SCHWA IN AMERICAN ENGLISH V+r/ SEQUENCES

MARÍA RIERA, JOAQUÍN ROMERO & BEN PARRELL

Universitat Rovira i Virgili

Abstract
This paper presents an acoustic study of word-final V+r/ sequences in American English. The objectives were (i) to identify the presence of a schwa-like element in the VC transitions, (ii) to investigate how this presence is related to the phonetic/phonological nature of V, and (iii) to determine whether the spectral and durational characteristics of this element vary as a function of speaking rate. Two speakers participated in the experiment. Formant and duration measurements accounted for (i) differences between the schwa-like element and canonical schwa and (ii) variability in the schwa-like element and V. One-way ANOVAs tested for formant and duration differences, while two-way ANOVAs tested for the relationship between formant variability in the schwa-like element and V. The results suggest that the presence of a highly variable schwa-like element in the V+r/ sequences is (i) a generalized process affecting all contexts and (ii) the result of coarticulation rather than epenthesis/insertion.

1. Introduction
In this study we investigate sequences of V+r/ in American English, in particular the schwa-like element that is often perceived between the two segments in these sequences. We are interested in finding out whether the presence of this schwa-like element is due to a process of epenthesis/insertion or to one of coarticulation.

Trask (1996:132) defines epenthesis as a phonological process consisting of “the insertion of a segment into a word in a position in which no segment was previously present.” Coarticulation is defined by Hammarberg (1976:357) as “a process whereby the properties of a segment are altered due to the

* This research was supported by a doctoral grant from the Universitat Rovira i Virgili (Tarragona, Spain) and by projects 2005SGR00864 of the Generalitat de Catalunya and HUM2005-02746 of the Ministerio de Educación y Ciencia (Spain). The authors would like to thank Maria-Josep Solé and two anonymous reviewers for their very useful comments and suggestions on an earlier version of this paper.

1 Currently at University of Southern California
influences exerted on it by neighboring segments.” Kühnert and Nolan (1999:7) state that coarticulation “refers to the fact that a phonological segment is not realized identically in all environments but often apparently varies to become more like an adjacent or nearby segment.” Finally, Ladefoged (2001:247) defines it as “the overlapping of adjacent articulatory gestures.”

Coarticulation is considered a phonetic process, its phonological counterpart being known as assimilation (Hammarberg 1976, 1982; Roach et al. 2006), though the distinction and the relationship between the two remain an issue of debate among scholars. In fact, authors like Hammarberg (1976, 1982) and Kühnert and Nolan (1999) consider it necessary to take both the phonetic and the phonological levels into account when studying coarticulation. According to these authors, in any instance of coarticulation, behind the phonetic representation of the allophone (a physical tangible unit) there is always the phonological representation of the phoneme (a mental abstract unit).

Along these lines, a less clear-cut distinction between the phonetic and phonological levels can be taken, however, as illustrated by the Articulatory Phonology approach (Browman & Goldstein 1986, 1989, 1990a, 1990b, 1992a, 1992b). Within this framework, articulatory gestures are simultaneously events of physical (traditionally phonetic) activity as well as units of linguistic (traditionally phonological) organization. Thus, coarticulation and assimilation could be viewed as instances of the same general process of gestural overlap and reduction. Language-dependent spatial and phasing relationships between gestures, as well as prosodic and contextual factors, would determine the extent to which a specific process remains at the level of colloquial, casual speech and is, therefore, under the control of the speaker, or rather, it becomes part of the lexical representation and is, supposedly, no longer affected by surface variability in aspects like speech rate. It is unclear, however, where the division between these two levels lies, as evidenced by much recent work on the phonetics/phonology interface. Thus, though for the sake of simplicity and convenience we will use the coarticulation vs. epenthesis/insertion dichotomy in this study, it is not our intention to imply a categorical distinction between the two or to present the study as a simple correlation between the phonetic and the phonological levels. Instead, we believe the two phenomena to be instances of the same generalized process of gestural conflict resolution, as an indication of the need to further our knowledge of the interaction between the two traditional levels.

In American English, a process of vowel neutralization takes place in the context of vowel sounds followed by /r/ (Avery & Ehrlich 1992; Giegerich 1992; Ladefoged 2001). The vowel contrast that is present in many other contexts between five pairs of vowel sounds (/i/ vs. /ɪ/, /e/ vs. /ɛ/, /a/ vs. /æ/, /o/ vs. /ɔ/ and /u/ vs. /u/) disappears in this context, with the resulting combinations of vowel+/r/ being /ɪr/, /ɛr/, /ɑr/, /ɔr/ and /ɜr/. Many authors (Avery & Ehrlich 1992; Baker & Goldstein 1990; Calvert 1986; Clark &
Hillenbrand 2003; Dauer 1993; Giegerich 1992; Ladefoged 2001) claim that, in
the existing V+/t/ combinations, the exact realization of the vowel varies rather
substantially between the two members of each contrasting pair. Thus, in the
case of /it/, the exact realization of the vowel could vary between /i/ and /i/,
based on speaker and/or dialectal differences. The same applies to the other
pairs.

Vowels before /t/ are often called rhotacized, retroflexed or r-colored as a
result of the influence exerted on them by the following /t/ (Avery & Ehrlich
1992; Clark & Yallop 1995; Davenport & Hannahs 1998; Ladefoged 2001;
Olive et al. 1993; Rogers 2000). In these rhotacized vowels, the basic tongue
configuration is retained, but the tongue tip is curled back in anticipation of the
/t/ (Clark & Yallop 1995; Ladefoged 2001; Olive et al. 1993; Rogers 2000).
The rhotacized effect may also be produced by keeping the tongue tip down
and bunching the tongue body upwards towards the roof of the mouth
may be in-between positions” to produce the rhotacized effect on the vowel
and that in all cases “there is a slight narrowing of the pharyngeal cavity.”
Rhotacized vowels before /t/ tend to show a considerable gradual and slow
lowering of the third formant as a result of the influence of the low-frequency
third formant of /t/ (Ladefoged 2001; Olive et al. 1993; Rogers 2000).

Sequences of V+/t/ do not occur syllable-finally in non-rhotic varieties of
English\(^2\). Instead, these varieties have the centering diphthongs /aʊ/, /eʊ/, and
/uə/, as well as the centering triphthongs /aʊə/, /eʊə/, /ʊə/, /aʊə/ and /uəə/. The
development of the /aʊ/ in all these cases is understood as the result of the
historic loss of the /t/ (Giegerich 1992; Ladefoged 2001; Laver 1994; Rogers
2000). However, the equivalence between the diphthongs and triphthongs of the
non-rhotic varieties and the V+/t/ sequences of the rhotic ones is not
straightforward. In a rhotic variety such as General American, even though /aʊ/,
/eaʊ/, /əʊ/, /aʊə/ and /uəə/ have their equivalents in /ir/, /er/, /ør/, /aʊr/ and /uər/,
the three triphthongs /eʊə/, /ʊə/ and /uəə/ have no V+/t/ counterpart.
According to standard descriptions (Kenyon & Knott 1953; Edwards 1997),
diphthongs in General American are limited to /aʊ/, /aʊ/ and /uə/, while /e/ and
/o/ are considered diphthongized monophthongs; in addition, General
American is described as having no triphthongs.

In his English pronunciation dictionary, Wells (2000) uses the superscript
symbol /r/ to show instances of schwa epenthesis/insertion (/t̚r/ /e̜r/, /u̜r/,
/aɪ̜r/ and /aʊ̜r/, as in here, hair, poor, fire and power). He refers to these
cases as examples of what he terms pre-r breaking. According to him, vowels
and diphthongs undergo pre-r breaking as the result of the development of a
glide sound /aʊ/ before /t/ and, consequently, become diphthongs and
triphthongs, respectively. The diacritic /ː/ that is present in the transcriptions

---

\(^2\) Except in the case of linking /t/ (here is fine, far and away, poor or rich).
/æ/ and /ə/ reflects the possibility of considering these words as composed of either one syllable or two. Similarly, Lavoie and Cohn (1999) state that monosyllables consisting of non-low tense pure vowels or diphthongs followed by a liquid, which they call *sesquisyllables*, can be pronounced with either one or two syllables. Finally, Roach et al. (2006:144) state that “opinions differ as to whether diphthongs should be treated as phonemes in their own right, or as combinations of two phonemes.”

In all the above-mentioned cases, Wells’s transcription system accounts for two possible pronunciations: one with schwa epenthesis/insertion and the other without it. According to him, the choice of one or the other is speaker-dependent as well as situation-dependent, with schwa epenthesis/insertion being more common in slow speech rates than in fast ones. As Wells (2000) points out, in his dictionary the superscript symbol /ə/ is also intended to help second and/or foreign language learners of English in their pronunciation. Other authors (Baker & Goldstein 1990; Calvert 1986; Celce-Murcia et al. 1996; Dauer 1993; Prator & Robinett 1985) also make a point of the inclusion of epenthetic schwa in the transcription of the V+/ə/ sequences as a way for the second and/or foreign learner of English to associate the phonological representation of the sequences with how they are both produced and perceived.

Some authors within the American tradition make use of the term *centering diphthong* to refer to either the vowel in the V+/ə/ sequences or the whole sequence. In their study, Clark and Hillenbrand (2003:1) define the combinations of V+/ə/ as diphthongs and choose “to use [ə] as the second part of the diphthong in question, based on the assumption that their first part is prominent and the second is an offglide to a relatively weaker centralized endpoint.” Dauer (1993) calls the V+/ə/ sequences *centering diphthongs* and uses a transcription system that is consistent with the term. Her phonological transcriptions include epenthetic schwa in all cases. Van Riper and Smith (1992) also refer to these sequences as *centering diphthongs*, but they present the transcription with /ə/ only as a secondary option. Kenyon (1989) also uses the term *centering diphthongs* to name the V+/ə/ sequences, but on no occasion does he suggest transcribing the sequences with epenthetic schwa, transcribing them instead with vowel+/ə/. Roach et al. (2006)³ transcribe the *centering diphthongs in here, hair and poor* without epenthetic schwa and with vowel+/ə/ as well, but they transcribe *fire* and *power* with final /ə/, and they do not consider the words *far* and *pour* to contain any *centering diphthong*. Prator and Robinett (1985) advocate for the term *centering diphthongs* as well, but their phonological representation of the sequences is with final /ə/ in all cases and

with epenthetic schwa included in /ɜər/ (here), /ɛər/ (hair), /ʌər/ (fire) and /aʊər/ (power), but not in /ɑr/ (far), /ɔr/ (pour) and /ɜr/ (pour). For them, the vocalic elements are the ones that make up the centering diphthong and the /r/ is not part of it.

Other authors (Baker & Goldstein 1990; Celcê-Murcia et al. 1996) do not talk of centering diphthongs, but they suggest epenthetic schwa as an option in at least some cases. Baker and Goldstein (1990), for instance, do so with /ɪər/ (here), /ɛər/ (hair), /ʌər/ (poor), /ɑər/ (fire) and /aʊər/ (power), not considering it possible in /ɑr/ (far) or /ɔr/ (pour). Celcê-Murcia et al. (1996:104) associate epenthetic schwa directly only with the diphthongs /aɪə/ (fire) and /aʊə/ (power), but indirectly with at least one other case when they state that the conventional transcription of the V+/r/ sequence in beard as /ɪər/ “may not completely represent or capture the precise articulation of the /r/-colored vowel.”

The schwa-like element that is often perceived between the two elements of some VC and CC sequences has received different names by different authors. Warner et al. (2001), in their analysis of Dutch schwa in /l+/C clusters, refer to it as epenthetic schwa. Authors such as Baker and Goldstein (1990), Celcê-Murcia (1996), Olive et al., (1993), Prator and Robinett (1985) and Wells (2000), despite not using the term epenthesis, refer to the process as one of insertion, a term that is considered synonymous with epenthesis. Hall (2003, 2006) makes a clear distinction between schwa intrusion and schwa epenthesis/insertion, claiming that the schwa-like element in the CC clusters of her studies belongs to the former type. According to this author, intrusive vowels are phonologically invisible, are inserted late in the phonological derivation, cannot act like syllable nuclei, do not add a syllable to the word and do not involve the addition of a vowel segment. Moreover, they are not likely to occur in the most marked types of CC clusters, tend to occur between heterorganic consonants, copy only over sonorants or gutturals and are either copy vowels or neutral and schwa-like in quality. Finally, they come in a restricted range of qualities, are often variable in duration and have a tendency to disappear in fast and/casual speech. Schmeiser (this volume) also provides justification for distinguishing between intrusive vowels and epenthetic vowels. Using data from Spanish, he claims that the result of diachronic phonologization of intrusive vowels is predictable, given the similarity in formant structure between the intrusive vowel and the nuclear vowel. Also, intrusive vowels are not specified lexically, as evidenced by syllable separation. In addition, the restrictions imposed by the three-syllable stress window of Spanish suggest the non-epenthetic nature of these vowels. Browman and Goldstein (1992b), in their analysis of CVCCVCV sequences, refer to this element as a targetless schwa, that is, a vocalic element with spectral and duration values similar to those that are often attributed to canonical schwa but, at the same time, somehow different from them, due
mainly to the influence exerted by neighboring vocalic segments on the vocalic element. Finally, Gick and Wilson (2001, 2006) call the schwa-like element in V+liquid combinations _exrescent schwa_ and consider it the result of the tongue movement required to pass through a schwa-like configuration on its way from the vowel to the /t/. Explanations along the same lines to account for the presence of a schwa-like element in American English V+/t/ sequences are provided by other authors like Calvert (1986), Prator and Robinett (1985) and Olive et al. (1993).

In a preliminary descriptive acoustic study, Riera and Romero (2006) investigated the presence of a schwa-like element in word-final V+/l/ and V+/t/ sequences in American English stressed monosyllables. The results showed that the VC transitions in these sequences are in some cases not easy to identify by means of visual spectrographic observation. The results also suggested a relationship between the phonological parameters that are used to classify vowels (tongue height, tongue advancement and the tense vs. lax distinction), as well as the clear presence in sequences containing high front tense vowels of a quite variable schwa-like element in terms of its duration and spectral characteristics. This study, however, did not provide acoustic measurements and was not able to account for speaking rate differences.

In a follow-up study, Riera and Romero (2007) looked at the phonetic and phonological nature of word-final V+/l/ sequences in American English stressed monosyllables as produced by only one speaker. The analysis involved acoustic (F1, F2 and duration) measurements. The results suggested the presence of a highly variable schwa-like element which differed from canonical schwa as a function of both the preceding vowel and speaking rate. This study concluded that the phenomenon under analysis was a generalized one affecting all V+/l/ contexts and that the presence of the schwa-like element was due to a process of coarticulation rather than to one of epenthesis/insertion.

Taking the results of previous studies into consideration, we designed an experimental acoustic study which expanded on those of Riera and Romero (2006, 2007) and replicated in part that of Riera and Romero (2007) in an attempt to provide further evidence for the phonetic and phonological nature of word-final V+/t/ sequences in American English stressed monosyllables.

1.1 _Objectives and hypotheses_

The first of our main objectives in designing the present study was threefold, since our intentions were (i) to discover, by means of acoustic measurements, which of the V+/t/ sequences under study contain a distinguishable schwa-like element in their VC transitions, (ii) to see whether our findings are consistent with those of previous work, and (iii) to determine whether the presence of this schwa-like element can be extended to VC transitions containing vowels other than high, front and tense. We hypothesize that acoustic measurements will make it possible to account for the presence of
a schwa-like element in sequences in which mere spectrographic observation renders it undetectable as well as in sequences not dealt with in previous studies. If this is the case, we will find ourselves in front of a generalized process affecting all contexts, in line with the results obtained in Riera and Romero (2007), rather than a process affecting just some or a few contexts, as seems to be suggested by the focus and/or the results of some previous studies.

Our second main objective was to investigate the extent to which the presence of this schwa-like element is related to the phonetic/phonological nature of the preceding vowel and whether the spectral and durational characteristics of the schwa-like element vary as a function of speaking rate. Two hypotheses stem from this objective. First, we hypothesize that F1, F2, F3 and duration values for the schwa-like element in the VC transitions will differ from those of canonical schwa. F1, F2 and F3 will show greater variability in the schwa-like element than in canonical schwa. Also, duration will be shorter in the former than in the latter. This will be taken as evidence that this element cannot be considered an epenthetic schwa. Second, we hypothesize that F1, F2 and F3 values for the schwa-like element will vary considerably across the different contexts and also as a function of the preceding vowel, resembling more the values of this preceding vowel the faster the speaking rate. If one or both of these hypotheses are confirmed, we will be able to claim that any presence of a schwa-like element between the vowel and the consonant in the V+/r/ sequences under study, far from being the result of a discrete process of epenthesis/insertion, can be attributed to a dynamic process of coarticulation.

2. Method
2.1 Subjects
The subjects that participated in the experiment were two native speakers of the Midwestern variety of American English. Speaker 1 was a 23-year-old male who had been born and had lived most of his life in the Chicago, Illinois, area. He was a graduate student in phonetics and had some specialized phonetic training. Speaker 2 was a 50-year-old female who had lived in different parts of the United States but who spoke with a clearly defined Midwest accent. She had no specialized phonetic training. Both speakers were unaware of the purposes of the experiment at the time of the recording.

2.2 Stimuli and data collection
The stimuli that were selected for the experiment consisted of seven meaningful English monosyllables containing the sequences C1VC2. The VC2 sequences consisted of /r/ preceded by each of the seven vowel sounds of American English that can appear before this consonant, namely, /i/, /ε/, /u/, /u/, /u/ and /u/. The choice of non-lingual (unlike /r/) and oral (like /r/) consonants was considered most appropriate and was made for the purpose of minimizing C1 coarticulatory influence on V, and even
on coda $C_2$. Seven additional monosyllables were chosen as distracters. In these, $C_1$ was /t/, /v/ or /h/, again non-lingual and oral consonants; V was once more one of the seven vowel sounds that can appear before /t/ in American English; and $C_2$ was either /t/ or /d/. Finally, the words arrive and ahead, both containing canonical schwa in their initial syllable, were also included, the former to serve as control and the latter as distracter. Table 1 shows the 16 stimuli (target words, control word and distracters) used for the experiment.

<table>
<thead>
<tr>
<th>Word</th>
<th>$C_1$, $C_2$, V</th>
<th>Distracters</th>
</tr>
</thead>
<tbody>
<tr>
<td>here</td>
<td>/hir/ /ir/ /i/</td>
<td>heat /hit/ /it/ /i/</td>
</tr>
<tr>
<td>hair</td>
<td>/her/ /er/ /e/</td>
<td>vet /vet/ /et/ /e/</td>
</tr>
<tr>
<td>far</td>
<td>/far/ /ar/ /a/</td>
<td>hot /hat/ /at/ /a/</td>
</tr>
<tr>
<td>pour</td>
<td>/por/ /or/ /o/</td>
<td>fought /fot/ /ot/ /o/</td>
</tr>
<tr>
<td>poor</td>
<td>/pur/ /ur/ /u/</td>
<td>hood /hud/ /ud/ /u/</td>
</tr>
<tr>
<td>fire</td>
<td>/fair/ /air/ /a/</td>
<td>hide /haid/ /aid/ /au/</td>
</tr>
<tr>
<td>power</td>
<td>/paur/ /aur/ /au/</td>
<td>vowed /vaud/ /aud/ /au/</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>arrive</th>
<th>/'arv/ /'ar/ /a/</th>
<th>ahead /'hed/ /'h/ /a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>$V_1C_1V_2C_2$</td>
<td>Word</td>
</tr>
<tr>
<td></td>
<td>$V_1C_1$</td>
<td>$V_1C_1$</td>
</tr>
<tr>
<td></td>
<td>$V_1$</td>
<td>$V_1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control word</th>
<th>Distracter</th>
</tr>
</thead>
</table>

Table 1: Stimuli (target words, control word and distracters) used for the experiment

A total of 40 stimuli —the stimuli that were selected for this experiment plus the stimuli that were selected for the experiment reported in Riera and Romero (2007), which looked at $V+/l/$ sequences— were inserted in the carrier sentence ‘Tell me ____ four times.’. The tokens were presented in a timed computer Power Point slide presentation for the speakers to read. Each speaker performed two readings of 12 randomized tokens of each sentence: first, at a slow rate and, second, at a fast rate. The speaking rate variable was controlled for by presenting the tokens in the slow reading at three-second intervals, with a three-second break every 20 tokens, and the tokens in the fast reading at one-second intervals, with a three-second break every five tokens. Each speaker provided a total of 960 (40x12x2) tokens, 384 (16x12x2) of which were selected and used for the experiment reported here. The total involvement time for each speaker was about one hour and thirty minutes, including the instructions period, the two readings, a 15-20 minute break between the two readings, and two trial periods before each reading, of 20 tokens each, which were not used for the analysis.

The data were recorded at a 22,050 Hz sampling rate directly into an HP Pavilion dv4000 laptop computer using an M-Audio Nova condenser microphone, an M-Audio Firewire Solo mobile audio interface and the Praat
(version 4.5) speech analysis software (Boersma & Weenink 2006). The latter was also used for the subsequent data processing and analysis.

2.3  Data processing and analysis

The procedure used for the data processing and analysis of the sequences under study replicates the one reported in Riera and Romero (2007). First, in order to account for general duration differences between the slow and fast productions, duration measurements were taken for the entire CVC sequence and the VC portion in the V+/t/ words, as well as for the entire word and the /ɔt/ portion in the control word arrive. These measurements, however, were not used for the final analysis, since this focused on the VC sequences under study only. Second, F1, F2 and F3 values were automatically calculated every five milliseconds for both the entire CVC sequence of the V+/t/ words and the /ɔt/ portion of the control word arrive. Third, the beginning and end of any VC transitions in the V+/t/ sequences and of the /ɔ/ in the /ɔt/ sequence in arrive were established on the basis of (i) F1, F2 and F3 values, (ii) spectrographic observation, and (iii) auditory perception. Fourth, mean F1, F2, F3 and duration values of the VC transitions were extracted for each V+/t/ sequence independently and were then compared with the mean values of the canonical schwa in arrive. Fifth, F1, F2 and F3 values for V were obtained from a single middle point in the most stable part of the pure vowels and in the offglide of the diphthongs. Finally, the differences between the values for V and those of the corresponding schwa-like element were calculated for each context in order to check for variability of the schwa-like element as a function of the preceding vowel. Figure 1 shows the measurement criteria followed for the analysis of the VC transitions and canonical schwa.

A first set of statistical analyses, consisting of one-way factorial ANOVAs, was aimed at obtaining overall significant differences in F1, F2, F3 and duration between the schwa-like element and canonical schwa. Fisher’s post-hoc tests were then carried out for individual comparisons between the schwa-like element in each of the contexts and the canonical schwa in the control word arrive, with context (each of the V+/t/ sequences) as the independent variable and F1, F2, F3 and duration as the dependent variables.

A second set of statistical analyses, using two-way factorial ANOVAs, was performed in order to test for F1, F2 and F3 variability between the schwa-like element in each of the contexts and the preceding vowel. In this case, the independent variables were rate (slow vs. fast) and context (each of the V+/t/ sequences), and the dependent variables were the F1, F2 and F3 values obtained from calculating the differences in F1, F2 and F3 between the schwa-like element and the preceding vowel.
Figure 1: Example tokens illustrating the measurement criteria for VC transitions (top) and canonical schwa (bottom) corresponding to one of Speaker 1’s slow readings of the words fire and arrive.
3. Results

3.1 Schwa-like element and canonical schwa

The first set of statistical analyses, which consisted of one-way factorial ANOVAs performed to test for general differences in overall comparisons between the schwa-like element in each of the contexts and the canonical schwa in the control word arrive, revealed highly significant differences for all the dependent variables (F1, F2, F3 and duration) for both speakers. Significance level was set at p<.01. These results are provided in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>F(7,184) = 51.366, p&lt;.0001</td>
<td>F(7,184) = 44.792, p&lt;.0001</td>
</tr>
<tr>
<td>F2</td>
<td>F(7,184) = 430.990, p&lt;.0001</td>
<td>F(7,184) = 300.945, p&lt;.0001</td>
</tr>
<tr>
<td>F3</td>
<td>F(7,184) = 17.507, p&lt;.0001</td>
<td>F(7,184) = 9.123, p&lt;.0001</td>
</tr>
<tr>
<td>Duration</td>
<td>F(7,184) = 10.726, p&lt;.0001</td>
<td>F(7,184) = 3.204, p=.0032</td>
</tr>
</tbody>
</table>

Table 2: Results of one-way factorial ANOVAs for overall significant differences in F1, F2, F3 and duration

Tables 3 and 4 show the results, for Speaker 1 and Speaker 2, respectively, of the Fisher’s post-hoc tests carried out to test for particular differences in individual comparisons between the schwa-like element in each of the contexts and the canonical schwa in the control word arrive. Context (each of the V+/t/ sequences) was the independent variable, and F1, F2, F3 and duration were the dependent variables. Significance level was set at p<.01. Significant differences were found for duration and, more strikingly so, for F2 in all contexts for both speakers, thus showing that the schwa-like element in the sequences under study differs from the canonical schwa in arrive as regards F2 and duration. The results for both speakers concerning F1 and F3 differences between the schwa-like element and canonical schwa were significant only in some cases. In addition, no clear relationship can be established between the results for both speakers. For Speaker 1, F1 showed significant differences in the case of here, far and fire, but not in the case of hair, pour, poor and power. For this speaker, F3 showed significant differences for hair, pour, poor and power, but not for here, far and fire. For Speaker 2, F1 yielded significant differences for hair, far, fire and power, but not for here, pour and poor. F3 for this speaker yielded non-significant differences in all cases except for pour.

---

4 We thank Maria-Josep Solé and one anonymous reviewer for pointing out that, because the V+/t/ sequences are tautosyllabic while the canonical schwa+/t/ combination in arrive is heterosyllabic, duration differences between the two schwa types may be influenced by syllable affiliation. We are currently carrying out an experiment that takes this issue into consideration.
3.2 Schwa-like element and preceding vowel

Figures 2, 3 and 4 show distribution graphs for F1, F2 and F3, respectively, of the variability between the schwa-like element in each of the contexts (V+/t/ sequences) and the preceding vowels. In the three figures, the top graphs display the results for the slow tokens while the bottom graphs display the results for the fast ones. Similarly, the left graphs reveal the results for Speaker 1 while the right graphs reveal the results for Speaker 2. In addition, the squares stand for the mean formant values of the schwa-like element in the VC transitions, the triangles represent the mean formant values of the vowel in the CVC sequences, and the solid horizontal line depicts the mean formant values of the canonical schwa in the control word arrive.

Overall, the graphs in Figures 2, 3 and 4 show similar results for both speakers. First, F1, F2 and F3 values for the schwa-like element in the VC
transitions are highly variable. Second, these values are also systematically
different from those of the canonical schwa in arrive. Third, these values tend
to resemble those of the preceding vowel. Finally, there is a higher degree of
variability the faster the rate, that is, there is a smaller distance between F1, F2
and F3 values of the schwa-like element and those of the corresponding
preceding value in the fast tokens than in the slow ones. Only a few tokens
depart from this tendency. For Speaker 1, the distance between the values of
the schwa-like element and of the preceding vowel is smaller in the slow
tokens than in the fast ones in the case of hair, far, pour, fire and power (F1)
and here and far (F3), and exactly the same in the case of here (F2). As regards
Speaker 2, the distance is shorter for here (F1) and far (F2), and exactly the
same for fire (F1). Once more, no clear relationship can be established between
the results for both speakers.

The results of the second set of statistical analyses show significant
effects in almost all cases and seem to corroborate the results shown by the
data in the graphs of Figures 2, 3 and 4. In this case, two-way factorial
ANOVAs testing for F1, F2 and F3 variability between the schwa-like element
in each of the contexts and the preceding vowel were performed, with rate and
context as the independent variables and with the F1, F2 and F3 values
obtained from calculating the differences in F1, F2 and F3 between the schwa-
like element and the preceding vowel as the dependent variable. Significance
level was set at p<.01. The results are provided in Table 5.

4. Discussion and conclusions

The purpose of this study was to investigate the phonetic/phonological
nature of word-final V+/r/ sequences in American English stressed
monosyllables. We designed an acoustic experiment in an attempt to answer
two main questions: (i) is there a schwa-like element in these V+/r/ sequences?
and (ii) if so, is this schwa-like element the result of a process of
epenthesis/insertion or of a process of coarticulation? So as to prove that the
latter is the case, we formulated two more questions: (i) does this schwa-like
element resemble canonical schwa? and (ii) does this schwa-like element vary
systematically as a function of the vocalic context? The results of our study
have provided answers to all these questions. Our findings corroborate those
reported in a previous study that focused on V+/l/ sequences (Riera & Romero,
2007). Moreover, they expand on earlier studies that were preliminary in nature
or that did not consider the full range of stressed vowels and diphthongs of
American English in the treatment of V+/r/ sequences. Finally, our work is
innovative in that it introduces the speaking rate variable to help explain the
phenomenon under analysis.

Acoustic measurements have made it possible to determine that a vocalic
schwa-like element can be distinguished, to a greater or lesser extent, in the VC
transitions of all of the V+/r/ sequences under study. The presence of this
element in all of the sequences leads us to conclude that we are dealing with a generalized process affecting all V+/r/ contexts, rather than a specific one concerning only a few, such as those involving high and/or front and/or tense vowels, which had been the object of previous research.

![Graphs showing F1 values for different vowels and speakers in slow and fast conditions.](image)

**Figure 2:** Distribution graphs for F1 values, in slow (top) and fast (bottom) tokens, for Speaker 1 (left) and Speaker 2 (right), with frequency shown in Hertz on the vertical axis and each of the contexts on the horizontal axis.

The results obtained by the one-way ANOVAs, which looked at overall comparisons between the schwa-like element and canonical schwa, yielded highly significant differences for both speakers. The results of the subsequent Fisher’s post-hoc tests, which focused on individual comparisons, despite not being significant for all contexts, corroborated the existence of these
differences for most cases and for both speakers. The fact that this schwa-like element differs from canonical schwa in its F2 and duration values in all contexts and for both speakers provides evidence to conclude that this element is different enough from canonical schwa for it not to be considered an epenthetic schwa. In a process of schwa epenthesis/insertion, the formant and duration values of this element would be expected to be more similar to those of canonical schwa than our results have proven it to be. Thus, we attribute the presence of this element to a process of coarticulation.

Figure 3: Distribution graphs for F2 values, in slow (top) and fast (bottom) tokens, for Speaker 1 (left) and Speaker 2 (right), with frequency shown in Hertz on the vertical axis and each of the contexts on the horizontal axis.
The analysis of the data that appear in the graphs of Figures 2, 3 and 4, as well as the subsequent two-way ANOVAs, were performed in order to explore the degree of F1, F2 and F3 variability between the schwa-like element in each of the contexts and the preceding vowel. The results obtained provide further evidence to confirm our hypothesis favoring coarticulation versus epenthesis/insertion. First, it is possible to detect the existence of a highly variable schwa-like element in terms of its spectral and duration values, on the one hand, and as a function of the preceding vowel, on the other. Second, this schwa-like element once more appears to be systematically different from canonical schwa. And third, the values of this schwa-like element tend to
resemble those of its preceding vowel, with an increase in variability, the faster the speaking rate. These observations, despite exceptions and the lack of significant results in some cases, apply to the great majority of contexts in our study.

<table>
<thead>
<tr>
<th>Rate</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>F(1,154) = 0.438, p=.5091</td>
<td>F(1,154) = 14.689, p=.0002</td>
</tr>
<tr>
<td>Context</td>
<td>F(6,154) = 9.297, p&lt;.0001</td>
<td>F(6,154) = 10.108, p&lt;.0001</td>
</tr>
<tr>
<td>Rate*Context</td>
<td>F(6,154) = 0.850, p=.5333</td>
<td>F(6,154) = 3.509, p=.0028</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rate</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2</td>
<td>F(1,154) = 52.928, p&lt;.0001</td>
<td>F(1,154) = 92.267, p&lt;.0001</td>
</tr>
<tr>
<td>Context</td>
<td>F(6,154) = 130.146, p&lt;.0001</td>
<td>F(6,154) = 90.806, p&lt;.0001</td>
</tr>
<tr>
<td>Rate*Context</td>
<td>F(6,154) = 6.762, p&lt;.0001</td>
<td>F(6,154) = 11.569, p&lt;.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rate</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>F3</td>
<td>F(1,154) = 6.887, p=.0096</td>
<td>F(1,154) = 202.374, p&lt;.0001</td>
</tr>
<tr>
<td>Context</td>
<td>F(6,154) = 48.183, p&lt;.0001</td>
<td>F(6,154) = 6.692, p&lt;.0001</td>
</tr>
<tr>
<td>Rate*Context</td>
<td>F(6,154) = 6.373, p&lt;.0001</td>
<td>F(6,154) = 2.993, p=.0379</td>
</tr>
</tbody>
</table>

Table 5: Results of two-way factorial ANOVAs for F1, F2 and F3 variability between the schwa-like element in each of the contexts (V+/r/ sequences) and the preceding vowel, with rate and context as the independent variables and with F1, F2 and F3 values obtained from the differences between the schwa-like element and the preceding vowel as the dependent variables

All in all, our results provide strong support for considering the schwa-like element in V+/r/ sequences as the result of a dynamic process of coarticulation rather than of a discrete process of epenthesis/insertion. A process of coarticulation requires two or more sounds to become in contact with each other and to influence one another, either in a progressive manner (from left to right) or in a regressive manner (from right to left). Consequently, a process of coarticulation implies as well that, far from being produced independently from each other and one after another, neighboring sounds overlap, blend and even, in extreme cases, disappear. Our understanding of the processes under analysis is more in accordance with this continuous nature of speech production than with any other involving a simple categorical process, such as that of vowel epenthesis/insertion. In addition, considering this issue from the point of view of articulation dynamics allows for a convincing explanation of the behavior of the schwa-like element due to speaking rate differences. An increase in speech rate entails a decrease in time for articulatory gestures to attain their targets and, thus, the faster the rate is, the more overlap and blending there is in the transitions and the more difficult it is to predict exactly what these will look like.

Our conclusions lead us to claim that the term *epenthetic schwa* is not appropriate to define this schwa-like element. Despite resembling schwa, this element also has spectral values similar to those of its preceding vowel, to the point of being turned into a different vocalic element in each of the different contexts, as a function of the vowel that precedes it. Other terms such as
targetless schwa (Browman & Goldstein 1992b), excrecent schwa (Gick & Wilson 2001, 2006) or intrusive schwa (Hall 2003, 2006) seem more appropriate, since they imply the existence of an element with differing spectral and duration values according to context. Moreover, a phonological/phonemic representation of this schwa-like element, i.e., /ə/ or /ə/ , such as the one present in some pronunciation dictionaries and manuals, aimed mainly at aiding foreign speakers’ English pronunciation, does not seem very adequate. In our view, a representation of this element would only be appropriate as an aid in second/foreign language pronunciation learning.

The present study has some limitations that could be overcome in further studies so as to lead to more consistent and reliable results as well as to a better understanding of the processes involved in the VC transitions of the V+/r sequences studied here. First, the number of participants should be increased, and speakers of different varieties of American English should be included. Second, a perception study would complement the production studies so far undertaken. Third, an analysis of the spectral and duration values of /r/ would reveal the extent to which this coda consonant influences the presence and magnitude of the schwa-like element, at the same time that it would show how this /r/ is influenced by the variability of the preceding schwa-like element. Fourth, an attempt to classify the preceding vowels, in terms of phonological parameters such as tongue height, tongue advancement, lip rounding and the tense-lax distinction, and according to the behavior of the schwa-like element, would provide a more accurate picture of the phenomenon. Fifth, canonical schwa should be studied and contrasted with the schwa-like element in greater and more varied contexts. Finally, no work on this subject would be complete without taking into consideration articulatory data such as the one gathered by means of electromagnetic midsagittal articulometry systems. The authors of this paper are currently working on a wider study which aims to take into account these suggestions.

References


