3

Manner and place conflicts in the articulation of accent in Australian English

JONATHAN HARRINGTON, JANET FLETCHER AND MARY E. BECKMAN

3.1 Introduction

Early experimental work on acoustic correlates of focal prominence in English and prosodically similar languages showed a leading role for pitch accent placement (e.g. Bolinger 1958, Bruce 1977), and we now know a great deal about the role of pitch target values in the cueing of relative prominence among accented words, particularly if the associated accents involve rises to or falls from peak targets for high tones (e.g. Terken 1991). At the same time, it has become increasingly clear that association with accent in these languages differs from the tonally analogous phenomenon in languages such as Japanese in that accent placement in the stress-accent languages can be accompanied by other, supralaryngeal markers of the rhythmic prominence of the associated syllable in the accented word (e.g. Beckman 1986). The more recent experimental literature on these supralaryngeal concomitants of accenting suggests that speakers have recourse to two different prominence-enhancing strategies for the production of accentually prominent syllables. On the one hand, studies of mandibular kinematics generally show larger and faster excursions into and out of accent-bearing syllables, which achieve significantly lower jaw positions at peak opening (e.g. Summers 1987). Beckman, Edwards & Fletcher (1992) interpreted these results in terms of a ‘sonority expansion’ strategy. The speaker intends to produce a louder vowel in the accented syllable, and does so in part by adjusting supralaryngeal articulation to make for a more open oral passage over a longer period of time. On the other hand, studies of lingual kinematics generally show a lower tongue body position in accented syllables only for low vowels, with non-low front vowels having higher tongue body and non-low back vowels having a backer tongue body (e.g. Engstrand 1988, de Jong 1995a). These results suggest a different strategy for enhancing accentual prominence, which de Jong (1995a) calls localized hyperarticulation, after Lindblom's (1990a) H&H model of a continuum of overall speech clarity. The speaker intends to produce a more peripheral vowel and thus make a clearer differentiation from any other vowel that might have occurred in the same position.

In a syllable with a low vowel nucleus, these two accounts are mutually compatible. If the general strategy of increasing the clarity of contrastive specifications for the accented syllable is understood more broadly to target structural contrasts as well as paradigmatic ones, then a lower jaw enhances the clarity of the manner contrast between the vowel and a preceding consonant, thus providing a clearer sonority rise from the edge of the syllable to its vocalic nucleus. Also, a more peripheral vowel enhances clarity not just of the paradigmatic contrast between the low vowel and any other full vowel that could have occurred in the accented syllable, but also of the prosodic contrast with any adjacent unaccented syllable, particularly if the nucleus of the unaccented syllable is a reduced vowel. Thus, if we consider how prosodic structure is parsed from the succession of vowel and consonant specifications, we might think of sonority expansion as a variant of localized hyperarticulation—namely, a hyperarticulation of the manner feature specification that makes a low vowel the ideal nucleus for a stressed syllable.

In other cases, however, enhancing the clarity of the sonority rise into the vowel seems incompatible with enhancing the paradigmatic specification of the accented vowel's place target. For example, when the syllable contains the vowel [i] after a stop, a more open oral tract that increases the sonority contrast between the vowel's manner specification and that of the preceding consonant cannot be produced simultaneously with a narrower palatal constriction that hyperarticulates the vowel's contrastive oral place specification as [+front] and [+high]. In this paper, we re-examine the sonority expansion and localized hyperarticulation accounts, looking at both mandibular and lingual articulations for both phonemically high and phonemically low vowels in Australian English. We explore how the articulatory differences relate to acoustic distinctions between prominence levels, paying particular attention to the timing of a localized hyperarticulation maneuver in the high vowel. We then speculate on the more general ramifications of the data for the phonological representation of vowel length and syllable weight.

3.2 Methods

The primary data are two sets of articulator movement records made thirteen months apart. The materials for the two recording sessions (henceforth corpus 1 and corpus 2) are the dialogue types illustrated in Table 3.1. These were intended to contrast nuclear accented (double underlined) with postnuclear and necessarily deaccented (single underlined) tokens of the same target word. Each corpus included other dialogues to vary the stressed vowel (and consonant place in corpus 2). In corpus 1, the one other dialogue interchanged *Babber* /bæb´/ with *Beaber* /biːb´/, and in corpus 2, the eight other dialogues replaced *Beeber* /biːb´/ by *Berber* /b‰ːb´/, *Barber* /baːb´/, *Deeder*, *Derder*, *Darder*, *Geeger*, *Gerger*, or *Garger* (although we report results only from the dialogues with high /iː/ or low /æ/ or /aː/ as the target vowel). We had native speakers of Australian English produce ten tokens of each dialogue at normal rate for corpus 2 and at each of slow and fast rates for corpus 1. We recorded up to five talkers for each corpus, and subsequent examination of the fundamental frequency patterns confirmed that all produced the intended intonational patterns around the target words. However, not all talkers produced clear non-tonal correlates of accentuation, and we report here only the results from two female talkers who participated in both recording sessions and who consistently produced a substantial lengthening of the vowel in the accented syllable relative to the unaccented. (One of these speakers is the second author of this paper.)

We used the MOVETRACK magnetometer system (Branderud 1985) to record four sets of vertical and horizontal positions from transducer coils attached to the chin, to the lower lip, to the tongue tip, and to the tongue dorsum about 4 cm behind the tip. Of these, the target parameters are jaw height (defined as the vertical position of the chin coil), and tongue height and fronting (the vertical and horizontal positions of the dorsum coil). The data were digitized directly to a SUN workstation and the ESPS/waves+ system was used for acoustic segmentation and labeling as well as to compute fundamental frequency, RMS amplitude, and formant frequencies. These data were read into the mu+ system to mark articulatory and acoustic extremum events and to use them in aligning and averaging the data traces. Differences were assessed using a two-level ANOVA function (which is equivalent to applying a *t*-test), and when results are reported as significant, the criterion is $p < 0.05$.

3.3 Results

We extracted the jaw height value at its extreme minimum within the target syllable in each token. Both talkers showed lower mean values for accented syllables. For speaker JMF, the difference was significant in all target words, and for MDB, it was significant in the target word *Deeder*, and in all bilabial contexts (i.e., in both target words in corpus 1 and in *Barber* and *Beeber* in corpus 2).

When the jaw is opened wider, there should be less energy lost to absorption by the vocal tract walls, as well as a larger radiating orifice at the lips. We evaluated this prediction by extracting the RMS amplitude level for each token at the time when the jaw height trace reached its extreme minimum value within the target vowel. The mean RMS value was significantly greater for the accented tokens in all target words for JMF's productions, and for all words except

Beaber in corpus 1 for MDB's productions. Although the RMS amplitude level is a very crude measure of loudness, this result suggests a heightened contrast in loudness between the vowel and the preceding stop closure—i.e., an expansion of the sonority rise that marks the nuclear mora of the accented syllable.

Table 3.1 *Speech materials. Tones above the target phrases are the expected intonation patterns, transcribed using ToBI*

	L+H* L+!H* L-	Corpus 1
A:	This is Laura Babber of the Babber-McDavis Company.	
	L+H* L+!H* L-	
	Could I speak to Dr Beaber, please?	
	L* H- L+H* L- L%	
B:	Do you want Dr Anna Beaber, or Dr Clara Beaber?	
	L+H* L+!H* L- L* L- H%	Corpus 2
A:	This is Hector Beeber, Colin. Would you order a nameplate for his desk?	
	H* L+H* L-	
B:	OK, should it be the plain "Hector Beeber", coloured blue on white,	
	L+H* L-	
	or the fancier "Dr Beeber", coloured gold?	

Since the tongue body rests on the jaw and the primary component of jaw opening is a rotational movement that swings the jaw backward as well as downward (Edwards & Harris 1990), we might expect the greater jaw lowering in accented syllables to expand the oral cavity at the expense of the pharyngeal cavity, and thus to result in a raising of first formant frequency as well as of RMS amplitude level. We extracted the F1 frequency value in each token at the time of the jaw opening extremum, and found that for JMF, the mean value was significantly higher in the accented context for all target vowels except the /a/ of *Garger*. For MDB, on the other hand, the mean F1 value at this point was not significantly higher in the accented syllable except in *Babber* in corpus 1.

Thus, for both the low and the high vowels, greater jaw opening in the accented syllable can be related to the strategy of sonority expansion. The mouth opens wider to reduce impedance in the oral cavity and to allow greater radiation of energy at the lips. This exaggerates the manner alternation between the vowel and preceding stop, to hyperarticulate the structural role of the sonority rise as marking the nucleus of a stressed syllable. For the low vowel targets in the *Babber* tokens produced by MDB and for all of the low vowels produced by

JMF, the lower jaw in the accented tokens can also be interpreted as a more local hyperarticulation of the syllable's paradigmatic feature specification. For these tokens, the more extreme jaw lowering in the accented syllable was associated with a raising of the first formant frequency which made the low vowel more peripheral in the vowel space. In the high vowel targets, on the other hand, raising the first formant makes the vowel less peripheral, and we must assess the other dimension of the vowel space to look for acoustic effects of hyperarticulation in /iː/.

Since listeners cannot separate perceptually the close second and third spectral peaks in [i] (cf. Chistovich 1985), the psychoacoustic consequence of raising either F2 or F3 should be the same percept of a sharpened high-frequency energy concentration. Syllables with labial or alveolar stops showed a consistent second formant peak somewhat after the middle of the vowel, whereas syllables with velar stops showed no F2 peak but a consistent third formant peak at about the same point. We extracted the both formant values at this F2 or F3 extremum time. In JMF's productions, F2 means were significantly higher for accented tokens of all three words, and F3 means were higher for *Beeber* and *Deeder*. In MDB's productions, F2 means were significantly higher in accented *Beeber* and *Deeder*, but not for *Geeger*. Thus there is evidence of increased peripherality of the vowel nucleus in labial and alveolar contexts for both talkers, and in the velar context for JMF.

In order to identify possible articulatory bases for these formant differences, we compared the horizontal and vertical positions of the tongue dorsum in accented and unaccented syllables. Figure 3.1 summarizes the data for the two talkers in all three consonantal contexts. It shows movement trajectories aligned for averaging at the acoustic 'target' (i.e., the F2 extremum in the top and middle panels and the F3 extremum in the lowest panels). This is the point marked 'T' on the plots. Note that there is usually a distinct elbow at this point, separating the path from the preceding consonant into the vowel target from the path back toward the following consonant. In the panels for JMF, the solid lines all lie below and to the left of the dotted lines, indicating a lower and more fronted tongue body at the formant target in the accented syllable. The mean differences in *x*-coordinate values at point 'T' are significant for all three target words. For subject MDB, on the other hand, the only significant difference in mean *x*-coordinate values is the fronter tongue position in the unaccented trajectory for *Beeber*, and the salient evidence of hyperarticulation is instead the significantly higher tongue at the F2 target for both the accented *Beeber* and the accented *Deeder* curves.

The lower tongue height in JMF's accented tokens accords with the greater jaw lowering and higher F1 values noted above. Fant's (1960) nomograms suggest that for a relatively close palatal vowel, increasing cross-sectional area at the constriction should have the effect of lowering the F2-F3 complex at the same time that it raises F1. The fronter tongue dorsum observed in JMF's panels in Figure 3.1 apparently more than compensated for this expected effect, to push the accented /iː/ toward the periphery of the vowel timbre space in the dimension that is orthogonal to this parameter of contrast with the preceding stop closure. MDB also made the accented /iː/ more peripheral in this dimension of the vowel space, but did so by bunching the tongue body further away from the jaw in accented tokens to decrease the cross-sectional area of the palatal constriction. Thus, the two speakers used different hyperarticulation maneuvers—in orthogonal dimensions of the articulatory space—to achieve the same sharpened timbre of the high vowel in the auditory space.

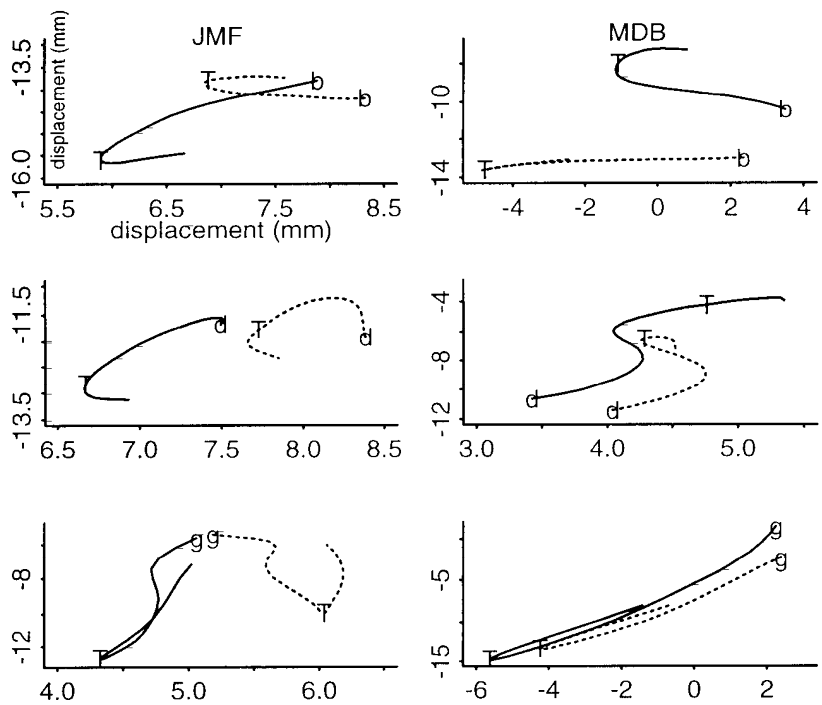

Figure 3.1 Averaged tongue dorsum trajectories for accented (solid lines) and unaccented syllables (dotted lines) produced by JMF (left panels) and by MDB (right). Trajectories were aligned for averaging at the F2 or F3 "target" (marked by 'T' in the plots). The beginning of the trajectory is marked by the phonetic symbol for the surrounding consonants. Decreasing values on the x- and y-axes correspond to more anterior and lower tongue positions, respectively.

The two speakers also differed in the relationship between the hyperarticulation and sonority expansion suggested by their tongue and jaw data. JMF seems to have reconciled the two prominence-enhancing strategies by assigning them to different articulatory and acoustic dimensions in just such a way that the listener can recover both the intent to lessen impedance in the vocal tract by opening the mouth wider and the intent to sharpen the timbre of the vowel by exaggerating the displacement of the tongue body away from a neutral resting position. MDB's hyperarticulation maneuver, on the other hand, was in the same dimension as the sonority contrast, but in the opposite direction, so that it undid any effect of the variation in jaw height on the F1 frequency at the time of the jaw opening extremum. However, it did not undo the effect on RMS amplitude. Perhaps MDB took advantage of another aspect of the vowel's paradigmatic specification to reduce the conflict between the two prominence-enhancing strategies. In Australian English, as in many other varieties, tense /iː/ contrasts with lax /i/ not just in length, but also in being a diphthong. It is specified to have a less peripheral onglide target followed by a more extremely palatal offglide target. If a speaker localized the hyperarticulation of the vowel's place specification even further, so that the effect is targeted narrowly for the later target, a palatal narrowing maneuver might be used to enhance the vowel's place specification without completely masking the acoustic effects of the sonority expansion.

Figure 3.2 gives some qualitative support for this interpretation of MDB's productions. It shows acoustic and articulatory traces for *Beaber*, aligned for averaging at the jaw extremum (t = 0 ms in the figure). At this point in the top panel, the F2 difference between the accented and unaccented traces is minimal. That is, the F2 extremum does not occur until later in the vowel—around the time of peak tongue dorsum raising in the middle panel, which is some 30–40 ms after the time of jaw height minimum value. Similar plots for MDB's *Beeber* and *Deeder* productions for corpus 2 show a comparable delay in reaching the F2 peak within the vowel. We evaluated this idea quantitatively by calculating the latency of the tongue height peak relative to the time of the jaw height minimum in all of MDB's productions of *Beaber* in corpus 1 and of *Deeder* in corpus 2. (As Figure 3.1 suggests, tokens of *Geeger* and of unaccented *Beeber* in corpus 2 could not be used in this evaluation because they did not show a dorsum peak within the vowel.) A comparison of this measure for the two accent conditions suggests that MDB actively exploits the sequencing of the peak palatal constriction after the syllable's sonority peak to reconcile the two prominence-enhancing strategies. Mean delay times were considerably later in the accented syllable. On average, the latency was 42.8 ms as opposed to 29.8 ms in accented versus unaccented *Beaber*, and 46.5 ms as opposed to 35.3 ms in accented versus unaccented *Deeder*. The difference is significant for the labial context, and almost significant for the alveolar context (where there were only half as many tokens).

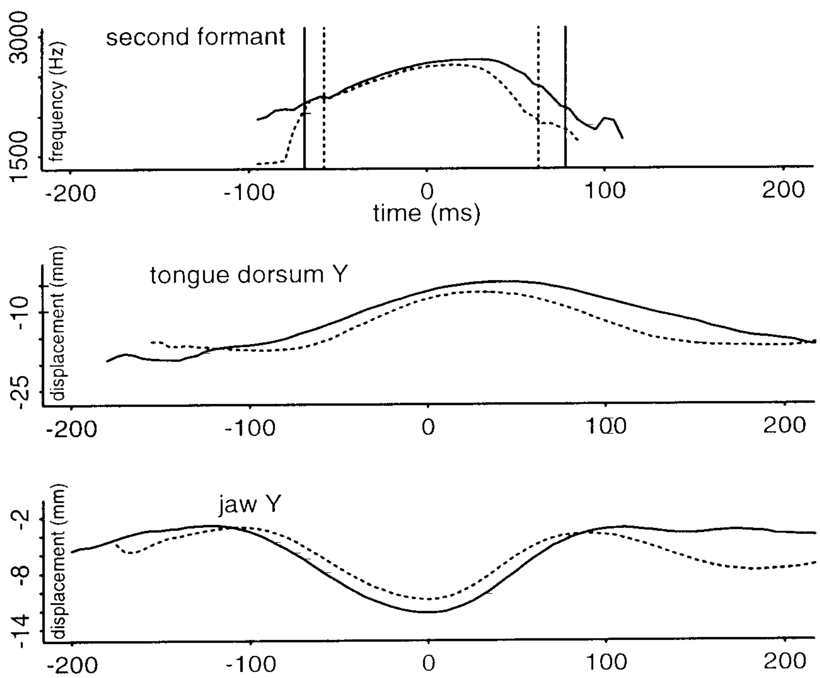

Figure 3.2 Averaged traces for F2 (top panel), tongue dorsum height (middle), and jaw height (bottom) in MDB's Beaber productions from corpus 1. All traces were aligned at the jaw displace-ment minimum prior to averaging. The vertical cursor lines in the formant trace panel show the average acoustic vowel onset and offset times relative to the alignment point. For all traces and cursors, solid lines indicate accented and dotted lines the unaccented tokens.

3.4 Discussion

The accented syllables in our data suggested two different ways in which the paradigmatic features of a high front vowel can be hyperarticulated. JMF created a sharper timbre by making a fronter palatal constriction, whereas MDB did so in the labial and alveolar contexts by making a narrower constriction instead of a fronter one. These two maneuvers are reminiscent of the two patterns of hyperarticulation seen in high and mid back vowels in de Jong (1995a, 1995b). For the word *put*, accented tokens had backer but not higher tongue body at the trajectory extremum point, whereas for *toes*, *toast*, *totes*, etc., accented tokens showed a backer and higher tongue body at the extremum point, which was consistently later than the vowel's midpoint. That is, just as in MDB's accented /iː/, the hyperarticulation in /oː/ by de Jong's Midwestern American speakers targeted the [w]-like offglide that is characteristic of the tense mid vowel of this variety of English. The timing of the hyperarticulation maneuver in MDB’s accented /iː/ is reminiscent also of the effects of focal accent on the articulation of long /iː/, /yː/, /ʉː/ and /uː/ in Stockholm Swedish, which “become diphthongized with a gesture toward a homorganic fricative, which usually causes a pronounced minimum in the intensity profile” (Fant 1991: 82).

We think these similarities are related to the diachronic trend for high and mid long vowels to break into rising diphthongs in the Germanic languages, and to the role that such long vowels and diphthongs play in the prominence hierarchy in these stress-accent languages. In both English and Swedish, the parsing of pitch accent placement is constrained by a contrast between ‘heavy’ and ‘light’ syllables that defines a lower level of stress. Associated consonant and vowel events must allow a syllable to be parsed as heavy if it is to be parsed as accent-bearing. Moreover, in both languages, open syllables containing long vowels are grouped together with closed syllables in parsing this weight contrast. In English, for example, part of the definition of a word is that it must be able to constitute a well-formed intonation phrase in isolation, which means it must be large enough prosodically to bear the obligatory nuclear pitch accent of the phrase. So heavy /biː/ and /bid/ are possible monosyllabic words, but light */bi/ is not. In syllables with non-low vowels, hyperarticulation of the height feature under accent can target the second mora. It thus can serve to make the first mora more sonorous by contrast to it and also brings out this prosodic equivalence between open syllables with long vowels and closed syllables with short vowels. That is, in MDB’s nuclear-accented /iː/ vowels in our study, in the accented /oː/ vowels in de Jong’s study, and in focally accented long high vowels in Swedish, the end of the vowel is articulated to be more consonant-like—hence, with a heightened sonority contrast to the beginning of the vowel.

In metrical phonology, these facts about the structural role of the offglide are captured by the device shown in Figure 3.3a. The prominence contrast at the lowest level of the stress hierarchy is represented by the differential association of vowel and consonant autosegments to structural units below the syllable, while the association of feet to syllables is determined by the number of morae (a foot can only associate to a heavy syllable i.e. one that dominates two morae). These structural units (mora, syllable, foot, prosodic word) are also used for the phonemic length contrasts of languages such as Japanese and Finnish (as shown in Figure 3.3b), and there are many similarities between the two sets of languages that argue for equating the weight contrasts to the length contrasts in this way. For example, in all four languages, a consonant does not participate in the mora count unless it can be parsed as associating after the first mora in the syllable.

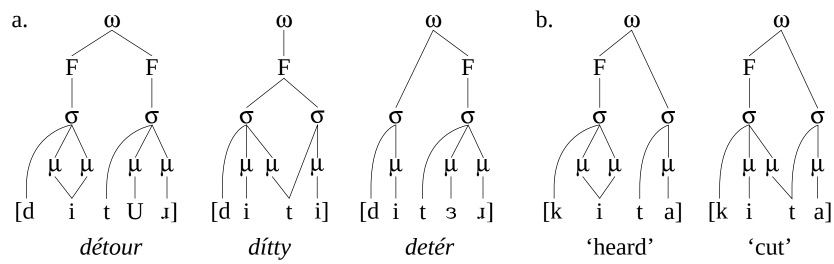

Figure 3.3 Prosodic tree representations of (a) English weight contrasts and (b) Japanese length contrasts (after Pierrehumbert & Beckman 1988)

However, there are also clear typological differences between the two sets of languages. For example, while both Japanese and Finnish have distributional constraints on tone placement that can be understood in terms of the notion of accent (see Pierrehumbert & Beckman 1988, Välimaa-Blum 1988), in neither language are pitch accents constrained to associate to bimoraic syllables. Finally, when we look at the diachrony of long vowels in Japanese and Finnish, we do not find the Germanic trend for high or mid long vowels to develop lower first targets. Instead, we find monophthongs developing higher onglides (/eː/ > /ie/, /øː/ > /yø/, and /oː/ > /uo/ in Eastern Finnish) or diphthongs becoming monophthongs (*ou > [oː] and *ei > [eː] in Tokyo Japanese).

In short, while the trees in Figure 3.3 are a convenient way to express the phonological universals, they do not fully predict the typological dependencies among size-related phenomena at different prosodic levels. This is because the arboreal phonological representation does not distinguish between the mora as a mere diacritic for syllable size and the more specialized use of the mora as a properly prosodic constituent—a place-holder in the syllable-internal dynamics of stress-related weight phenomena in the two stress-accent languages. That is, when the bimoraic notation expresses the stress contrast between heavy and light syllables, the first mora in the heavy syllables is the head of the syllable that is the head of the stress foot. By contrast, when the bimoraic notation expresses a pure length contrast, the first half of a long vowel does not have this status. To capture the typological differences, the trees need to be fleshed out by specifying the phonetic patterns that allow moras and syllables to be parsed from the speech signal.

To see how the phonology is grounded in potentially language-specific phonetic interpretations, consider how the prosodic tree is used to account for the fact that in many languages a consonant constriction gesture can be interpreted as increasing the mora count for the syllable that bears a preceding but not a following vowel target. The asymmetry can be represented by allowing a coda consonant to associate to the second mora of a syllable, but prohibiting association of an onset consonant below the syllable node, as illustrated by the consonant associations in Figure 3.3. However, there is nothing inherent to the representation that predicts these association patterns. To explain the asymmetry, we must refer instead to real-world events such as the specific consonant constriction and vowel target, and the phonetics for parsing the locations of a syllable's edges and head mora from the acoustic consequences of forming and releasing the consonant constriction. In a canonical stop-vowel syllable, the release of the stop into the vowel is the locus of a clear spectral edge, of a sharp rise in amplitude, of a rapid increase in first formant frequency, and of several other dynamic properties that help to demarcate the nuclear mora of the syllable as a prosodic subconstituent. While not all consonant-vowel pairs provide all of these properties, there are few pairs for which the acoustic consequences of forming the consonant constriction in the VC order mark the edge of a mora as clearly as do the consequences of releasing the consonant in the CV order. This means that perfectly alternating VCVCV sequences are more conducive to being parsed as open syllables than as closed. It also means that in languages where syllables can contrast in mora count, a second mora might be parsed from the signal if there is an increased interval before the next good sonority rise. In languages where the mora functions primarily as a length diacritic, the source of the increased interval may not be important just so long as the sonority rise into the first mora is no less steep than the subsequent fall. That is, for parsing a long (bimoraic) syllable, it may not matter whether the increased interval to the next sonority rise is due to the prolongation of the vowel target or to the elongation of the closure phase of a following stop. In Japanese, for example, this equivalence is used to define a relatively pure paradigmatic length contrast, in which the initial sonority rises in the long first syllables of [kiːta] or [kitːa] are not markedly different from that in [kita].

In languages such as English and Swedish, on the other hand, the phonetics for parsing a bimoraic syllable interacts with the parsing of structure at other levels of the hierarchy. A word can begin with long open first syllable, as in *detour*, or with a closed first syllable, as in *ditty*. However, it cannot begin with a short open first syllable unless the sonority rise is considerably reduced to contrast syntagmatically with a heavy second syllable, as in *deter*. In other words, in *detour* and *ditty*, the steep sonority rise and shallower sonority fall that delimit the initial mora of the first syllable function to signal that these syllables are heavy (i.e. stressed) and not just long. The sonority profile of this first mora must accord with its structural role as the head of a stress foot. It is not just its contribution to the syllable's mora count, but the position of a mora that is structurally significant. It is in this sense that we can say that the super-palatal off-glide in the long /iː/ in many dialects plays the same role in the sonority profile for the heavy syllable as the requisite consonant does in the syllable with the short /i/. We might conjecture that these higher offglides develop diachronically when a strategy for accommodating sonority expansion and [+high] hyperarticulation within the same accented long syllable is 'phonologized' as an inherent feature of the long vowel as the nucleus of the stressed syllable in a word. Conversely, in the synchronic phonology of dialects where this structural parse has been encoded in the production of the lexical contrast between long and short high vowels generally, one strategy for hyperarticulation under accent can then be simply to exaggerate that aspect of the long vowel's articulation that makes the syllable containing it inherently stressed.

Notes

This research was supported by two Australian Research Council small grants (1995–1996) to the authors of this paper and by a Macquarie University Visiting Research Scholarship (1995) to Mary Beckman. Preliminary results for corpus 1 were presented at the 131st Meeting of the Acoustical Society of America, and we thank Sallyanne Palethorpe (our coauthor on that talk) for her many contributions in the analysis of those data. Thanks also to Chris Callaghan for technical assistance with the MOVETRACK system; to Belinda Taylor, Joanne Arcuili, Megan Bradley, and Louise Loomes for segmenting and labelling the data; and to Steve Cassidy for the segmentation and label file organization.