Consistency in /r/ Trajectories in American English

Carol Y. Espy-Wilson

Electrical, Computer and Systems Engineering, Boston University, Boston, MA, USA

Research Laboratory of Electronics , Massachusetts Institute of Technology, Cambridge, MA, USA

ABSTRACT

We discuss the results of an acoustic study of the influence of postvocalic /r/ on neighboring segments in American English. The data suggest that a certain amount of time is needed to articulate an /r/ and that different speakers begin to produce /r/ at different times, depending on rate and context.

1. INTRODUCTION

A salient acoustic characteristic of American English /r/'s is a low third formant (F3) which is close in frequency to the second formant (F2). F3 for /r/ is usually around 2000 Hz or below, whereas for other segments, F3 is usually above 2400 Hz. There is substantial downward movement in F3 from a canonically articulated /ɑ/ to an /r/ in words like "car." However, in words like "cart" and "carwash," other F3 trajectories sometimes occur where the downward F3 movement is seen earlier [1]. In this study, we investigate the effects of speaking rate, context and speaker differences on the anticipation of /r/.

2. CORPUS

To conduct this acoustic study, the words "car," "cart," "carve," "card" and "carp" were embedded in the carrier phrase "Say _____ for me" and spoken by six speakers, four females (AF, LW, LT and MH) and two males (MR and JR). As a neutral case, the speakers also said the word "Nadav" (/Nədɑv/) in the sentence "Nadav was here." The speakers were recorded in a quiet room and instructed to speak at a slow and fast rate. The utterances were low pass filtered at 4800 Hz, sampled at 10 kHz and preemphasized. F3 tracks were obtained from DFT and LPC spectra with a 25.6 ms Hamming window.

3. ANALYSIS

In this section, we present measurements of speaking rate, a characterization of the F3 trajectories and a measure of how early speakers anticipate the /r/.

3.1 DURATION

Measurements of the sonorant interval showed that the average /ar/ duration across speakers (except for subject JR) was 295 ms for the words spoken at the slow rate and 182 ms for the words spoken at the fast rate. (JR did not always show durational differences.) As expected, average /ar/ durations were longer before voiced consonants (293 ms - slow, 221 ms - fast) than before unvoiced consonants (170 ms - slow, 142 ms -fast).

3.2 F3 trajectories

F3 trajectories observable during the sonorant region had four basic shapes. These shapes are shown schematically in Figure 1 with spectrograms for different pronunciations of "cart" which illustrate the corresponding F3 trajectory. First, as shown in part (a), F3 can start from a high position and move to a lower position (L-like). In this case, the vowel and /r/ appear to be produced canonically, with the /r/ articulation appearing at the end of the sonorant region. In part (b), F3 is rather flat and at a low position throughout. In this case, the /r/ and vowel appear to be completely coarticulated. In part (c), F3 moves from a low position at the beginning of the sonorant region to a higher position towards the end (J-like). Thus, as in part (b), it appears as if the /r/ is coarticulated with the vowel; however movement away from the /r/ to the

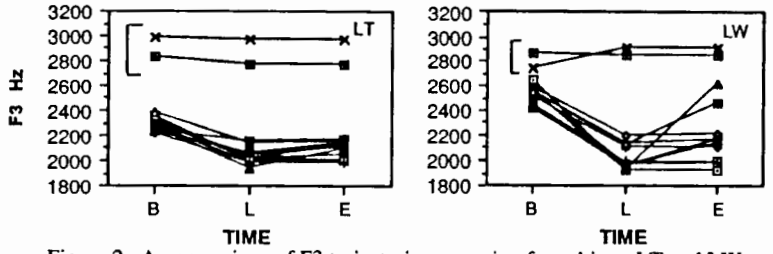

Figure 2. A comparison of F3 trajectories occurring for subjects LT and LW

F3 trajectory for MR always moves downward toward the end of the sonorant region (L-like) when the final consonant is labial, and always moves upwards towards the end of the sonorant region (J-like or U-like) when the final consonant is alveolar. On the other hand, LW shows a rate effect. All of the F3 trajectories show downward and upward movement when speaking slow and only downward movement when speaking fast.

As noted above, all of these patterns are consistent with a theory that the /r/ has a stable trajectory, but variable timing. The implication is that a U-shaped trajectory is always present but not visible. To support the theory, we compare in Figure 3 the /t/ bursts in the slow and fast pronunciations of "cart" by subject LW. The major spectral prominence of the /t/ burst in the fast pronunciation is around 1500 Hz lower than it is in the slow pronunciation. This substantial spectral difference suggests that the /r/ in the fast pronunciation of "cart" is still being articulated during the following /t/.

3.3 Anticipation of /r/

To develop a criterion by which it can safely be said that the /r/ is being produced, we used the F3 minimum (Fn) during the neutral case,the /ɑ/ in "Nadav." The beginning of r-coloring in the test words was taken as the time (TR) at which F3 during the test word fell 500 Hz below Fn. The difference of 500 Hz was chosen since other factors which can lower F3 such as the influence of a labial consonant should not result in such a large change. To measure when speakers started to produce an unambiguously r-colored sound, we subtracted B from TR, the time at which the sonorant region began. This difference was divided by the total duration of the sonorant region to normalize for speaking rate. Thus, the resulting values lie between 0 and 1. If F3 is 500 Hz below Fn at the beginningof the sonorant region, the normalized difference

Table 1. Shapes of F3 trajectories across all speakers as a function of rate and context. The words are specified by the final consonant.

		Shape of F3 Trajectories			
	Subjects	L-like	Flat	J-like	U-like
Slow Repetitions	AF	P			V,T,D
	LW				all
	MR	V,P			T,D
	MH	all			
	LT		all		
	JR	all			
Fast Repetitions	AF	P			V,T,D
	LW	all			
	MR	V,P		T	D
	MH	all			
	LT		all		
	JR	all			

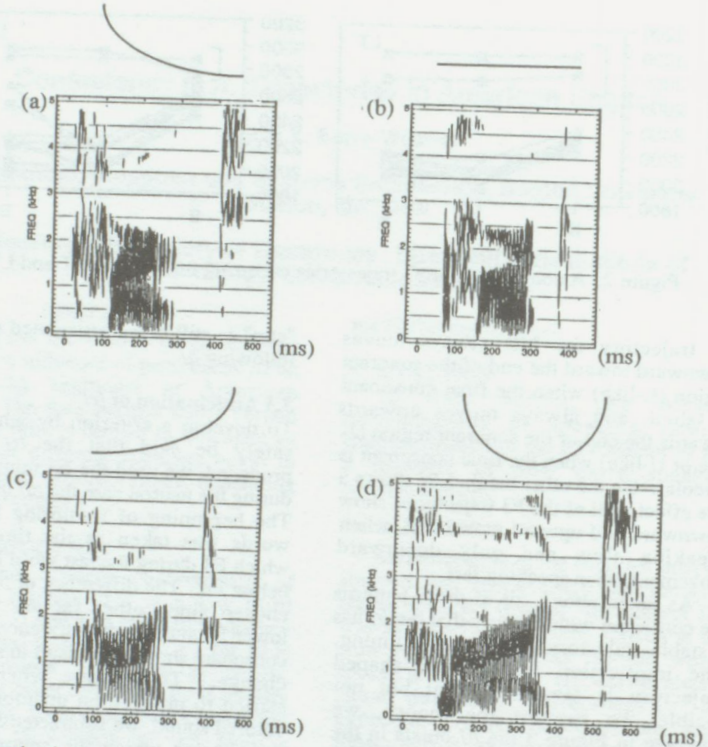

Figure 1. A comparison of F3 trajectories during different pronunciations of "cart."

following consonant is also evident. Finally, as in part (a), the F3 trajectory of part (d) starts from a high position and drops to a minimum. However, as in part (b), F3 rises towards the end of the sonorant region (U-like).

As we will discuss below, these data support the possibility that the U-like F3 trajectory occurs in all cases; however, there appear to be differences because the full F3 trajectory does not always occur within the sonorant region where the formants are visible. The other cases can be derived from the U-like trajectory. In the case of the L-like trajectory, the latter part of the F3 trajectory is coarticulated with the final consonant so that the upward F3 movement from the F3 minimum is not visible. For the flat trajectory, the beginning and end of the full F3 trajectory occur outside the sonorant region so that only the region around the F3 minimum is visible. Finally, for the J-like trajectory, anticipation of the /r/ occurs during the initial consonant so that the downward F3 movement occurs during the aspiration noise.

Figure 2 shows three F3 measurements for each word: the beginning of the sonorant region (B), the end of the sonorant region (E) and the F3 minimum (L) for subjects LT (left) and LW (right). The upper two trajectories in each graph are measurements of F3 during the /ɑ/ in the fast and slow pronunciations of "Nadav," which serves as the neutral case.

The plots for subject LT illustrate that some speakers have fairly uniform behavior across rate and context. The F3 trajectory is always relatively flat. On the other hand, other speakers like subject LW show more variability.

The shape of the F3 trajectories as a function of rate and context are summarized in Table 1 for each speaker. The data show that different speakers have different tendencies for when they begin to produce the /r/. For example, the

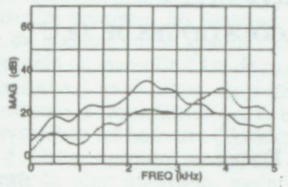

Figure 3. A comparison of the /t/ burtsts in the slow (dotted) and fast (solid) pronunciations of "cart" by subject LW.

speaking at a faster rate (the exception for subject AF is "carp").

4. CONCLUSIONS

The data in this study suggests that 1) the articulation of /r/ requires a minimum time of execution and 2) the acoustic consequence of the articulation of /r/ is a downward movement in F3 into the /r/ and an upward movement in F3 away from the /r/. However, this full F3 trajectory is not always observable because the formants are generally visible only during the sonorant regions of the

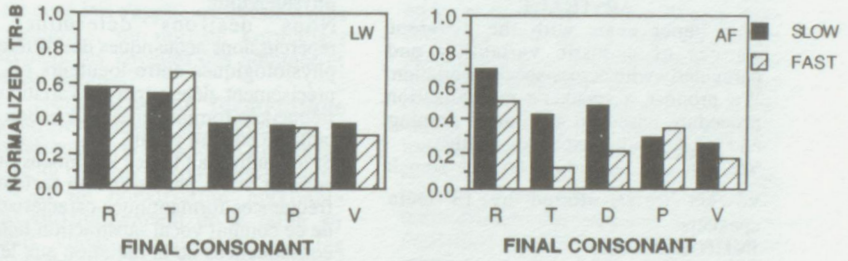

Figure 4. A plot of the normalized difference TR-B for subjects LW and AF

will be 0. If F3 falls 500 Hz below Fn at the end of the sonorant region, the normalized difference will be 1. Values in between indicate where in the sonorant region speakers started to produce the /r/.

A bar graph of this normalized difference is shown in Figure 4 for subjects LW (left) and AF (right). Recall that LW shows two distinct patterns of F3 trajectories when speaking slow and fast. The near equivalence of the column heights for subject LW means that there is little difference in when F3 begins to lower, relative to the beginning of the vowel. Thus, one explanation for the difference in the F3 trajectories is that the articulation of /r/ requires a minimum time of execution so that the upward movement in F3 away from the /r/ gesture is not seen in the case of the fast pronunciations because the sonorant region is too short. In fast speech, then, LW starts to produce the final consonant before finishing the production of /r/, as was seen in Figure 3. Interestingly, data from subject AF, whose F3 lowers at a more variable point in the word, shows that some speakers generally start to produce the /r/ much earlier when speech signal. The sonorant regions tend to change in duration due to factors such as speaking rate and the voicing of any final consonant. Furthermore, what portion of the F3 trajectory occurs during the sonorant region depends on how early speakers begin to produce the /r/. This study has shown that different people have different timing for an /r/ and that these tendencies can vary depending on speaking rate, and the voicing and place of articulation of a following consonant.

5. ACKNOWLEDGEMENTS

This work was supported in part by NSF Grant BNS-8920470. I also want to acknowledge Shawn Williams who made many of the measurements during her bachelor's thesis and Suzanne Boyce who gave many helpful comments.

6. REFERENCES

[1] Espy-Wilson, C. (1987), "An acoustic-phonetic approach to speech recognition: application to the semivowels," RLE Tech. Rep. 531, MIT.