The temporal location of rms peak in coarticulated vowels

Ewa Jacewicz and Robert Fox

The Ohio State University, 110 Pressey Hall, 1070 Carmack Road, Columbus, OH 43210, USA
jacewicz.1@osu.edu

The acoustic energy peak of the vowel, here defined as the rms peak, occurs typically before the temporal vowel midpoint. This study examines whether and how the location of the rms peak varies as a function of consonantal context and vowel duration. Four measures of rms peak location were explored: its distance (in ms) from vowel onset and from CVC-word onset (absolute measures) and its relative position (ranging from 0 to 100%) within the vowel and within the CVC-word (relative measures). Analysis of eight American English vowels produced in ten consonantal contexts by twenty speakers shows that the two relative measures yielded the most comparable and consistent patterns. The relative location of rms peak varied significantly as a function of consonantal context. It occurred earlier (i.e. closer to vowel or word onset) in the context of fricatives (and not stops), voiced consonants (as opposed to voiceless), and labials (as opposed to alveolars). The temporal location of rms peak was also closer to the vowel onset (or word onset) for long vowels as opposed to short. Overall, the results show that temporal location of rms peak can be to a large extent predicted on the basis of vowel duration.

1 Introduction

The center of a syllable, widely referred to as the syllable nucleus, consists in most cases of a nuclear vowel. One might consider the acoustic center of the vowel (e.g., in speech recognition applications) to be its peak energy, usually defined as the rms peak. It is widely recognized that the rms peak occurs typically before the temporal vowel midpoint. However, little is known about a systematic variation in its temporal location as a function of segmental and suprasegmental factors. This study explores the potential role of immediate consonantal context of the nuclear vowel on the variation in the temporal location of its rms peak. It examines how, for a given vowel, the location of the rms peak is affected by consonant voicing, place and manner of articulation.

The location of the rms peak can be expressed in several ways. First, the rms peak position can be determined in terms of its termporal distance (in ms) from vowel onset (absolute rms_loc_{vonset}) or word onset (absolute rms_loc_{wonset}). Alternatively, the peak rms can be expressed in terms of its relative position (with values ranging from 0 to 100%) with regard to the duration of the vowel (relative $rms\_loc_{vowel} = 100*rms\_loc_{vonset}$/vowel duration) or duration of the word (relative $rms\_loc_{word} = 100*rms\_loc_{wonset}$/word duration). All four approaches were explored in the present study to determine which measure best characterized the variation in the location of rms peak as a function of vowel category and consonantal context.

The present work is a part of a larger project undertaken to characterize the nature, size and range of acoustic amplitude variation in naturally produced coarticulated vowels [1]. The amplitude of rms peak and a set of measurements including frequency and amplitude of the first four vowel formants obtained at rms peak are reported in that study. The present focus is on characterizing the variation in the temporal location of the energy peak of the vowel using the data set reported in [1].

2 Methods

2.1 Speech materials

Eight American English vowels were selected: /i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u/. Each vowel was embedded in a symmetrical C_1VC_1 context. The selected consonant set consisted of ten oral consonants /p, t, k, b, d, g, f, v, s, z/. The CVCs were produced as monosyllabic words in a stressed position in a short phrase "It's a _____" [ɪtsə ____]. Each speaker recorded 240 utterances in a single session lasting approximately one hour (8 vowels × 10 consonants × 3 repetitions).

2.2 Speakers and procedure

Twenty speakers (ten men and ten women) of Midwestern American English participated. They were students aged 18-36 years enrolled in a variety of majors at The Ohio State University. Recordings were controlled by a custom program written in MATLAB. The speaker was seated in a sound-attenuated booth. A head-mounted Shure SM10A dynamic microphone was used, positioned at a distance of 2 inches from the speaker's lips. Speech samples were recorded and digitized at a 44.1-kHz sampling rate directly onto a hard disc drive. A word model containing the vowel of interest was first displayed on the computer monitor (e.g., "heat"). Next to the word, the stimulus sentence appeared (e.g, "It's a **beeb**"). The speaker read the sentence placing stress on the CVC-word. The presentation was blocked by vowel type. The order of consonantal contexts for each vowel type and the order of vowels themselves were randomized for each speaker.

2.3 Acoustic measurements

Prior to acoustic analysis, the tokens were digitally filtered and downsampled to 11.025 kHz. Vowel onsets and offsets were located by hand, primarily on the basis of a waveform display with segmentation decisions checked against a spectrogram. For vowels in the context of fricatives, vowel onset was measured from the cessation of noise in the periodic waveform following a frication offset and vowel offset was determined by the onset of noise in the waveform signaling the start of final frication. For vowels in the context of stops, vowel onset was measured from onset of periodicity (at a zero crossing) following the release burst. Vowel offset for voiceless stops was defined at the point at which the amplitude of the vowel dropped to near zero (which was also coincident with elimination of all periodicity in the waveform). The vowel offset for voiced stops was defined as that point when the amplitude dropped to near zero and any periodicity in the waveform contained no high frequency components. Vowel duration and the duration of the whole CVC-word obtained from these acoustic landmarks served as input for subsequent automated analysis of rms peak location.

To find the location and size of the amplitude peak, the rms values of a series of overlapping 25-ms windows were calculated whose number varied with the duration of the vowel. The onset of a series of 25-ms windows was incremented by 1 ms from the onset of the previous window. The onset of the first window began with the onset of the vowel and the offset of the final window ended with the offset of the vowel. The temporal location of the center of the 25-ms with the highest rms value was defined as the *location of the rms peak*, which provides an estimate of the location of the peak energy of the vowel.

3 Results

All four approaches to locating rms peak were initially tested, i.e., absolute rms_loc_{wonset}, absolute rms_loc_{vonset}, relative rms_loc_{word}, and relative rms_loc_{vowel}. Based on the statistical results, it became clear that the two relative measures yielded the most comparable patterns. The results for the absolute rms_loc_{wonset} were in line with the relative measures only for the effect of consonant voicing but not for manner or place. The most conflicting patterns for the effects of voicing, manner and place were found between absolute rms_loc_{vonset} and each of the three other measures. We therefore decided to present here only the results for the two most dependable measures, i.e. relative rms_loc_{vowel}, and relative rms_loc_{word}.

3.1 Rms peak within vowel

As shown in Fig. 1, the relative positions of rms peaks, ranging from 28% to 42%, occur earlier (i.e. closer to vowel onset) for long vowels and later (i.e., closer to vowel midpoint) for short vowels. The exception here is the vowel /ɔ/ whose peak is relatively late (37%-point in time) for its long duration of 279 ms.

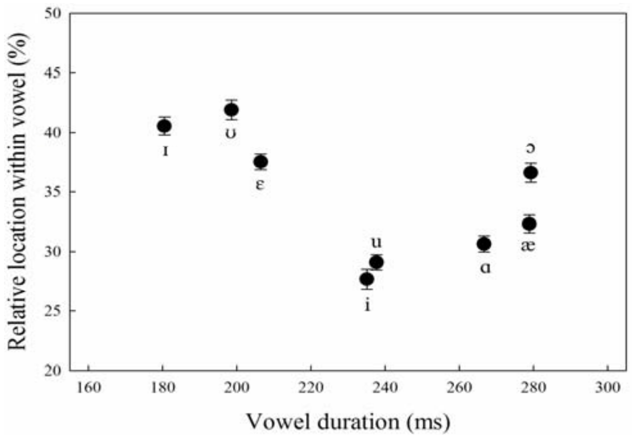

Fig. 1 Mean relative location of rms peak within vowel (st.error). Each vowel's mean duration is averaged across ten consonantal contexts.

The variation of rms_loc_{vowel} for each vowel as a function of individual consonant is shown in Fig. 2. As can be seen, consonantal context effects are considerable. For example, for the vowel /æ/, rms peak occurs much earlier in the b_b context (22%-point in time) than in the p_p context (47%-point in time).

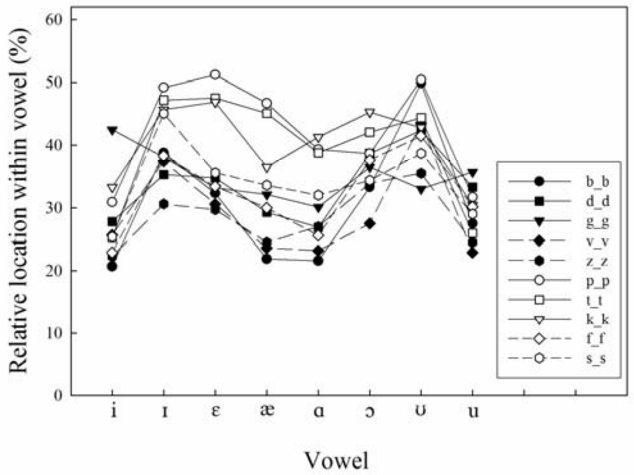

Fig. 2 Relative location of rms peak within vowel for each vowel and each individual consonantal context.

To assess the significance of this variation, a within-subject ANOVA was first conducted for rms peak location within vowel (relative rms_loc_{vowel}) with vowel and consonant as within-subject factors, and gender as a between-subjects factor. All three main effects were significant, indicating that rms peak location differs as a function of vowel [$F(7, 70) = 11.64$, $p < 0.001$, $\eta^2 = 0.538$] and consonant [$F(9, 90) = 7.7$, $p < 0.001$, $\eta^2 = 0.435$]. In terns of the significant gender effect [$F(1, 10) = 6.29$, $p = 0.031$, $\eta^2 = 0.386$], mean rms peaks in the female vowels occurred slightly later in the course of vowel duration (37%) than for male vowels (31%). As expected, there was also a significant vowel × consonant interaction [$F(7.2, 72.4) = 3.43$, $p = 0.003$, $\eta^2 = 0.255$].

Separate repeated-measures ANOVAs were subsequently completed on each individual vowel. The analyses were done separately for the stop and fricative set /b, p, d, t, v, f, z, s/ with the factors manner, voicing and place, and for the stops only set /b, p, d, t, k, g/ with the factors voicing and place. Significant consonantal effects for each vowel in terms of consonantal features manner, voicing, and place are summarized in Table 1 for the front vowels /i, ɪ, ɛ, æ/ and in Table 2 for the back vowels /ɑ, ɔ, ʊ, u/.

The effect of manner was significant for all vowels but /i/ and /u/. Uniformly, rms peaks for vowels in the context of stops occurred relatively later than in the context of fricatives. This was most likely due to the fact that vowels in the context of stops were shorter than in the context of fricatives.

The effect of consonant voicing, albeit significant only for /ɪ, ɛ, æ, ɑ/, was again uniform across all vowels in that rms peaks for vowels in the context of voiceless consonants occurred later in time than in the context of voiced ones. This again can be explained on the basis of differences in vowel durations, which were shorter in the context of voiceless consonants as compared to voiced.

The effects of consonant place were more variable. In the analyses of stops and fricatives (labial and alveolar contexts) the rms peaks in alveolar contexts were generally later than in the labial contexts. For the stops-only analysis (labial, alveolar, and velar contexts), the effect of place was mostly not significant and consistency was found only for the vowels /i, ɑ, ɔ/, where the peaks in velar contexts were later than in alveolar and velar contexts, respectively.

	/i/	/ɪ/	/ɛ/	/æ/
/ b, d, p, t, v, z, f, s/				
Manner	----	.402**	.700†	.635†
		st>fr	st>fr	st>fr
Voicing	----	.336**	.516†	.653†
		vl>vd	vl>vd	vl>vd
Place	----	----	----	.217*
				alv>l
/b, d, g, p, t, k/				
Voicing	----	.291*	587†	.604†
		vl>vd	vl>vd	vl>vd
Place	.540†	----	----	----
	vel>alv>l			

Table 1 Significant main effects of consonantal features on rms peak location for front vowels. Shown are partial eta squared values (η^2). * p<0.050, ** p<0.010, † p<0.001; ---- not significant; st = stop, fr = fricative, vd = voiced, vl = voiceless, l = labial, alv = alveolar, vel = velar.

	/ɑ/	/ɔ/	/ʊ/	/u/
/ b, d, p, t, v, z, f, s/				
Manner	.323*	.274*	.523†	----
	st>fr	st>fr	st>fr	
Voicing	.494**	----	----	----
	vl>vd			
Place	.393**	.429**	.325*	----
	alv>l	alv>l	alv>l	
/b, d, g, p, t, k/				
Voicing	.536**	----	----	----
	vl>vd			
Place	----	----	.430†	----
			l>alv>vel	

Table 2 Significant main effects of consonantal features on rms peak location for back vowels. Shown are partial eta squared values (η^2). * p<0.050, ** p<0.010, † p<0.001; ---- not significant; st = stop, fr = fricative, vd = voiced, vl = voiceless, l = labial, alv = alveolar, vel = velar.

It is noteworthy that there were no significant main effects of consonantal features for the vowel /u/. Similarly, the rms peak location for the vowel /i/ was mostly unaffected by consonantal context with the exception of place for the stop only analysis. Given that the rms peaks for these two high vowels occurred very early (see Fig. 1), it is tempting to speculate that the proximity to vowel onset (rather than to vowel midpoint) had a smaller effect on the variation in the temporal location of rms peak. However, this interpretation requires verification in light of additional data.

The effect of speaker gender was significant only for the vowel /ʊ/ [$F(1, 17) = 6.28$, $p = 0.023$, $\eta^2 = 0.270$]. For this vowel, the rms peaks for female speakers occurred later than for males.

3.2 Rms peak within word

The results for the second measure of rms peak location, i.e., relative to the duration of the entire CVC-word (relative rms_loc$_{word}$) are generally in line with the pattern found for the relative location of rms peak within vowel. Overall, the peaks within words occurred later as compared to the results for rms peaks within vowels, ranging from 39% to 47%. This outcome can be attributable, for the most part, to the duration of the consonants surrounding the vowel.

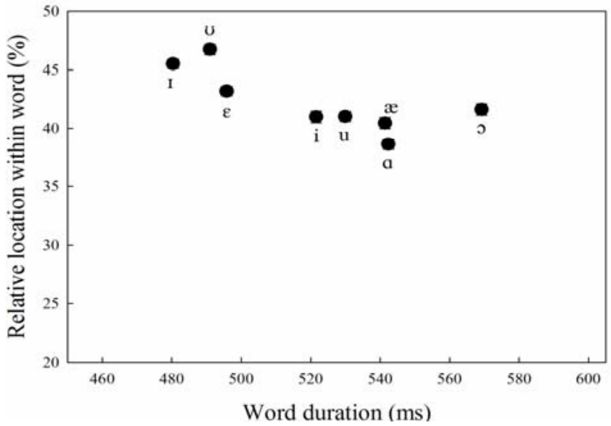

Fig. 3 Mean relative location of rms peak within CVC-word (st.error). Each words's mean duration is averaged across ten consonantal contexts.

As shown in Fig. 3, the peaks occurred again later for short vowels /ɪ, ɛ, ʊ/. For the remaining long vowels, the locations of rms peaks within word were less variable than within vowel, including the vowel /ɔ/ (compare Fig. 1).

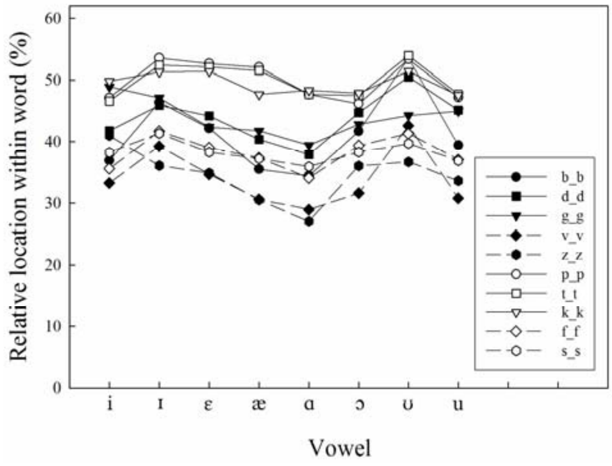

Fig. 4 Relative location of rms peak within word for each vowel and each individual consonantal context.

For each vowel, the pattern of variation in the relative positions of rms peak within word as a function of consonantal context is shown in Fig. 4. As it was found for relative rms_loc$_{vowel}$, the peaks for words in the context of stops occurred later than in the context of fricatives. They were also later in the context of voiceless consonants rather than voiced (compare Fig. 2). Since the variation across vowels was generally less dramatic for relative rms_loc_{word}, the separation between peaks in the contexts of stops and fricatives displayed in Fig. 4 can be observed more directly.

Results of an overall ANOVA for relative rms_loc_{word} indicated that significant main effects were consistent with those for relative rms_loc_{vowel}. The effect of manner was strong [$F(1, 17) = 306.16$, $p < 0.001$, $\eta^2 = 0.947$], showing that the rms peaks occurred later in time in the context of stops than fricatives. The rms peaks in the context of voiceless consonants occurred significantly later in time [$F(1, 17) = 82.09$, $p < 0.001$, $\eta^2 = 0.828$] than in the context of voiced. The effects of place were more variable, consistently with the results for relative rms_loc_{vowel}.

4 Discussion

The results for the relative measures indicate that consonantal context affects the location of rms peak in a systematic way. There is a relationship between vowel's duration (which also varies as a function of consonantal features) and the location of rms peak within vowel.

Exploring this relationship, we will discuss the correspondence between vowel duration data from the present speakers and relative locations of rms peaks for these vowels. Table 3 shows durations of individual vowels as a function of consonant voicing and manner. The durations were averaged across speaker gender. As can be seen, all vowels were longer in the context of voiced consonants as opposed to voiceless. The differences were of the range 87-102 ms. The durations of all vowels tended to be longer in the present data because they were produced in a stressed position and included female vowels which typically have longer durations [2]. The present duration data are fully comparable with those in [3] for the voiced and voiceless consonantal contexts. As shown in Table 3, all vowels are also longer in the context of fricatives as opposed to stops. However, the differences were smaller than for the voicing contrast, ranging from 29-54 ms.

Vowel	Vd_Vd	Vl_Vl	Fr_Fr	St_St
/ɪ/	227 (42)	134 (35)	198 (65)	169 (62)
/ʊ/	242 (48)	155 (44)	213 (59)	179 (60)
/ɛ/	251 (51)	162 (44)	230 (72)	191 (63)
/i/	281 (55)	189 (61)	263 (81)	217 (71)
/u/	288 (57)	187 (48)	256 (82)	226 (70)
/ɔ/	316 (66)	217 (62)	309 (90)	259 (89)
/ɑ/	328 (64)	226 (73)	294 (78)	248 (82)
/æ/	332 (69)	230 (63)	311 (75)	257 (82)

Table 3 Durations of individual vowels (in ms) in consonant environments shown for voiced (Vd_Vd) and voiceless (Vl_Vl), fricatives (Fr_Fr) and stops (St_St) symmetrical contexts. Standard deviations are in parentheses.

The corresponding locations of rms peaks for the individual vowels as a function of consonant voicing and manner are shown in Fig. 5 and Fig. 6.

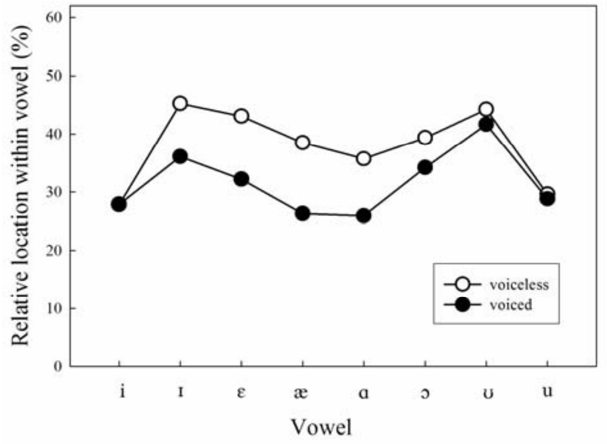

Fig. 5 Relative location of rms peak within vowel in voiceless and voiced consonantal context.

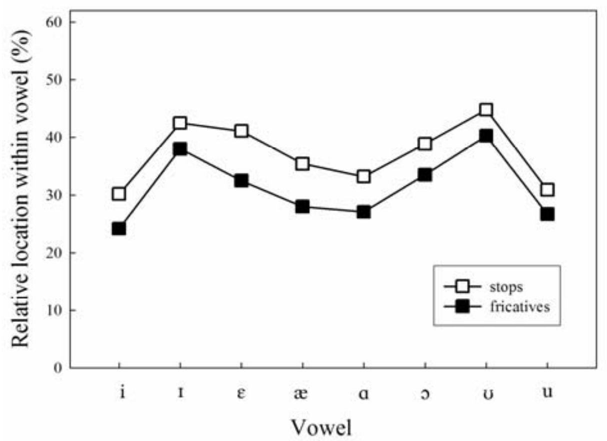

Fig. 6 Relative location of rms peak within vowel in the context of stops and fricatives.

It is apparent that rms peaks occur later for vowels having shorter durations, i.e., in the context of voiceless consonants and in the context of fricatives. For the voicing contrast, the differences in the location of rms peaks for the vowels /ɔ, ʊ/ were too small to reach statistical significance. However, the location of the peaks conforms to the general pattern. As an exception, consonant voicing did not have an effect on the variation in rms peak location for the vowels /i, u/.

With regard to consonant manner, the peaks were again later in the context of stops, in which vowel duration is shorter, as opposed to the context of fricatives. This effect was uniform across all vowels although the differences for the vowels /i, u/ did not reach statistical significance.

The effects of consonant place on rms peak location were more consistent for the labial-alveolar distinction when both stops and fricatives were compared. The peaks occurred later for the alveolar contexts as opposed to labial. However, there was less consistency in the patterns for the velar consonant when stops only were analyzed. The results were mixed (and the differences were mostly not significant). This suggests that vowel duration differences as a function of consonant place distinction may not be large enough to affect the location of rms peak in a systematic way.

5 Conclusion

This study shows that the relative location of the rms peak of a vowel varies significantly as a function of consonantal context. There is a relationship between vowel's duration and the location of rms peak within vowel. Specifically, the peaks occur earlier (i.e. closer to vowel or word onset) for longer vowels such as in the context of fricatives (and not stops) and voiced consonants (and not voiceless). Averaged across all consonantal contexts, the rms peaks occur earlier for long vowels than for short. This indicates that the specific temporal location of rms peak can be to a large extent predicted on the basis vowel duration.

The effects of consonant place of articulation were generally smaller, with the tendency for the peaks to occur earlier in the context of labials as opposed to alveolars.

These systematic effects were obtained when the temporal location of rms peak was measured in relative terms, i.e. with regard to the duration of the vowel or to the word. Measured as absolute temporal locations (in ms from vowel or word onset), the results were highly variable and no clear effects of consonantal context could be found.

Acknowledgments

This study was supported by the research grant No. R03 DC005560 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. The authors would like to thank Chiung-Yun Chang for assistance at various stages of this research.

References

[1] E. Jacewicz, R. A. Fox, "Amplitude variations in coarticulated vowels", *J. Acoust. Soc. Am.* 123 (2008)

[2] E. Jacewicz, R. A. Fox, J. Salmons, "Vowel duration in three American English dialects", *American Speech* 82, 367-385 (2007)

[3] J. Hillenbrand, M. J. Clark, T. M. Nearey, "Effects of consonantal environment on vowel formant patterns", *J. Acoust. Soc. Am.* 109, 748-763 (2001)