetic Midsagittal Articulometer (EMMA) apparatus (Perkell *et al.*, 1992). The acoustic signal was recorded with a directional microphone. The subject was seated in a quiet room and produced the experimental stimuli in randomized order by visual reference to a list suspended at eye level. The subject's head was restrained from movement by a specially designed headpiece. Transducers were attached to the subject's tongue at approximately 1 and 5 cm from the apex of the tongue tip along the tongue midline. Again, the speaker was instructed to maintain a consistent speaking rate. Movement in the anterior-posterior (X) and superior-inferior (Y) dimension was recorded separately for the tongue tip (TT) and the tongue dorsum (TD) transducer. Movement signals were digitized simultaneously with the audio signals at a rate of 312.5 Hz. Acoustic formant track data and movement signal frame rates were matched by recomputing the acoustic formant tracks with a 51.2-ms window and a 3.2-ms frame rate. These were aligned by taking

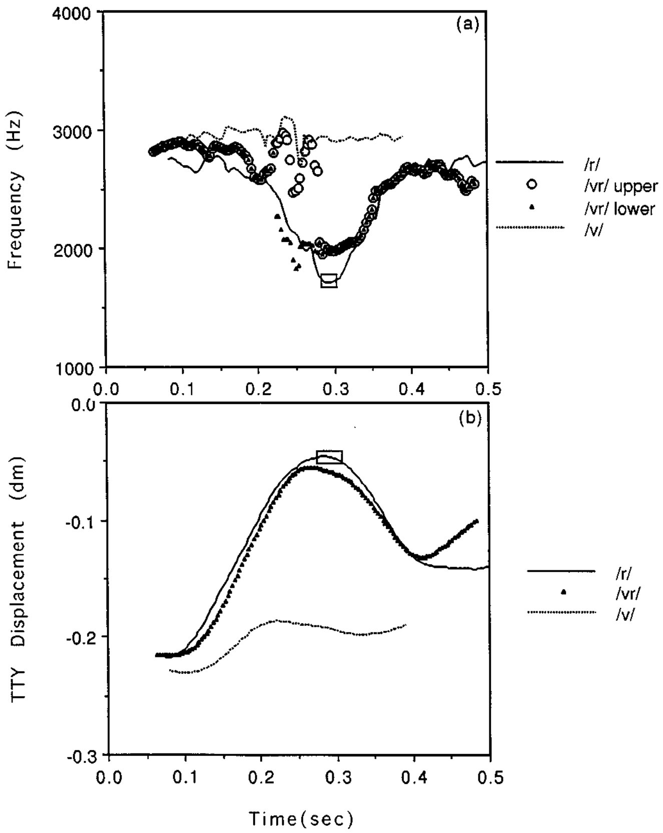

FIG. 9. (a) *F*3 tracks shown in part (c) of Fig. 5 for speaker RD. (b) Corresponding articulatory data showing the time course of the upward (superior-inferior) movement of the tongue tip during the /ɑrɑ/ of /’wɑrɑv/, the /ɑvrɑ/ of /’wɑvrɑv/, and the /ɑvɑ/ of /’wɑvɑv/.

into account a shift of half the window length (i.e., eight sample points).

B. Correspondence of acoustic and articulatory time course for singleton /r/ words

Subject RD generally raised his tongue tip to make the primary constriction for /r/ [except for the /gr/ context, where tongue dorsum was raised (Espy-Wilson and Boyce, 1994)]. The correspondence between articulatory movement and *F*3 trajectory for /r/ proper was explored in detail as follows: For 11 tokens of intervocalic /r/ (six tokens of /wɑrɑv/, two tokens of ‘‘pariah,’’ two tokens of ‘‘barometer’’ and one token of ‘‘bar’’), maximum tongue tip transducer (TTY) and minimum *F*3 values were determined by automatic procedure. These were an average of 21.5 ms apart in time, with the *F*3 minimum occurring after the TTY maximum in all cases (range: 6–38 ms). When measurement error was taken into account (i.e., ranges of TTY or *F*3 points to the left and right of extrema within a measurement error of 0.3 mm for TTY and 10 Hz for *F*3),[6] the TTY maximum overlapped with the leftmost *F*3 minimum in all cases. This is illustrated in Fig. 9, which shows the acoustic *F*3 trajectory for tokens of /’wɑrɑv/, /’wɑvrɑv/, and /’wɑvɑv/ together with time-aligned superior-inferior (*Y*) movement of the tongue tip. Ranges for *F*3 minimum and TTY maximum are indicated by rectangular boxes. As in previous figures, *F*3 and TTY values start at the beginning of the stressed /ɑ/ and end with the end of the unstressed /ɑ/. It is clear from Fig. 9 that the acoustic *F*3 trajectory and the TTY trajectory parallel each other. No clear demarcation exists in the articulatory TTY data to correspond with the inflection points on the *F*3 trajectory, i.e., where *F*3 became unambiguously associated with /r/, but the tongue tip and *F*3 tracks show broad peaks and valleys occurring at congruent points in time. These data confirm that the *F*3 trajectory is a reasonable index of articulatory movement for /r/.

Articulatory data from /Cr/ tokens also confirmed that movement related to /r/ was reflected in ‘‘lower’’ rather than ‘‘upper’’ path *F*3 trajectories. This is shown in Fig. 9, which combines the acoustic trajectories shown in Fig. 5 with their associated TTY movement. As noted above, the TTY maxima and *F*3 minima for the ‘‘lower path’’ occur at synchronous points in time. In contrast, the same comparison for the ‘‘upper path’’ trajectory in /’wɑvrɑv/ reveals no visible congruence of timing. Rather, the ‘‘upper path’’ resembles the TTY trajectory for /’wɑvɑv/ in that both show little change from *F*3 and tongue movement position for /ɑ/. These data confirm that the acoustic ‘‘upper path’’ data reflect articulatory shaping of the vocal tract specific to /v/.

Altogether, congruent relationships between the ‘‘lower’’ *F*3 trajectory track and maximum TTY movement (or maximum TDY, for /gr/ words) were characteristic of all real and nonsense words containing /r/ produced by subject RD. We interpret this trajectory congruence (F3 lowering and TTY raising for /dr/, *F*3 lowering and TDY raising for /gr/) as showing that, everything else being equal, ‘‘lower path’’ *F*3 lowering in the acoustic domain is a reasonable index of the time course of articulatory movement specific to /r/.

IV. CONCLUSION

The data in this study demonstrate that *F*3 trajectories for /r/, for any one subject, show relatively consistent duration and shape across a number of variables that might be expected to affect the way /r/ is articulated. Notably, the qualitative similarity in trajectory shape suggests that duration for all components of *F*3 trajectories—early onset (lowering), extremum, and offset (raising)—remains consistent across phonetic context. Similarly, the measured duration of the full *F*3 trajectory is consistent across phonetic contexts. Thus it appears that whether the segment preceding /r/ is alveolar, velar, labial, or vocalic does not affect the essential shape or duration of the *F*3 trajectory. In a global sense, this result is more consistent with the coproduction model of coarticulation, which predicts that articulatory trajectories for a particular segment will tend to show stable profiles across segmental contexts, than with the traditional ‘‘feature spreading’’ model, which predicts that *F*3 trajectories will change shape and duration (i.e., lengthen or shorten) according to the articulatory and acoustic requirements of the adjoining segments. Given the results reported here, it is unclear whether stress change accompanied by vowel reduction operates to change *F*3 trajectory duration (primarily by flattening the lowering curve from the initial vowel). Note that an effect of this nature due to stress is consistent with both coarticulation models. However, the minimal effect of vowel reduction and/or stress on trajectory shape seen in the present data suggests that any effect is minor at best.

As noted above, /r/ is an articulatorily complex segment, involving variant forms of tongue tip and tongue body movement as well as varying combinations of these with labial and pharyngeal narrowing. Subjects may have a number of strategies available to deal with articulatory difficulty in combining /r/ with alveolar or velar contexts. For instance, a speaker may alternate between use of /r/ variants so as to use the tongue dorsum for /r/ in alveolar contexts, and the tongue tip for /r/ in velar contexts (Espy-Wilson and Boyce, 1994). From our articulatory data, it appears that subject RD used compatible forms of retroflex /r/ in /dr/ contexts and bunched /r/ in /gr/ contexts. The acoustic evidence presented for the subjects examined here does not allow us to identify the precise articulatory mechanisms involved in producing /r/ for the remaining subjects, words, and contexts examined here. However, even if we assume that the subjects in this study used varied articulatory strategies, and that these strategies varied both idiosyncratically and by context, we might expect such variation in articulatory strategies to result in some degree of variability of acoustic patterns for /r/ coarticulation between speakers and across contexts. Certainly, we know of no external articulatory factors that would prevent /r/ variation in duration and shape over context. In view of this articulatory complexity, the consistency in duration and shape exhibited in this study by the acoustic $F3$ trajectory for /r/ across contexts is notable. This is particularly the case when we consider the similarity between singleton /r/ and labial contexts, and between labial and lingual consonant environments. Altogether, these results suggest that articulatory movement for /r/ may be organized specifically to achieve a consistent acoustic pattern. In other words, the maintenance of acoustic (and articulatory) movement profiles over segmental context may be an organizing principle of the speech system.

This study started with an observation of consistency in $F3$ trajectories, accompanied by speculation that much of what appears to be variability in the acoustic record is more apparent than real. Consequently, we suggest that much of what has been described as phonological and coarticulatory interaction between /r/ and surrounding segments can be attributed to trajectory overlap and ‘‘sliding’’ rather than ‘‘spreading’’ of /r/-related characteristics and attendant change in the articulatory plan for /r/. With regard to possible articulatory mechanisms of trajectory maintenance, we can only speculate. One possibility is that subjects ‘‘swap’’ between bunched and retroflex articulations of /r/ according to the requirements of context. Alternatively, the fact that overlapping constrictions for obstruents such as /g/ and /d/ interfered so little with $F3$ trajectories for /r/ suggests that constriction location is less important than some other manipulation of the vocal tract affecting resonance. That is, perhaps the primary producer of /r/-related movement, and the source of its uniquely low $F3$, is not the tongue tip or tongue dorsum place of articulation *per se*, but simultaneous narrowing at some more peripheral portion of the vocal tract, or the formation of an extra resonating cavity, or some complex interaction between these factors. It will be necessary, in the future, to expand these findings to additional data from real words, with different articulators, and with different numbers of syllables and different vocalic contexts.

ACKNOWLEDGMENTS

The articulatory data discussed in this paper were collected with the generous help of Dr. Joseph Perkell, Dr. Melanie Matthies, Dr. Mario Svirsky, and Jane Wozniak. We are indebted to them and to many others for comments and suggestions, including Dr. Kenneth Stevens, Nabil Bitar, Demetrios Paneras, Dr. Caroline Huang, Dr. Sharon Manuel, Dr. Rena Krakow, Dr. Patricia Keating, Dr. Daniel Recasens, Dr. Jan Edwards, one anonymous reviewer, and associate editor Dr. Anders Lofqvist. In addition, we thank Dr. Raymond Kent for sharing his recollection of past x-ray studies of /r/. This work was supported in part by: NSF Grants No. 9296032 and No. IRI-9310518, a Clare Boothe Luce fellowship to the second author, and a Boston University College of Engineering Dean’s postdoctoral fellowship, and NIH Grant No. 1R030C-02576 to the first author.

[1] It is notable that except for the vowel /i/, where $F3$ is around 3000 Hz, and for rounded vowels, where $F3$ may be as low as 2200 Hz, $F3$ for vowels tends to remain within a 2300–2500 Hz band (Ladefoged, 1982; Zue, 1985). Individual speakers may vary proportionately; subjects BS and RD in the present study, for instance, had $F3$’s of 2700–3000 Hz during /a/, and trajectory ‘‘bends’’ some 300–500 Hz lower. Apart from /r/, obstruent-type constriction at specific points on the palate, in the pharynx, constriction at the lips, and/or extension of the vocal tract lengthwise such as occurs during lip protrusion, will lower $F3$, but probably by no more than 200–300 Hz (Kewley-Port, 1982; Espy-Wilson and Boyce, unpublished modeling study).

[2] In situations where there was no clear acoustic landmark separating a semivowel from an adjacent vowel, the heuristic rule applied for segmentation was to assign 2/3 of the vowel and semivowel region to the vowel and the remainder to the semivowel.

[3] HSS and MS had participated in an earlier pilot study of four subjects whose results resemble those produced here. At that time, neither were aware of the purpose of the study. BS was a researcher associated with the study.

[4] A three-factor analysis of variance using tokens, in which speaker- and item-specific characteristics were treated as uncorrelated variables, yielded a similar pattern of results. The one exception was a significant difference between singleton /r/ and /Cr/ words. This effect was traced to an interaction between context and subject, whereby speakers WJ and HD showed longer measured trajectories in /Cr/ words (see Sec. II A 2 for a possible explanation). This effect was not significant in the repeated measures analysis, which takes account of intraspeaker correlations.

[5] As with the duration measures, a three-factor analysis of variance using tokens, in which speaker- and item-specific characteristics were treated as uncorrelated variables, yielded a similar pattern of results.

[6] Frequency resolution for the ESPS/WAVES formant tracker in the $F3$ frequency band was empirically determined using synthesized /ɑ/ and /r/. When an all-pole model was assumed, frequency was matched to within 2 to 3 Hz. When one antiresonance was included in the four-formant model, frequency was matched to within 55 Hz. Since whether the acoustics of /r/ includes an antiresonance is not known, we used 10 Hz as a conservative compromise.

Alwan, A., Narayanan, S., and Haker, K. (**1997**). ‘‘Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II: The rhotics,’’ J. Acoust. Soc. Am. **101**, 1078–1089.

Bell-Berti, F., and Krakow, R. (**1991**). ‘‘Anticipatory velar lowering: A coproduction account,’’ J. Acoust. Soc. Am. **90,** 112–123.

Bell-Berti, F., Krakow, R. A., Gelfer, C. E., and Boyce, S. E. (**1995**). ‘‘Anticipatory and carryover effects: Implication for models of speech production,’’ *Producing Speech: A Festschrift in honor of Katherine Safford Harris, 77–97,* edited by F. Bell-Berti and L. Raphael (American Institute of Physics, Woodbury, NY).

Bernthal, J., and Bankson, N. (**1993**). *Articulation and Phonological Disorders* (Prentice-Hall, Englewood Cliffs, NJ), pp. 5–60.

Boyce, S. E., Krakow, R. A., Bell-Berti, F., and Gelfer, C. (**1990**). ''Converging sources of evidence for dissecting articulation into core gestures,'' J. Phon. **18**, 173–188.

Bronstein, A. (**1967**). *Your Speech and Voice* (Random House, New York).

Browman, C. P., and Goldstein, L. M. (**1986**). ''Towards an articulatory phonology,'' Phonology Yearbook **3**, 215–252.

Browman, C. P., and Goldstein, L. M. (**1990**). ''Gestural specification using dynamically-defined articulatory structures,'' J. Phon. **18**, 299–320.

Daniloff, R., and Moll, K. (**1968**). ''Coarticulation of lip-rounding,'' J. Speech Hear. Res. **11,** 707–721.

Delattre, P. (**1967**). ''Acoustic or articulatory invariance?,'' Glossa, **1**, 3–25.

Delattre, P., and Freeman, D. (**1968**). ''A dialect study of American r's by x-ray motion picture,'' Language **44**, 29–68.

Espy-Wilson, C. Y. (**1987**). ''An acoustic-phonetic approach to speech recognition: Application to the semivowels,'' Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Espy-Wilson, C. Y. (**1992**). ''Acoustic measures for linguistic features distinguishing the semivowels in American English,'' J. Acoust. Soc. Am. **92,** 736–757.

Espy-Wilson, C. Y. (**1994**). ''A feature-based semivowel recognition system,'' J. Acoust. Soc. Am. **96,** 65–72.

Espy-Wilson, C., and Boyce, S. (**1994**). ''Acoustic differences between ''bunched'' and ''retroflex'' variants of American English /r/,'' J. Acoust. Soc. Am. **95,** 2823(A).

Espy-Wilson, C., and Boyce, S. (unpublished).

Fowler, C. (**1993**). ''Phonological and Articulatory Characteristics of Spoken Language,'' in *Linguistic Disorders and Pathologies: An International Handbook*, edited by G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, and C.-W. Wallesch (Walter de Gruyter, New York), pp. 34–46.

Gelfer, C., Bell-Berti, F., and Harris, K. (**1989**). ''Determining the extent of coarticulation: Effects of experimental design,'' J. Acoust. Soc. Am. **86,** 2443–2445.

Giergerich, H. (**1992**). *English Phonology: An Introduction* (Cambridge U.P., Cambridge, England).

Gracco, V., and Lofqvist, A. (**1994**). ''Speech motor coordination and control: Evidence from lip, jaw and laryngeal movements,'' J. Neurosci. **14**(11), 6585–6597.

Hagiwara, R. (**1995**). ''Acoustic realizations of American English /r/ as produced by women and men,'' UCLA Phonetics Laboratory Working Papers **90**, 55–61.

Hammarberg, R. (**1976**). ''The metaphysics of coarticulation,'' J. Phon. **4**, 353–363.

Harris, K., and Bell-Berti, F. (**1984**). ''On consonants and syllable boundaries,'' in *Language and Cognition*, edited by L. J. Raphael and M. R. Valdovinos (Plenum, New York).

Keating, P. (**1988**). ''Underspecification in phonetics,'' Phonology **5**, 275–292.

Kent, R., and Minifie, F. (**1977**). ''Coarticulation in recent speech production models,'' J. Phon. **5**, 115–133.

Kewley-Port, D. (**1982**). ''Measurement of formant transitions in naturally produced stop consonant-vowel syllables,'' J. Acoust. Soc. Am. **73,** 379–389.

Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich, San Diego).

Lehiste, I. (**1962**). ''Acoustical characteristics of selected English consonants,'' University of Michigan Communication Sciences Laboratory Report No. 9.

Lehiste, I., and Peterson, G. E. (**1961**). ''Transitions, glides, and diphthongs,'' J. Acoust. Soc. Am. **33,** 268–277.

Lindau, M. (**1985**). ''The story of /r/,'' in *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, edited by V. Fromkin (Academic, Orlando).

Munhall, K., and Lofqvist, A. (**1992**). ''Gestural aggregation in speech: Laryngeal gestures,'' J. Phon. **20**, 111–126.

Nolan, F. (**1983**). *The Phonetic Bases of Speaker Recognition* (Cambridge U.P., Cambridge, England).

Olive, J., Greenwood, A., and Coleman, J. (**1993**). *Acoustics of American English Speech* (Springer-Verlag, New York).

Perkell, J. S., and Matthies, M. L. (**1992**). ''Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability,'' J. Acoust. Soc. Am. **91,** 2911–2925.

Perkell, J. S., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). ''Electro-Magnetic Midsagittal Articulometer (EMMA) systems for transducing speech articulatory movements,'' J. Acoust. Soc. Am. **92,** 3078–3096.

Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24,** 175–184.

Recasens, D. (**1985**). ''Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences,'' Language Speech **28**, 97–114.

Seneff, S., and Zue, V. (**1988**). ''Transcription and alignment of the TIMIT database,'' documentation distributed with the TIMIT database by NBS.

Shoup, J., and Pfeifer, L. (**1976**). ''Acoustic characteristics of speech sounds,'' in *Contemporary Issues in Experimental Phonetics*, edited by N. Lass (Academic, New York).

Westbury, J. M., Hashi, M., and Lindstrom, M. J. (**1995**). ''Differences among speakers in articulation of American English /r/: An x-ray microbeam study,'' in Proceedings of the XIIIth International Conference on Phonetic Sciences, 1995, Vol. 4, 50–57.

Zue, V. (**1985**). Speech Spectrogram Reading, Summer Course, MIT.