Journal of Phonetics (1995) **23**, 305–322

Coarticulation and the accented/unaccented distinction: evidence from jaw movement data

Jonathan Harrington

Speech Hearing and Language Research Centre, Macquarie University, Sydney, Australia

Janet Fletcher

Department of Linguistics, University of Melbourne, Melbourne, Australia, and Speech Hearing and Language Research Centre, Macquarie University, Sydney, Australia

Corinne Roberts

Speech Hearing and Language Research Centre, Macquarie University, Sydney, Australia

Received 31st May 1994, and in revised form 3rd December 1994

Both coproduction and the more recent task dynamic models of speech production have advanced an explanation for certain kinds of vowel shortening in terms of coarticulatory overlap between neighbouring speech sounds. In this study, the extent to which coarticulatory overlap between opening and closing jaw movement gestures accompanies the accented/unaccented vowel distinction is considered. The authors begin by quantifying the salient differences between truncation, which is caused by a close phasing of articulatory gestures, and linear rescaling of jaw movements. In both truncation and linear rescaling, the duration and amplitude of movement decrease, and the peak velocities remain the same: the main differences occur in the resulting shape of the waveform between the temporal location of the peak velocities in the opening and closing gestures. Three parameters that encode these shape differences are then applied to the speech movements produced by three subjects. This study shows that the accented/unaccented differences are more appropriately modelled as a consequence of truncation, than linear rescaling. © 1995 Academic Press Limited

1. Introduction

A very well-documented finding in the phonetics literature is that the duration of speech segments, and of vowels in particular, is influenced by a variety of different factors, including the rate of speech, the prosodic organisation and grammatical structure of the utterance, and intrinsic phonetic effects, to mention but a few (Klatt, 1975; Umeda, 1975; Crystal & House, 1988). One of the innovative aspects in the development of coproduction accounts of speech articulation (Fowler, 1980, 1983, 1986; Fowler & Saltzman, 1993) is that some kinds of vowel shortening are directly explained as the result of coarticulatory processes. A central aspect of coproduction

0095-4470/95/030305 + 18 $12.00/0

models, which is taken over from the earlier acoustic studies of Öhman (1966), is that consonants and vowels are characterised by independent articulatory strategies that overlap in time. Furthermore, these strategies can overlap to a variable degree and the greater the overlap, the more extensively they are coarticulated, or coproduced. Although an empirical demonstration is far from straightforward (Fowler, 1981), coproduction models predict that a greater extent of coarticulatory overlap is directly associated with increased vowel shortening. This is for two reasons. Firstly, because a segment's articulatory duration is presumed to be invariant (therefore the greater the extent of overlap, the shorter the non-overlapping extent of the segment); and secondly, because phoneticians conventionally place acoustic boundaries at the edges of the non-overlapping section of the segment, which are the points where the segment first becomes, and then ceases to be, acoustically prominent (Fowler, 1984; Fowler & Smith, 1986).

This relationship between the relative timing of articulatory strategies, coarticulation, and vowel shortening is also an integral part of the more recent task dynamic model of speech production (Browman & Goldstein, 1989, 1990; Saltzman & Munhall, 1989). In this model, speech production is described in terms of *gestures*, which are functionally equivalent ways of achieving the same phonetic goal (e.g., *bilabial closure*, in which the jaw and lips work together to achieve the same goal, but in different ways, depending on factors such as context). In a task dynamic model, the duration of a gesture is influenced by two main parameters, one of them intragestural, the other intergestural. The intragestural parameter is analogous to stiffness in a mass-spring system: essentially, the stiffer the gesture, the faster, and therefore the shorter, the movement is. The intergestural parameter is defined by the relative phasing between two gestures: the earlier a gesture is phased with respect to a following gesture, the greater the extent of gestural overlap, and therefore the greater the measured shortening of the phonetic segment with which the gesture is associated (see Browman & Goldstein, 1990; and Hawkins, 1992 for more detailed reviews).

The articulatory correlate of a closer phasing of gestures in the task dynamic model depends on the extent to which the overlapping gestures share articulators (Fowler & Saltzman, 1993). When the gestures are defined by different sets of tract variables, they have no, or only some, articulators in common: consequently, when such gestures overlap spatially, they interfere minimally with each other. As Fowler & Saltzman (1993) show, a representative example would be vowel-symmetric /VdV/ sequences in which, despite the articulatory overlap between the alveolar and the flanking vowels, the tongue-tip constriction goal is nevertheless met in the context of different phonetic vowels, precisely because the vowel and consonant gestures control (to a large extent) different articulators. However, when overlapping gestures share most of the articulators, then the potential for mutual interference is created which can result in *intergestural blending.* An example of blending is the advanced, pre-alveolar realisation of /t/ in the first word of *Put the toe on the table* (with stress on *toe*) in de Jong (1991). In this case, the stop and following dental compete for the same articulator (the tongue tip), and the result is a single gesture for the cluster [tð] which is intermediate between the alveolar and dental positions.

The mathematical model of gestural coordination in Saltzman & Munhall (1989) allows for many different kinds of blending and some of these are supported by experimental data (see e.g., Boyce, 1990 and Löfqvist, 1990 for evidence for additive blending). One kind of blending that has recently been discussed in relation to measured vowel shortening and articulatory undershoot is *truncation.* In this type of blending, which is similar to the truncation effects that have been observed in limb movement by Bullock & Grossberg (1988), the conflicting demands that are made on the same articulator by opposing gestures causes one to be "cut off" by the other. Some relevant data in support of truncation is presented in Munhall, Fowler, Hawkins, & Saltzman (1992) who compared jaw trajectories in vowels preceding single and multiple consonant codae (e.g., *pap vs. paps*). As others have found (e.g., Lindblom & Rapp, 1973), the acoustic vowel duration was shorter before consonant clusters; however, since the extent of jaw lowering (for the vowel) was the same before the single consonants and clusters, and since there was also no difference in the peak velocity of jaw movement, Munhall *et al.* (1992) reasoned that the decrease in duration was the result of the offset of the opening gesture being truncated by the closing gesture in syllables with final clusters.

In Munhall *et al.*'s (1992) experiments, the extent of gestural overlap was sufficiently small that truncation did not produce vowel undershoot (as determined from the jaw trajectory) for vowels before clusters. However, when the overlap between the vowel's opening and closing gestures is extensive, it is possible for the entire relatively steady-state, final section of the opening gesture (corresponding to the "vowel target" in more traditional terminology) to be cut off. The result of this more extreme form of truncation is a reduction in the extent of displacement of the supralaryngeal articulators (analogous to target undershoot). This type of truncation characterises the distinction between accented and unaccented vowels in the studies of Beckman and her colleagues (Edwards, Beckman, & Fletcher, 1991; Beckman, Edwards, & Fletcher, 1992; see also Beckman, de Jong, Jun & Lee, 1992; de Jong, Beckman, & Edwards, 1993). In Beckman, Edwards, & Fletcher (1992), the significantly smaller jaw displacement in unaccented, compared with accented, vowels is directly attributed to the considerable truncation of the opening gesture by the following closing gesture.

The interpretation of phonetic differences in terms of gestural overlap and truncation is interesting because it implies a form of articulatory invariance which may well underlie listeners' perception of phonetic segments (e.g., Fowler & Rosenblum, 1989). For example, it is possible that listeners perceive the same phonetic vowel before single consonants and clusters, despite considerable variation in acoustic duration, precisely because vowels before single consonants and clusters have the same underlying opening and closing gestures (but phased differently). However, a major difficulty in establishing evidence for truncation, at least from jaw movement data, is in knowing what constitutes a truncated vowel. For example, although a truncated vowel is likely to be accompanied by a reduction in both duration and displacement, together with minimal changes in the peak velocity of movement, these articulatory characteristics are also compatible with making the vowel "smaller" by linear rescaling (analogous to looking at a movement waveform through a zoom lens, and zooming out, producing smaller durations and displacements, but maintaining the same overall shape, and therefore the same peak velocities).

In the first part of this study, we consider some of the principal articulatory parameters that can be used as evidence for truncation; the null hypothesis against which this evidence is assessed is a rescaling model, in which the sizes of the gestures are scaled in proportion to changes in duration. In the second part of the study, we reconsider whether the accented/unaccented difference is appropriately modelled in terms of truncation, or whether this prosodic difference is more accurately modelled by linear rescaling. The motivation for testing truncation against this model is that linear rescaling is one possible interpretation of the production differences between accented and unaccented vowels. For example, it may be that the production of unaccented vowels is accompanied by a smaller gestural amplitude resulting in vowel centralisation. Some evidence for this second interpretation is presented (for French data) in Fletcher & Vatikiotis-Bateson (1990), in which unaccented vowels are considered to be "smaller" versions of accented vowels. Another reason for evaluating a model of linear rescaling is that, while a non-significant change in peak velocity values is compatible with a model of truncation, it is also consistent with a model of linear rescaling in which the gestures are made smaller. If vowel shortening in contexts such as unaccented vowels and before consonant clusters is to be attributed to truncation effects, then it is also necessary to consider articulatory parameters which can adjudicate between models of truncation and rescaling.

2. Articulatory characteristics of truncated vowels: simulated data

2.1. *Talkers and materials*

Four adult talkers of Australian English (3 female, 1 male) with no known speech or hearing disorders, participated in the experiment. The subjects' accents can be described as intermediate between Cultivated and General Australian (Bernard, 1970; Mitchell & Delbridge, 1965). The materials were designed to elicit an accented, and an unaccented, production of *barb* in sentences (1) and (2) respectively:

(1) say *barb* naturally, not bub
(2) say barb *naturally*, not slowly

Talkers were instructed to produce each sentence as two phrases, with an intermediate phrase boundary (Beckman & Ayers, 1994) after *naturally*. In the first phrase of sentence (1), talkers were instructed to accent *barb*, and not to accent any other words in that phrase. In the first phrase of sentence (2), talkers were instructed to accent *naturally*, and not to accent any other words in that phrase. Thus the distinction between sentences (1) and (2) that is relevant for this experiment is that *barb* is (nuclear) accented in (1), and unaccented in (2). In Australian English, *barb* and *bub* have a similar vowel quality and are mainly distinguished by length. Appropriate phonetic transcriptions are [bä:b] and [bäb], respectively (Australian English is non-rhotic).

Each sentence was produced 15 times at two rates (slow and fast) resulting in 2 (sentences) × 2 (slow/fast) × 15 = 60 utterances per talker. Only the slow productions of *barb* will be considered in this paper (30 utterances per talker). Furthermore, since the male talker was not judged to have produced reliably the accent distinction on which this experiment is based (the talker produced an accented *barb* on several occasions, in sentence (2)), that talker's data were discarded. Two of the remaining three (female) talkers (JF, CR) are authors of this paper.

2.2. *Physiological and acoustic recordings*

Recordings of the speech waveform and 6 articulatory channels (lower lip, upper lip, jaw, all in the vertical and horizontal dimensions) were made at the Speech Hearing and Language Research Centre, Macquarie University using the Movetrack system interfaced to the speech signal processing package Waves+. The sampling rates for the articulatory and acoustic data were 200 Hz and 20 kHz, respectively.

The acoustic boundaries of the target word were marked using Waves+, and the data was read into the mu+ speech database analysis system (Harrington, Cassidy, Fletcher & McVeigh, 1993) for kinematic and all subsequent analyses. In the first stage of the kinematic analysis, opening and closing gestures of the jaw were marked for each target word. These articulatory boundaries were calculated automatically by differentiating the movement waveform to derive the jaw velocity, and then obtaining the time values at which the velocity was zero. The peak velocities of the opening and closing gestures were obtained from the peaks of the velocity waveform (Fig. 1). If an automatically calculated displacement, or velocity, peak was considered to be inaccurate, it could be changed interactively; since, for the present

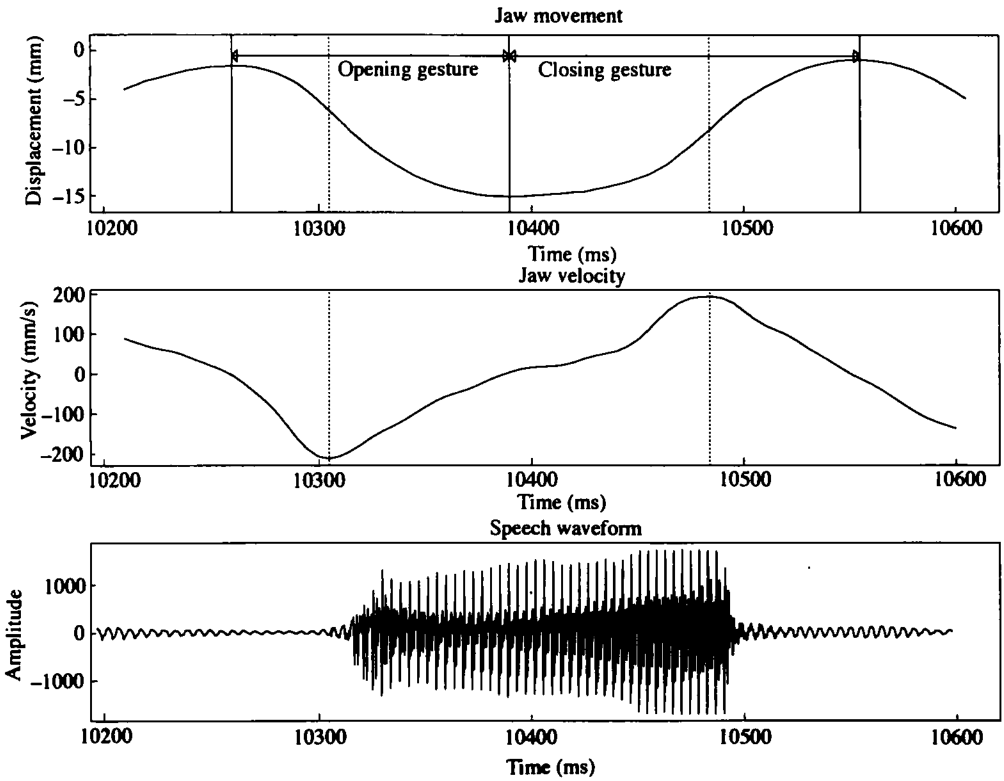

Figure 1. Movement and velocity waveforms with the boundaries of opening and closing gestures. The times of the peak velocities are shown as vertical dotted lines.

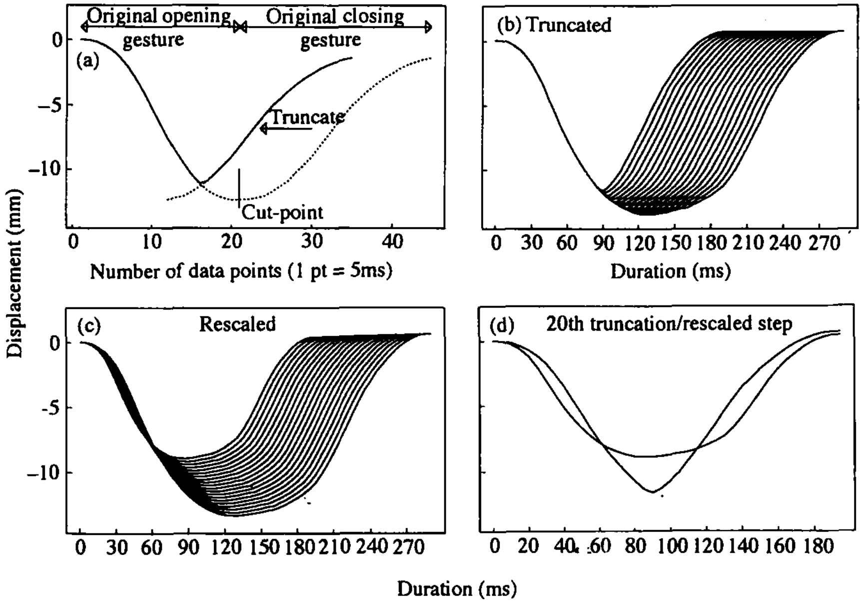

Figure 2. (a), the process by which a waveform is truncated; (b), truncations at 5 ms intervals; (c), linearly rescaled waveforms at 5 ms intervals; (d), superimposed truncated and rescaled waveforms.

data, all the automatically calculated displacement and velocity peaks were considered to have been reliably placed, none were changed manually.

2.3. *Simulations of truncation and rescaling*

A single accented *barb* token from talker JF was truncated in steps of 5 ms (1 data point at a 200 Hz sampling rate) for 20 steps. The token was selected on the basis that it had both clear displacement peaks, and smooth opening and closing trajectories between the peaks.

The process of truncation is illustrated in the top left panel of Fig. 2 in which a jaw movement waveform is truncated by 45 ms (9 data points at a 200 Hz sampling rate). The truncation is made by cutting the waveform at the peak displacement for the vowel (the cut point) and sliding the closing gesture to the left by a specified number of data points. The new waveform, indicated by the solid line, has a peak vowel displacement where the opening and closing gestures bisect (top left panel, Fig. 2). The top right panel of Fig. 2 shows the resulting waveforms that are produced when the waveform is truncated in 20 steps of 5 ms (5–100 ms truncation).

Rescaling involves resizing (shrinking) the movement waveform in linear proportion to changes in duration. Specifically, if the duration of the waveform is i ms (the summed duration of the opening and closing gestures), and if the duration is reduced by j ms, then the jaw displacement values are reduced (multiplied) by a factor of $(i-j)/i$. The resulting rescaled waveforms in steps of 5 ms for 20 steps are shown in the bottom left panel of Fig. 2. The bottom right panel of Fig. 2 shows superimposed truncated and rescaled waveforms when the original waveform was

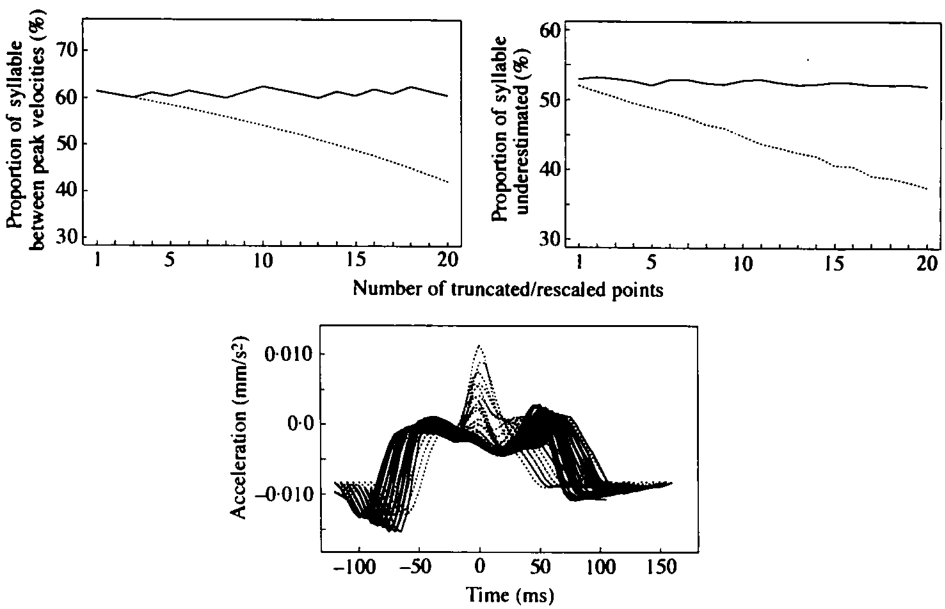

Figure 3. Truncated (· · · · ·) and rescaled (——) waveforms on three articulatory parameters.

reduced by 100 ms. It should be noted that neither truncation nor rescaling produces any significant changes in the peak velocity values.

2.4. Quantification of the differences between truncated and rescaled vowels

The most important difference between truncation and rescaling lies in the shape of the displacement waveform between the peak velocity times. Specifically, the opening and closing gestures of a truncated waveform converge to a sharper peak than those of a rescaled waveform, as the extent of the truncation increases. This is clearly shown in the bottom right panel of Fig. 2. Accordingly, three articulatory parameters were devised which are all based on this observed difference between truncated vowels and their rescaled counterparts.

The first of these parameters (top left panel in Fig. 3) is the percentage of the syllable's duration which lies between the time points of the opening and closing gestures' peak velocities. As the syllable is increasingly truncated, the proportion of the syllable which lies between the peak velocities decreases. However, the proportion remains constant when the syllable is rescaled. The proportional duration is given by:

(1) $$100(v_{tc} - v_{to})/t\%$$

where v_{tc} and v_{to} are the times of the peak velocities in the closing and opening gestures respectively, and t is the articulatory syllable duration (the sum of the durations of the opening and closing gestures).

The second parameter is closely based on predicting the syllable duration from the ratio of the displacement to the peak velocity; a metric of this kind has been used by Beckman, Edwards & Fletcher (1992) in support of their evidence that unaccented vowels are truncated. Essentially, the ratio of the displacement to the peak velocity is equal to the time taken to travel the displacement *if the velocity is constant* (deviates negligibly from the peak velocity). This is because if a displacement s is covered in t seconds, the velocity is constant at s/t, and consequently the displacement/peak velocity ratio is $s/(s/t) = t$. If, however, the velocity deviates considerably from the peak velocity, then a discrepancy is produced between the displacement/peak velocity ratio and the actual time it takes for the displacement to be covered. Applying this reasoning to jaw movement waveforms, the more the opening and closing gestures resemble a triangular shape with an apex at the peak lowering for the vowel (i.e., with constant velocities in the opening and closing gestures), the better the displacement/velocity ratio predicts the time taken to traverse the displacement. Therefore, since truncated vowels are more peaked than rescaled vowels, their displacement/velocity ratios should correspond more closely to the actual syllable duration. In confirmation of this, the top right panel of Fig. 3 shows that the proportion of the actual syllable duration which is underestimated by the displacement/velocity ratio decreases with increasing truncation; for rescaled vowels, on the other hand, the proportion remains constant. This proportion is given by:

(2) $$100[t - (s_o/v_o + s_c/v_c)]/t\%$$

where t is the actual syllable duration (sum of the durations of the opening and closing gestures), s_o/v_o the displacement/peak velocity ratio in the opening gesture, and s_c/v_c the displacement/peak velocity ratio in the closing gesture.

The final parameter is based on an examination of the acceleration of the jaw (obtained by double differencing the jaw movement waveform). Since truncated vowels have a sharp peak for large truncation values, the rate of change of velocity, i.e., acceleration, should be high at this point. Since on the other hand, rescaled vowels have a plateau at the maximum jaw lowering for the vowel, the rate of change of velocity is minimal. In confirmation of this, the bottom panel shows increasingly higher acceleration values for the increasingly truncated vowels, whereas for the rescaled vowels, the acceleration waveforms are falling, and are close to zero.

In the following section, the three parameters are applied to the accented vowels. Specifically, the aim of the experiment described in the next section is to determine whether unaccented vowels resemble more closely (i) truncated accented vowels or (ii) rescaled accented vowels.

3. The accented/unaccented distinction

3.1. *Method*

The accented vowels of the three talkers were truncated or rescaled following the techniques described in Section 2.3. In an attempt to match as closely as possible the modified accented vowels to the unaccented vowels, each accented vowel was truncated, or rescaled, by the talker's average durational difference between the accented and unaccented vowels (to the nearest data point, or 5 ms). Specifically,

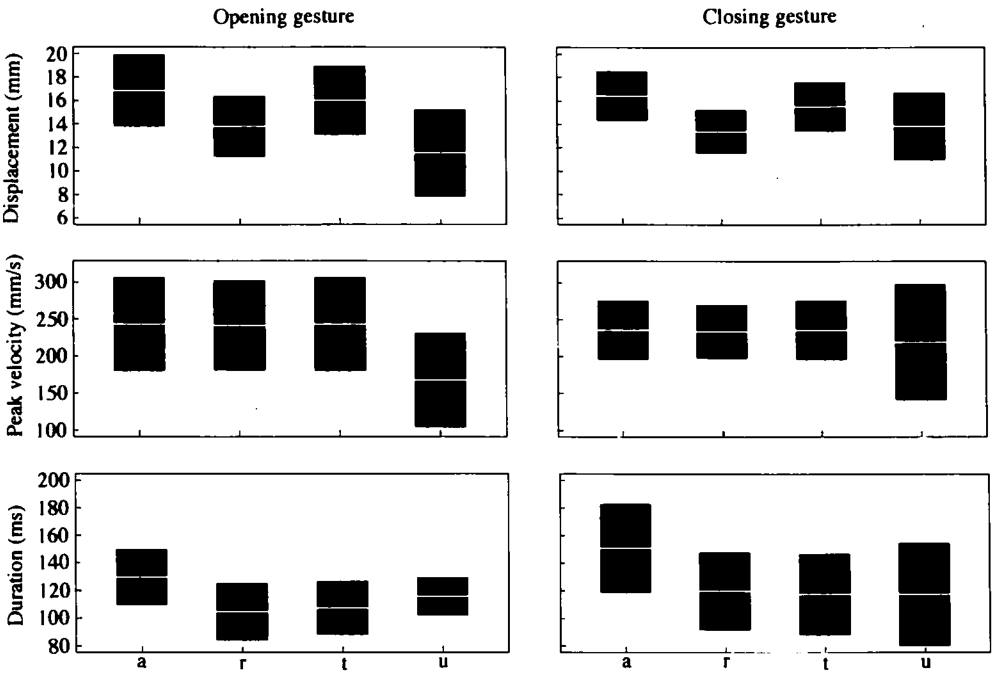

Figure 4. Displacement, peak velocity, and duration of the opening and closing gestures for subject CR. *a*: accented; *r*: rescaled accented; *t*: truncated accented; *u*: unaccented. The boxes represent two standard deviations on either side of the mean, which is shown as a horizontal white line.

the accented vowels were truncated/rescaled, by 50 ms (talker CR), 120 ms (talker KC), and 100 ms (talker JF). A total of 42 accented vowels were modified (13 for CR; 14 for KC; 15 for JF). Three accented vowels (2 from CR, 1 from KC), and three unaccented vowels (2 from CR, 1 from KC) were discarded, because of problems in the temporal alignment of the articulatory and acoustic channels during recording and digitisation.

3.2. *Results*

The displacement, peak velocity, and duration values for the opening and closing gestures for the three subjects on the four conditions (accented, rescaled accented, truncated accented, and unaccented) are shown in Figs 4–6. The displacement values for the opening/closing gestures were calculated as the absolute differences between peak jaw displacements associated with the initial/final consonants and the medial vowel (see Fig. 1).

Focusing firstly on displacement, Figs 4–6 show that the magnitude of displacement in the opening and closing gestures was greater for accented than unaccented vowels for all three subjects: the results of a one-way ANOVA (Table I) showed that the displacement values were significantly greater for accented than unaccented vowels in both the opening and closing gestures for two subjects (CR and JF), but only in the opening gesture for subject KC.

With regard to peak velocity (middle panels, Figs 4–6), subjects CR and JF had

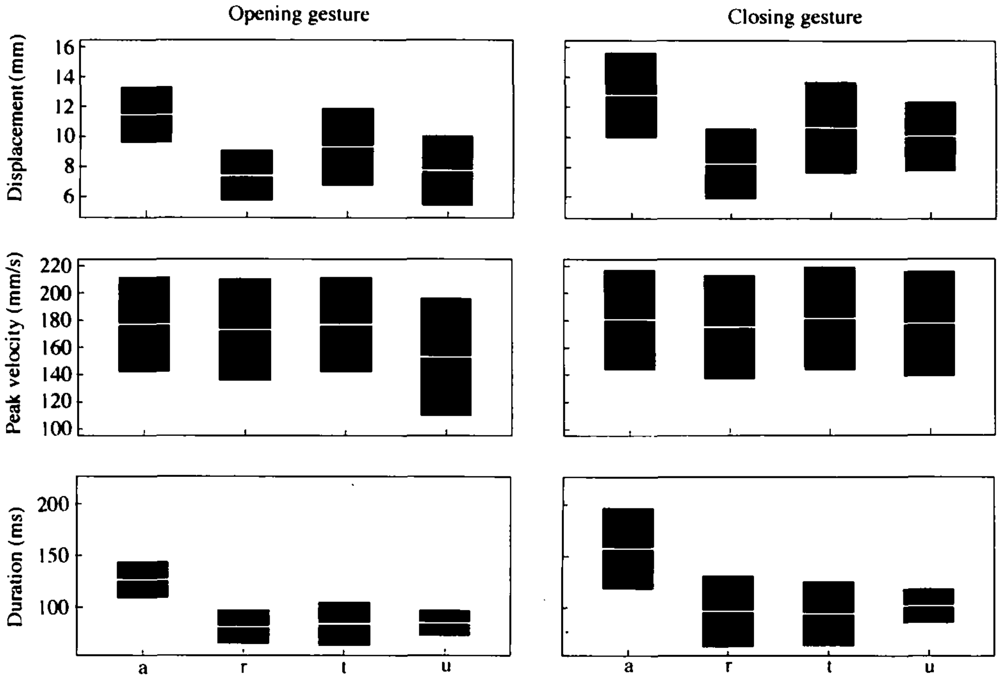

Figure 5. Displacement, peak velocity, and duration of the opening and closing gestures for subject JF. *a*: accented; *r*: rescaled accented; *t*: truncated accented; *u*: unaccented. The boxes represent two standard deviations on either side of the mean, which is shown as a horizontal white line.

significantly greater peak velocity values for accented vowels in the opening gesture, and non-significant differences between accented and unaccented vowels in the closing gesture. Subject KC's peak velocity was also significantly greater for accented vowels in the opening gesture, and significantly greater for *unaccented* vowels in the closing gesture (Table I). Finally, for all three subjects, the duration of accented vowels was significantly greater than that of unaccented vowels in both the opening and closing gestures (bottom panels, Figs 4–6 and Table I). The general pattern of all these results corresponds quite closely to that of Edwards *et al.* (1991).

Turning now to the main focus of the study, which is to compare unaccented with rescaled accented and truncated accented vowels, the results from the first two parameters described in the simulation study (proportion of the syllable between the peak velocities; proportion of the syllable underestimated) are displayed as ellipses for truncated, rescaled, and (unmodified) unaccented vowels in Fig. 7. In all cases, the ellipses span 4.9 standard deviations along the major and minor axes, and thereby include at least 95% of the data points.

For all three talkers, the ellipses for the unaccented vowels are closer to those of the truncated accented vowels, than to those of the rescaled accented vowels. In order to quantify the relative distances between the ellipses further, for each talker separately, Bayesian distances were calculated from each unaccented token to the centroids of the truncated and rescaled ellipses (thus $n \times 2$ calculations per talker, here n is the number of unaccented vowels). For all three talkers, the Bayesian distance to the truncated centroid was significantly less than to the rescaled centroid

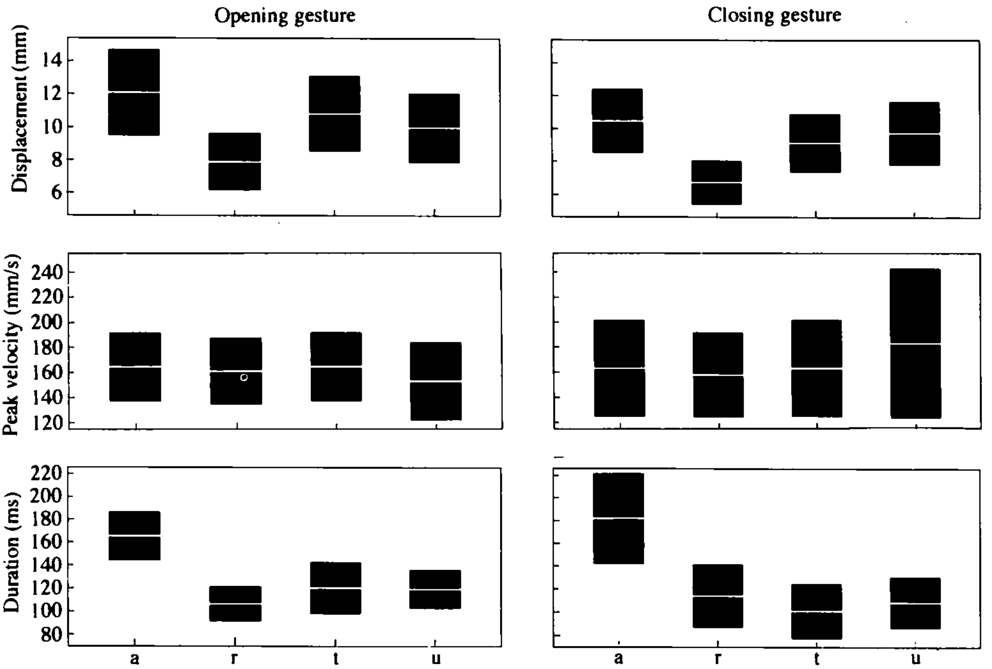

Figure 6. Displacement, peak velocity, and duration of the opening and closing gestures for subject KC. *a*: accented; *r*: rescaled accented; *t*: truncated accented; *u*: unaccented. The boxes represent two standard deviations on either side of the mean, which is shown as a horizontal white line.

TABLE I. Analyses comparing accented with unaccented vowels on displacement, peak velocity, and duration in the opening and closing gestures

	Opening	Closing
Subject CR		
Displacement	$F(1, 24) = 65.2$**	$F(1, 24) = 29.0$**
Peak velocity	$F(1, 24) = 37.5$**	$F(1, 24) = 1.0$
Duration	$F(1, 24) = 17.4$**	$F(1, 24) = 24.5$**
Subject JF		
Displacement	$F(1, 28) = 95.9$**	$F(1, 28) = 34.2$**
Peak velocity	$F(1, 28) = 11.3$**	$F(1, 28) = 0.2$
Duration	$F(1, 28) = 235.5$**	$F(1, 28) = 104.0$**
Subject KC		
Displacement	$F(1, 26) = 22.9$**	$F(1, 26) = 4.2$
Peak velocity	$F(1, 26) = 4.6$*	$F(1, 26) = 4.5$*
Duration	$F(1, 26) = 166.8$**	$F(1, 26) = 147.9$**

Significance levels: ** $P < 0.01$; * $p < 0.5$.

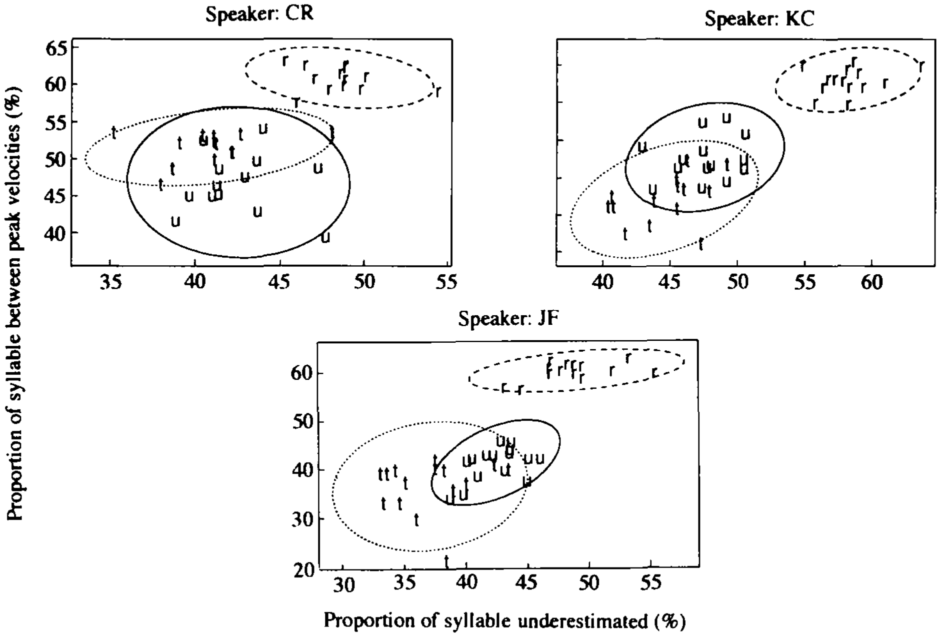

Figure 7. Ellipse plots fitted separately to truncated (*t*), rescaled (*r*), and unaccented (*u*) vowels. The ellipses include at least 95% of the data points. (——), unaccented; (· · · · · ·), truncated; (- - - -), rescaled.

(CR: $t = 2.24$; $p < 0.05$; d$f = 24$. KC: $t = 3.59$; $p < 0.002$; d$f = 26$; JF: $t = 3.81$; $p < 0.001$; d$f = 28$).

The third parameter to be considered is the shape of the acceleration waveform at the peak displacement point for the vowel. Fig. 8 shows averaged acceleration waveforms for the unaccented, truncated and rescaled data for each talker.

These waveforms were derived in the following way. Firstly, the separate jaw displacement waveforms were aligned, in both amplitude and time, at the peak displacement time for the vowel, and the aligned waveforms were double-differenced to obtain the corresponding acceleration waveforms. Secondly, the values of the acceleration waveforms were averaged at equal time intervals (step size = 5 ms) to obtain three averaged waveforms for the truncated, rescaled, and unaccented data ($t = 0$ ms in Fig. 8 corresponds to the point of peak jaw lowering for the vowel).

Following the reasoning in Section 2.4., truncated accented vowels should show a clear peak near $t = 0$ ms, but there should be no such peak for rescaled accented vowels. This difference is evident for talkers CR and JF (top left and bottom left panels of Fig. 8), but the peak is smaller at $t = 0$ ms in the truncated waveform for talker KC (top right panel).

Concerning the unaccented vowels, their averaged waveforms correspond more closely to those of the truncated vowels for both talkers CR and JF. For both these talkers, the unaccented acceleration waveform rises to a peak, although the peak occurs later than that of the truncated vowel for talker CR. For talker KC, the

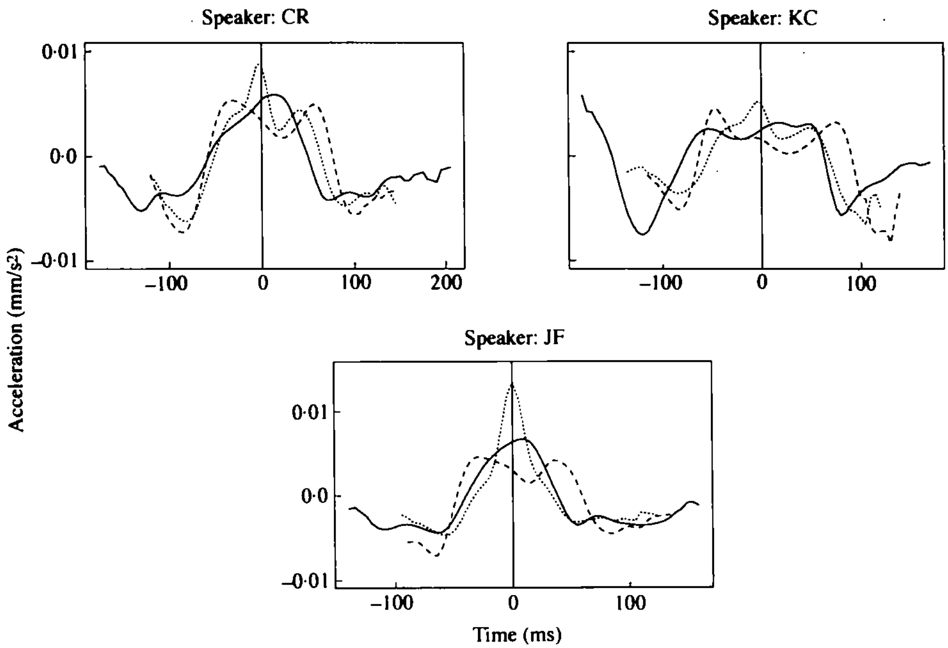

Figure 8. Averaged acceleration waveforms for truncated, rescaled, and unaccented vowels. The vertical line in each panel corresponds to $t = 0$ ms, the time of the peak jaw lowering for the vowel. Key as for Fig. 7.

unaccented waveform shows a minor peak at $t = 15$ ms, but in general its resemblance to the truncated waveform is considerably less clear than for the other two talkers.

In order to quantify more precisely the relationship between the different kinds of waveforms, a measure was obtained of how close (in time) the biggest acceleration peak was to the time of peak jaw lowering for the vowel. Only acceleration peaks which occurred between the peak velocity points were considered (Fig. 9, top left panel). This measure was calculated separately for all truncated, rescaled, and unaccented vowels of the three talkers.

The remaining three panels of Fig. 9, which show fitted normal curves of the absolute difference between the time of the biggest acceleration peak and the time of the peak vowel displacement, suggest that the unaccented vowels pattern more closely with the truncated vowels than the rescaled vowels for all three talkers. Once again, a calculation was made of the Bayesian distances from the unaccented vowels to the means of the truncated and rescaled vowels. The results showed that unaccented vowels were closer on this parameter to the truncated mean than to the rescaled mean for talkers CR and JF, but not for talker KC (CR: $t = 5.34$, $p < 0.001$, $df = 24$; JF: $t = 2.17$, $p < 0.05$, $df = 28$; KC: $t = 0.63$, $p > 0.5$, $df = 26$). For two of the three talkers, therefore, the evidence suggests that the shape of the acceleration waveform close to the peak vowel displacement corresponds more closely to a truncated accented waveform, than to a rescaled accented waveform.

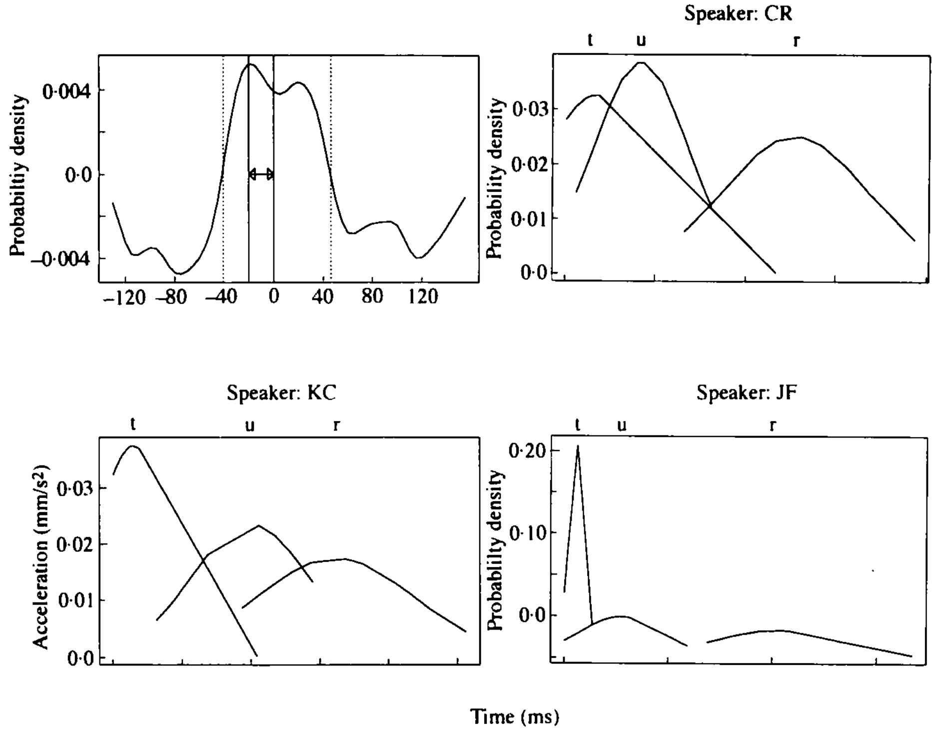

Figure 9. The top left panel illustrates the procedure that was used to parameterise further the acceleration data. The biggest acceleration peak (left solid line) was located between the peak velocity times (vertical dotted lines), and its absolute duration calculated from $t = 0$ ms (right solid line). The remaining three panels show, for each of the three talkers, the normal curves that were fitted separately to the truncated, rescaled, and unaccented vowels on this parameter.

4. Summary and conclusions

Three parameters have been proposed for distinguishing between truncated and rescaled vowels, which are all based on differences in the predicted shape of the movement waveform between the time points of the peak velocities. The parameters are: the proportion of the syllable which occurs between the peak velocity time points; the ratio between predicted and actual articulatory duration; and, the shape of the acceleration waveform at the time of peak vowel lowering. Truncated accented vowels were shown to be clearly different from rescaled accented vowels on the first of these two parameters, and for two of the three talkers on the third parameter. The study has also shown that unaccented vowels resemble more closely truncated, than rescaled, vowels, on the first two parameters for all three talkers; on the third parameter, the correspondence between truncation and unaccented vowels only emerged for two out of three of the talkers.

The study has provided some evidence in favor of the view that one form of coarticulation can be understood as the coproduction of autonomous opening and closing gestures. These gestures can be timed relatively to each other in different ways: as the temporal overlap between the gestures increases, acoustic and articulatory vowel duration decrease and the opening gesture is truncated by the closing gestures—i.e., the resulting movement waveform has a progressively smaller displacement amplitude. This explanation therefore establishes a relationship between coarticulation, vowel duration, and target undershoot, which, as Beckman, de Jong, Jun & Lee (1992) have noted, bears many similarities to Lindblom's (1963) classic analysis of vowel reduction and undershoot.

The present study also validates to a certain extent the explanation by Beckman and her colleagues that accented/unaccented differences can be modelled in terms of coarticulatory differences—or more specifically, that the progression from an accented to an unaccented vowel can be better understood as a change in the relative timing of articulatory gestures, than as a change to their underlying amplitudes. Under the relative timing theory, the progression from an accented to an unaccented vowel corresponds to an increasing overlap between the opening and closing articulatory gestures. The apparent reduction in the displacement of the vocal tract gestures is therefore not brought about by instructions to decrease the extent of supralaryngeal articulator movement, but is, instead, a consequence of the earlier timing of the closing gesture which "cuts off" the maximum extent of the opening gesture.

Although the results of the present experiment validate a model of truncation as an explanation of accented/unaccented vowel differences, a further consideration of the raw data in Figs 4–6 might suggest that unaccented vowels can be modelled by a *combination* of the truncation and rescaling effects discussed in this paper. This is because for two of the three talkers (CR and JF), jaw displacement values for the unaccented vowels were smaller than those that would be predicted by the truncation model, and are in fact sometimes closer to those of the rescaled accented vowels (see the top two panels of Figs 4 and 5).

Although these data might suggest that unaccented vowels conform to a combined truncation and rescaling model, we would like to consider an alternative explanation of the data in Figs 4 and 5 that takes into account the acceleration data in Fig. 8. These acceleration trajectories show that, although unaccented vowels are characterised by acceleration peaks close in time to the maximum jaw lowering for the vowel, the peaks are not as sharp, nor as high as those of the truncated accented vowels. Once again, this might be evidence that the production of unaccented vowels can be explained by a mixed truncation-rescaling model. Alternatively, the peaks for the unaccented vowels may be flatter because of limitations on the jaw's capacity for acceleration. More specifically, the simulated truncated vowels have high and peaked acceleration values because the opening and closing gestures tend towards a triangular shape in which the quasi steady-state jaw trajectory corresponding to part of the vowel target is cut out; and furthermore, as discussed earlier, a triangular shaped displacement trajectory necessarily implies high and peaked acceleration values. In reality, the jaw may never be able to attain the high acceleration values which correspond to the sharply peaked displacement trajectories of the truncation data. One possibility, therefore, is for the theoretically derived peak to be undershot as a smoothed trajectory as shown in Fig. 10.

This explanation can now account for the smaller displacement values for unaccented vowels compared with truncated accented vowels observed in the raw data in Figs 4 and 5. That is, because of the inertia of the jaw, the (sharp)

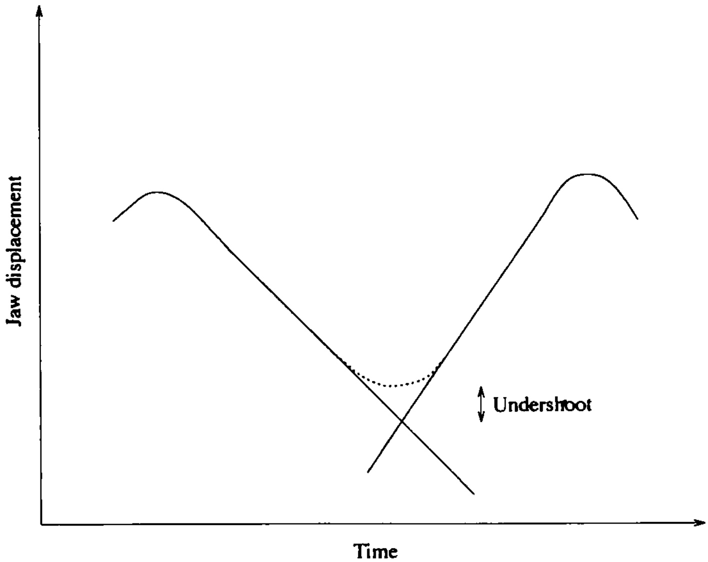

Figure 10. Schematic outline of the jaw trajectory of a truncated accented vowel (——) and the predicted actual trajectory of an unaccented vowel (· · · · · ·).

displacement peak that is predicted by the model of truncated accented vowels is undershot, resulting in a smoother movement trajectory, as shown in Fig. 10.

In summary, our model of the differences between unaccented and accented vowels incorporates two principal stages. Firstly, the closing gesture is timed to occur earlier relative to the opening gesture resulting in truncation of the central part of the jaw trajectory, close to the vowel target; secondly, this truncated trajectory is itself slightly undershot due to the limitations of the jaw to track a sharply peaked displacement trajectory which is derived from the preceding stage of truncation. There are various assumptions that underlie this model of unaccented vowels, the most important of these being that we have only demonstrated that unaccented vowels are more like truncated accented vowels than linearly rescaled vowels; the possibility remains, of course, that other articulatory models, which we have not simulated and tested, may capture the accented/unaccented distinction just as effectively. Another possibility is that, since one of our carrier phrases included a contrast between *barb* and *bub* (which are contrasted in Australian English primarily by the length of the vowel), the truncation effects that have been observed in our paper may be partly attributable to this long/short vowel opposition, as well as to the accented/unaccented distinction that has been tested (see also de Jong, 1991). Clearly, subsequent studies should be directed towards investigating the extent to which other forms of vowel shortening (e.g., phonemic long/short vowel oppositions, rate, vowel shortening before voiceless consonants) can be explained by the kind of model of truncation that has been proposed here.

Our thanks to Ken de Jong and an anonymous reviewer for many helpful comments on this paper, as well as to Steve Cassidy for earlier discussions of the data and to Chris Callaghan for technical assistance. This research was supported by Australian Research Council grant AC9330706.

References

Beckman, M. E., Edwards, J. & Fletcher, J. (1992) Prosodic structure and tempo in a sonority model of articulatory dynamics. In *Paper in Laboratory Phonology II: Gesture, Segment, Prosody* (G. J. Docherty & D. Robert Ladd, editors), pp. 68–86. Cambridge University Press: Cambridge.

Beckman, M. E., de Jong, K., Jun, S.-A. & Lee, S.-H. (1992) The interaction of coarticulation and prosody in sound change, *Language and Speech*, **35**, 45–58.

Beckman, M. E. & Ayers, G. (1994) Guidelines for ToBI labelling (version 2.0). Unpublished manuscript released with the materials from the Tones and Break Indices Workshop, Columbus Ohio (July, 1993).

Bernard, J. (1970) Towards the acoustic specification of Australian English, *Zeitschift für Phonetik*, **23**, 113–128.

Boyce, S. (1990) Coarticulatory organization for lip rounding in Turkish and English, *Journal of the Acoustical Society of America*, **88**, 2584–2595.

Browman, C. P. & Goldstein, L. (1989) Articulatory gestures as phonological units, *Phonology*, **6**, 205–251.

Browman, C. P. & Goldstein, L. (1990) Tiers in articulatory phonology with some implications for casual speech. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (J. Kingston & M. E. Beckman, editors), pp. 341–376. Cambridge University Press: Cambridge.

Bullock, D. & Grossberg, S. (1988) The VITE model: A neural command circuit for generating arm and articulator trajectories. In *Dynamic Patterns in Complex Systems* (J. A. S. Kelso, A. J. Mandell and M. S. Shlesinger, editors) pp. 305–326. World Scientific Publishers: Singapore.

Crystal, T. H. & House, A. S. (1988) The duration of American-English vowels: an overview, *Journal of Phonetics*, **16**, 263–284.

De Jong, K. (1991) An articulatory study of consonant-induced vowel duration changes in English, *Phonetica*, **48**, 1–17.

De Jong, K., Beckman, M. E. & Edwards, J. (1993) The interplay between prosodic structure and coarticulation, *Language and Speech*, **36**, 197–212.

Edwards, J., Beckman, M. E. & Fletcher, J. (1991) The articulatory kinematics of final lengthening, *Journal of the Acoustical Society of America*, **89**, 369–382.

Fletcher, J. & Vatikiotis-Bateson, E. (1990) Prosody and intrasyllabic timing in French, *Proceedings of the 3rd Australian Speech Science and Technology Conference*, pp. 318–323.

Fowler, C. A. (1980) Coarticulation and theories of extrinsic timing, *Journal of Phonetics*, **8**, 113–133.

Fowler, C. A. (1981) A relationship between coarticulation and compensatory shortening, *Phonetica*, **38**, 35–50.

Fowler, C. A. (1983) Converging sources of evidence on spoken and perceived rhythms in speech: cyclic productions of vowels in monosyllabic stress feet, *Journal of Experimental Psychology: General*, **112**, 386–412.

Fowler, C. A. (1984) Segmentation of coarticulated speech in perception, *Perception and Psychophysics*, **36**, 359–368.

Fowler, C. A. (1986) An event approach to the study of speech perception from a direct-realist perspective, *Journal of Phonetics*, **14**, 3–28.

Fowler, C. A. & Rosenblum, L. D. (1989) The perception of phonetic gestures, *Haskins Laboratories Status Report on Speech Research*, **SR-99/100**, 102–117.

Fowler, C. A. & Saltzman, E. (1993) Coordination and coarticulation in speech production, *Language and Speech*, **36**(2, 3), 171–195.

Fowler, C. A. & Smith, M. (1986) Speech perception as "vector analysis": an approach to the problems of segmentation and invariance. In *Invariance and Variability in Speech Processes* (J. Perkell & D. Klatt, editors), pp. 123–136. Academic Press: New York.

Harrington, J., Cassidy, S., Fletcher, J. & McVeigh, A. (1993) The mu+ system for corpus based speech research, *Computer Speech and Language*, **7**, 305–331.

Hawkins, S. (1992) An introduction to task dynamics. In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (G. J. Docherty & D. Robert Ladd, editors), pp. 9–25. Cambridge University Press: Cambridge.

Klatt, D. H. (1975) Vowel length is syntactically determined in a connected discourse, *Journal of Phonetics*, **3**, 129–140.

Lindblom, B. (1963) Spectrographic study of vowel reduction, *Journal of the Acoustical Society of America*, **35**, 1773–1781.

Lindblom, B. & Rapp, K. (1973) Some Temporal Regularities of Spoken Swedish. Publication No. 21, Institute of Linguistics, University of Stockholm.

Löfqvist, A. (1990) Speech as audible gestures. In *Speech Production and Speech Modelling*, (W. J. Hardcastle & A. Marchal, editors), pp. 289–332. Kluwer Academic Publishers: Dordrecht.

Mitchell, A. G. & Delbridge, A. (1965) *The Speech of Australian Adolescents*. Angus and Robertson: Sydney.

Munhall, K., Fowler, C. A., Hawkins, S. & Saltzman, E. (1992) "Compensatory shortening" in monosyllables of spoken English, *Journal of Phonetics*, **20**, 225–239.
Öhman, S. E. G. (1966) Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America*, **39**, 151–168.
Saltzman, E. & Munhall, K. (1989) A dynamical approach to gestural patterning in speech production, *Ecological Psychology*, **1**, 333–382.
Umeda, N. (1975) Vowel duration in American English, *Journal of the Acoustical Society of America*, **58**, 434–445.