Proceedings of the IRCS Workshop on Prosody in Natural Speech, August 5-12 1992

Pitch Accent Placement within Words

S. Shattuck-Hufnagel[1] M. Ostendorf[2] K. Ross[2]

October 29, 1992

[1]Massachusetts Institute of Technology, Cambridge, MA 02139 USA
[2]Boston University, Boston, MA 02215 USA

Abstract

Two aspects of prosodic structure have been suggested as factors contributing to the early placement of prominence, sometimes called 'stress shift', in late-main-stress words: these two factors are *rhythmic regularity* and *onset location of the pitch accent* in its prosodic phrase. This paper reports data from a corpus of FM radio news stories showing support for both factors. When listeners label phrase-level prominences for each syllable in an utterance, they tend to report that the speaker has placed a prominence early in a late-main-stress target word when either (a) the first syllable of the following word also bears a prominence (rhythmic clash) or (b) the target word carries the first prominence of the prosodic phrase (onset marking). We also examined some of the acoustic correlates of early prominence labeling in a subset of the same target words. When a syllable early in the word (i.e. before the main-stress syllable) is labeled with a phrase-level prominence, it shows a substantial F0 movement compared with the same syllable in non-early-accent examples. These findings support the hypothesis that apparent stress shift is, at least in part, a matter of early pitch accent placement within the word.

1 Introduction

For many years observers have noted that some words of American English can be produced with their major prominence on a syllable to the left of their main-stress syllable, under certain conditions. The set of words that can undergo this apparent 'stress shift' include those with a) main stress located late in the word, and b) a full-vowel syllable earlier in the word, which can carry the prominence; examples often cited in the literature include *Massachusetts* (as in "Massachusetts miracle"), *thirteen* (as in "thirteen men"), and *Japanese* (as in "Japanese food").[1] Dictionaries

[1]Gussenhoven suggests that not all pre-main-stress full-vowel syllables are candidates for stress shift [1].

sometimes capture this phenomenon by marking both unreduced vowels as main-stress vowels, or by putting a main-stress marker in parenthesis on the earlier vowel. Phoneticians occasionally described this phenomenon [2], but the first systematic treatment in the context of a general theory of prosody was by Vanderslice and Ladefoged in 1972 [3]. They suggested that in words like *telegraphic*, both of the full-vowel syllables should be marked as accentable, "leaving the determination of which accents are ultimately to be realized phonetically to the care of a late rhythm rule..." (p. 829). In subsequent work, the two major accounts of apparent shift have been based on rhythm and on intonation. We will briefly summarize these two approaches before presenting the results of our perceptual and acoustic analyses.

2 Phonological Accounts of Apparent Stress Shift

2.1 Rhythm-based theories

Liberman (1975) [4] gave the phenomenon of stress shift an important role as evidence for his theory of metrical phonology. Since then, a number of theoretical proposals have suggested that stress shift provides a persuasive argument in support of the metrical grid, a representational device which indicates the degree of stress on a syllable by the number of marked cells in the vertical column above that syllable in a nucleus-based matrix [5, 6, 7, 8, 9]. In these models, lexical stress corresponds to the marking of a certain number of cells in the column; when words are concatenated into a phrase and phrase-level stress is assigned, the resulting pattern of columns of cell markings in the grid may create a *stress clash*. That is, two adjacent or nearly-adjacent syllables may be rhythmically strong, so that a tendency toward the placement of heavy stresses at more equal intervals requires that one of them be moved. Observation has shown that the left-hand member of a pair of clashing stresses, which (according to the Nuclear Stress Rule of English) is normally the weaker one, appears to shift further left, away from the clashing (and stronger) stress on its right. An example that is often cited to illustrate the claims of this approach compares *the Mississippi legislator* with *the Mississippi legislation*. The former phrase embodies a stress clash, and so might be expected to undergo stress shift; the latter would not.

In these stress-based theories, rhythmic stress is viewed as a unitary phenomenon, whether it occurs at the lexical or the phrasal level; all x-marks in the cells of the grid are of the same variety, and differences in degree of stress are reflected simply in the number of such marks in the vertical columns. Phrase-level stress occurs on the main-stress syllables of words, except when this results in rhythmic irregularity, in which case it can move to an earlier syllable in its word. The acoustic correlates of the moved prominence are not specified.

2.2 Pitch-accent-based theories

Concurrent with these developments in phonological theory, relevant developments were also occurring in a different domain: models of intonation. In 1965, Bolinger [10] suggested that there were two main kinds of markers for phrasal prominence: one pitch marker or accent early in the phrase, and another late in the phrase. He noted that the tendency to put the first pitch accent as early as possible in the phrase might lead to its placement on a syllable to the left of the main-stress syllable, for late-stress words. In 1975, 't Hart and Collier [11] described the IPO model of intonation for Dutch, based on perceptual equivalence of simplified F0 contours, and noted that the common 'hat pattern' for declarative sentences included an 'onset rise'. Although they did not explicitly discuss the relevance of the onset rise to apparent stress shift, their model for Dutch reflects the same observation that Bolinger reports for English: the common occurrence of a pitch marker early in the phrase.[2]

Shattuck-Hufnagel (1988) [15], building on this observation, proposed an account of apparent stress shift in speech production planning. In this model, speakers have the option of placing pre-nuclear pitch accents on full-vowel syllables to the left of the main-stress syllable of a late-main-stress word, and they exercise this option with particular frequency when the target word carries the first pitch accent of its phrase. Subsequent expansions of this proposal [16, 17] have suggested that this tendency, combined with the location of phrase-final pitch accents on the main-stress syllable, might help the listener to identify the first and last pitch accents of a phrase.

In 1980, Pierrehumbert [18] proposed a grammar of intonation for American English that combined an account of pitch accent prominence with an account of the tonal markers of prosodic boundaries. In 1986 Beckman and Pierrehumbert [19] elaborated this model to include two kinds of prosodic phrases: the **intermediate phrase**, marked by at least one pitch accent and by the presence of a phrase accent that determines the intonation contour between the final pitch accent and the right boundary, and the **intonational phrase**, consisting of one or more intermediate phrases and marked by an additional boundary tone at its right boundary, realized on the final syllable of the phrase. Beckman and Edwards (forthcoming) [20], working in this framework, have suggested that apparent shift may occur because speakers tend to accent the first accentable syllable in a new intermediate phrase. Gussenhoven (1991) [1] has proposed a phonological model which reaches a similar result via a different mechanism. In his model, all accentable syllables receive an accent and the accents on alternate syllables are removed starting from the right-most accent. Monaghan (1990) [21] has also proposed an algorithm of alternate-accent deletion for prosody synthesis.

In all of these approaches, whether from the point of view of cognitive processing, phonological theory or synthesis algorithms, apparent stress shift is said to result

[2]The IPO model was subsequently extended to British [12, 13] and to American [14] English.

from a) the placement of a pre-nuclear pitch accent on the early full-vowel of a late-main-stress target word, plus b) the disappearance of the nuclear (or final) pitch accent from its main-stress syllable. For example, in the single-word phrase

```
Massachusetts
*     **
```

the pre-nuclear accent might occur on **Ma-** and the nuclear accent on **-chu-**. But in the longer phrase

```
The Massachusetts miracle
    *             **
```

the nuclear pitch accent no longer occurs on **-chu-**; by the Nuclear Stress Rule of English, it occurs on **mir-** in the following word. Since the pre-nuclear accents in both the single-word phrase and the longer phrase are optional, predictions about acoustic comparisons between the two cases are not straightforward; before testable predictions can be made, we must to determine which options the speaker selected for pitch accent location in the particular pair of utterances being compared.

In these pitch-accent-based theories, prominence is not viewed as a unitary phenomenon. Instead, the type of prominence (and its acoustic correlates) varies systematically with the level of constituent in the prosodic hierarchy; phrase-level prominence is a matter of pitch accents, whether those pitch accents occur on the lexically main-stressed syllable or on other full-vowel syllables in the word. Such theories predict that apparent stress shift will be associated with the kinds of F0 markers that normally signal a pitch accent.

Empirical studies relevant to the stress-based and pitch-accent-based approaches to apparent stress shift rely on two kinds of observations: judgments of prominence placement [22], and measurements of acoustic parameters believed to be correlates of perceived prominence, e.g. duration and F0 [23, 24]. Few studies have combined these two approaches [25]. As a result, if the theories do not predict with complete accuracy where shift will occur, the acoustic measurements that have been reported may not always correspond to actual cases of perceived shift. As part of a larger study of prosody, we have begun to address this problem in an extended analysis of a corpus of FM radio news stories. We have begun with perceptual labeling of perceived prominence, major boundary tones and the prosodic boundaries between each pair of adjacent words, as described in Price *et al.* (1992) [26]. Eventually we plan to have the corpus labeled for pitch accent type, boundary tones and break-index values, as described in [27], and to have acoustic analyses of segment and syllable duration and F0 patterns. Here we report results from the initial prominence and break index labeling effort for a portion of the corpus, and from acoustic analyses of a smaller subset of stress-shift candidate words, that are relevant to the issue of where pitch accents appear within words. The questions we asked were 1) Does onset position in the prosodic phrase, as well as rhythmic clash, influence the early

placement of pitch accents in the word? 2) Does the final pitch accent of the phrase occur on the main-stress syllable of its word? and 3) When a syllable is judged to have early accent placement, does it show a substantial F0 marker, compared to non-early-accent utterances of the same word?

3 Analysis of Radio News Speech

3.1 Methods

Both studies reported here are based on a corpus of recorded FM radio news broadcasts spoken by two female radio announcers. A total of twenty-five radio news stories (7880 words) were available for this study. The database is described in more detail in Ross *et al.*, (1992) [28] where some of these results are also presented. For the acoustic measurements, we focus on a small subset of this corpus: 44 instances of the early-accent candidate word *Massachusetts* that occurred in these stories. This data set is useful because it offers the control of a single word, and yet the word appears in naturally occurring paragraphs of speech, as opposed to isolated laboratory sentences. The stories were initially hand-labeled for phrasal prominences (syllables are labeled either prominent or unmarked) and seven levels (0-6) of prosodic boundaries, based on listening only. For this study, only the intermediate phrase boundaries (3 or higher) are relevant. The words in these stories that were candidates for early accent placement (i.e. late-main-stress words with an earlier unreduced vowel[3]) were then verified by listeners with access to visual displays of F0 contours, and in some cases additional pitch accents were labeled.

3.2 Study 1: Phrasal prominence location within words

We classified the words that were candidates for early accent placement according to their accent labeling. Table 1 shows the categories in columns 1, 2 and 3: accent on the early syllable only, on both the early and main-stress syllables and on the main-stress syllable only. A fourth category, for words that did not receive a phrasal prominence label, is omitted here. In addition, we classified the target words according to the serial position of their accent in the intermediate phrase: first accent (but not last), middle accent (neither first nor last), last accent (but not first), with a separate category for target words that contain all the accents in the phrase. Finally, we distinguished among different contexts: clash context (prominence on the initial syllable of the following word in the phrase), possible clash (prominence later in the following word) and no clash (no prominence on the following word). These distinctions allow us to ask the following questions.

1) Does early accent placement occur with greater frequency under conditions of rhythmic clash?

[3]See Shattuck-Hufnagel (1992) [16] for more detailed discussion of the criteria for words included as early accent candidates.

Accent Location in Phrase	Accent Location in Next Word		Early Accent	Two Accents	Main Stress	Total
			1	2	3	
First	First Syll.	A	20	3	6	29
	Non-first Syll.	B	3	4	8	15
	No Accent	C	12	7	13	32
Medial	First Syll.	D	12	1	2	15
	Non-first Syll.	E	5	2	7	14
	No Accent	F	6	0	10	16
Final	N/A	G	4	7	62	73
All	N/A	H	5	12	33	50

Table 1: Pitch accent placement in early-placement candidate words in a corpus of FM radio speech. Horizontal divisions show the serial position of target-word pitch accents in phrases (first (but not last), middle (neither last nor first), last (but not first), all). Rows within horizontal divisions show position of the pitch accent in the following word, reflecting clash context. Columns 1, 2 and 3 indicate the position of the pitch accent within the target word (early syllable only, main-stress syllable only, double accent). Deaccented words are not included.

In our terms, rhythmic clash occurs when the word following the late-main-stress target word has a phrasal prominence on its initial syllable. The entries in rows A and D, column 1, of Table 1 show that when the following word has a prominence on its initial syllable, early placement is likely to occur (32 of 44 words, or 73%). When the following word does not have initial prominence (rows B, C, E and F), early placement (column 1) is less likely to occur (26 of 77 words, or 34%). Clearly, early placement is more likely to occur under conditions of accent clash in this corpus.

2) Does early accent placement occur, even without a clash, when the accent is the initial accent in an intermediate phrase? n Comparing the entries in row C, columns 1 and 2, with row F, columns 1 and 2, we see phrase-initial accents occur early in their words in 59% of cases (19 of 32 words), while middle accents occur early in only 38% of cases (6 of 16 words.) For this comparison, we include the double-accented words, because the early-pitch-accent model gives the speaker the option of placing multiple accents on a single word. The stress-based account has more difficulty accounting for these instances, since the prominence on the early syllable is assumed to have been moved there from its former location on the main-stress syllable.

As others have pointed out [3], phonological theory suggests that phrase-final or nuclear accents will occur on the main-stress syllable of their words. The entries in rows G and H of Table 1 show that this is the case: 93% (114 of 123) of the

Data set	Early Accent	Two Accents	Main Stress	No Accents
Massachusetts	15	3	17	9
25 Stories	67	36	141	65

Table 2: Number of occurrences of different classes of accenting for early accent candidate words in two data sets.

phrase-final accents in this corpus appear on a main-stress syllable. We include the double-accented words here, because their second accent always appears on the main-stress syllable. Finally, when the target word contains all of the accents in its intermediate phrase (row H of Table 1), it often has two accents (24%, or 12 of 50, compared with 12% for the rest of the examples.) This finding supports the claim that speakers tend to place an onset accent on an early full-vowel syllable, and a final accent on a main-stress syllable of the words in an intermediate phrase.

3.3 Study 2: Acoustic correlates of early-accent syllables

If perceived stress shift is at least partly a matter of early pitch accent placement within the word, then syllables labeled with early prominence should show the acoustic correlates of a pitch accent. To test this prediction, we carried out acoustic analyses of 44 instances of the word *Massachusetts* that occurred in the labeled corpus. The phrases containing *Massachusetts* were analyzed interactively using the Klattools developed by Dennis Klatt for use on the Vax. Time offsets were obtained for each segment in the target word, and F0 contours were estimated using Klatt's method, which compares the spacing between individual harmonics in adjacent pitch periods. From these analyses we obtained 1) the size and direction of the largest continuously-measurable F0 change in each syllable, and 2) the duration of each syllable. Here we will compare these measures for the word-initial syllable *Ma-* in two types of utterances: those labeled with a prominence only on the main-stress syllable *-chu-* (17 examples) and those marked with a prominence on the early syllable *Ma-* (15 examples on *Ma-* only, 3 examples with double accent.) The distributions of accents within words for the larger set of early-accent candidates and for the subset of occurrences of *Massachusetts* are shown in Table 2.

F0 comparisons. One of the 18 examples of early accent labeling was omitted from this analysis because diplophonia made it difficult to estimate the F0 contour on the initial syllable of the target word. The remaining 17 showed a substantial F0 rise in the first syllable of the word. The extent of this rise ranged from 7 Hz to 86 Hz, with a mean of 44 Hz. A typical example is shown in Figure 1. In contrast, F0 in the initial syllable of non-early-accent examples did not show a consistent pattern. Here again, one utterance had to be omitted because of diplophonia. Of the remaining 16 examples, 6 had a rising F0, 7 showed a fall, one showed a fall-rise

and two showed little or no change. The absolute value of the largest continuous F0 change in these syllables, for those that had one, ranged from 6 to 32 Hz, with a mean of 16 Hz. A typical example is shown in Figure 2. For these two FM radio speakers, there appears to be no question that listeners' judgments of early prominence location corresponds to a substantial F0 marker of the kind that might be expected for a pitch accent. Corresponding syllables in utterances without an early accent do not have this F0 marker.

For all of the 17 examples with early pitch accent placement, including the three with double accent placement, the F0 pattern for the main-stress syllable *-chu-* showed a substantial fall (see Figure 1). The size of the fall ranged from 11 Hz to 46 Hz, with a mean of 23 Hz. Since the break index following these target words was 2 or less, it is unlikely that this fall in F0 was associated with a low boundary tone or phrase accent. Moreover, the F0 values for the unaccented main-stress syllable generally overlapped with those in the accented first syllable. In contrast, for the 16 examples with main-stress-only accent placement, the F0 of the accented main-stress syllable was substantially higher than in the first syllable; in most cases, the lowest measurable F0 in the accented syllable 3 was 20 or more Hz higher than the highest measurable F0 in the unaccented syllable 1 (see Figure 2). This contrast suggests that, if voicing had not been interrupted by the *-ch-* at the syllable onset, we might have seen a substantial pitch marker (i.e. a rise in F0) in the accented third syllable. The F0 of an accented syllable is not, of course always higher than for an unaccented syllable. But in this case it does suggest that the contrast between early-prominence-only and main-stress-prominence-only words lies in the placement of pitch accents.

Duration comparisons. Preliminary comparison of syllable durations for *Ma-* in *Massachusetts* showed no noticeable difference between the early-accent and non-early-accent cases. Since other analyses we have carried out both in this corpus and other corpora suggest that there may be a difference in duration between accented and unaccented syllables, the question requires further study.

4 Discussion

Results of our analysis of perceived prominence location and of syllable duration and F0 contours for 44 utterances of the word **Massachusetts**, as well as for the larger corpus of FM radio speech, show that early-prominence location in early-accentable words is influenced by two tendencies exhibited by speakers: to avoid the rhythmic clash that would arise if two pitch accents occurred on nearly-adjacent syllables in adjacent words, and to place the first accent of an intermediate phrase on the earliest possible syllable of its word. In addition, results show that this early prominence corresponds to a substantial rise in F0 on the prominent syllable, suggesting that apparent stress shift is, in many cases, the result of patterns of pitch accent placement.

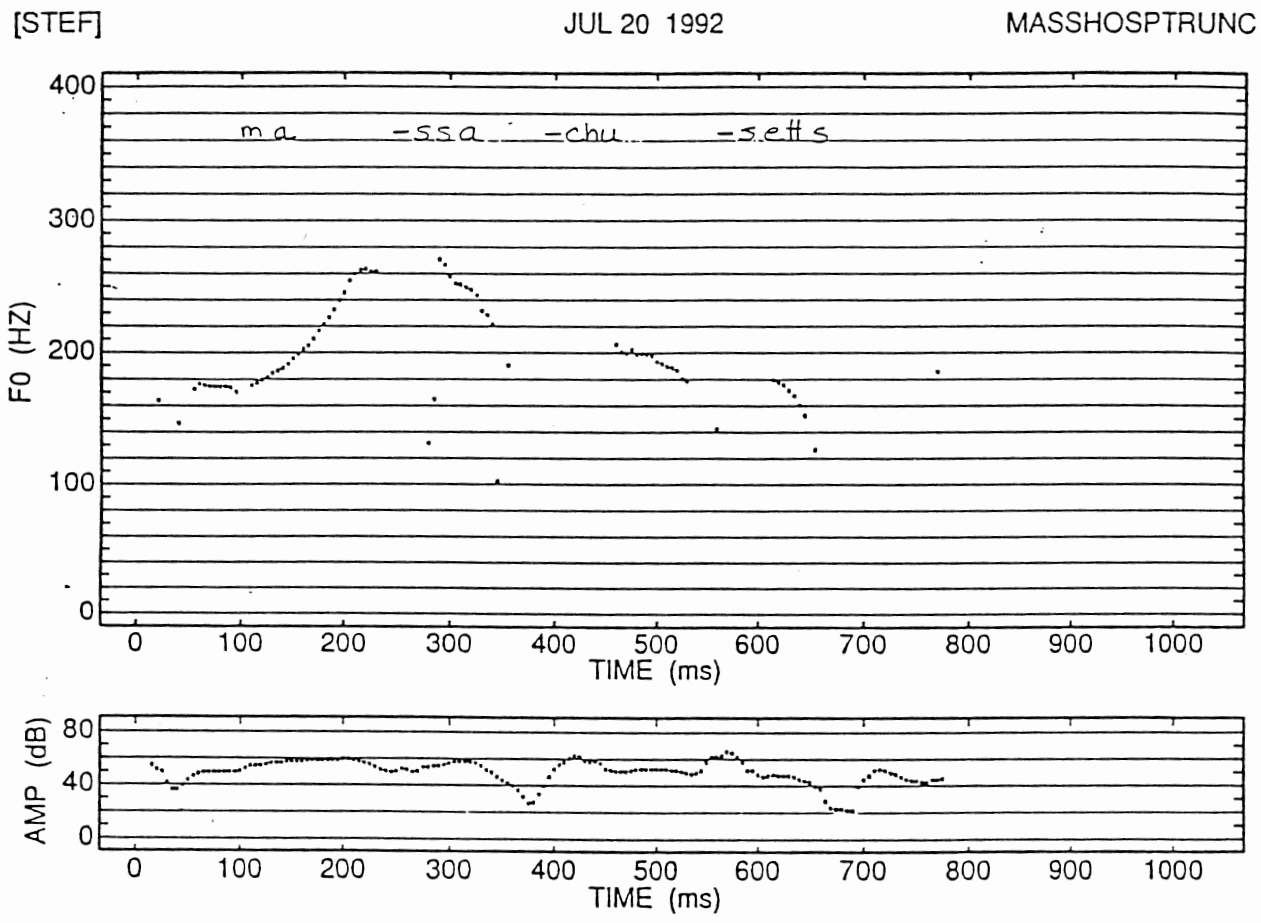

Figure 1: F0 contour for an example of the word *Massachusetts* labeled with an accent on the first syllable. Note the large F0 rise on the accented syllable *Ma-*. The small fall on the non-accented vowel *-u-* lies within the F0 values defined by the first-syllable rise.

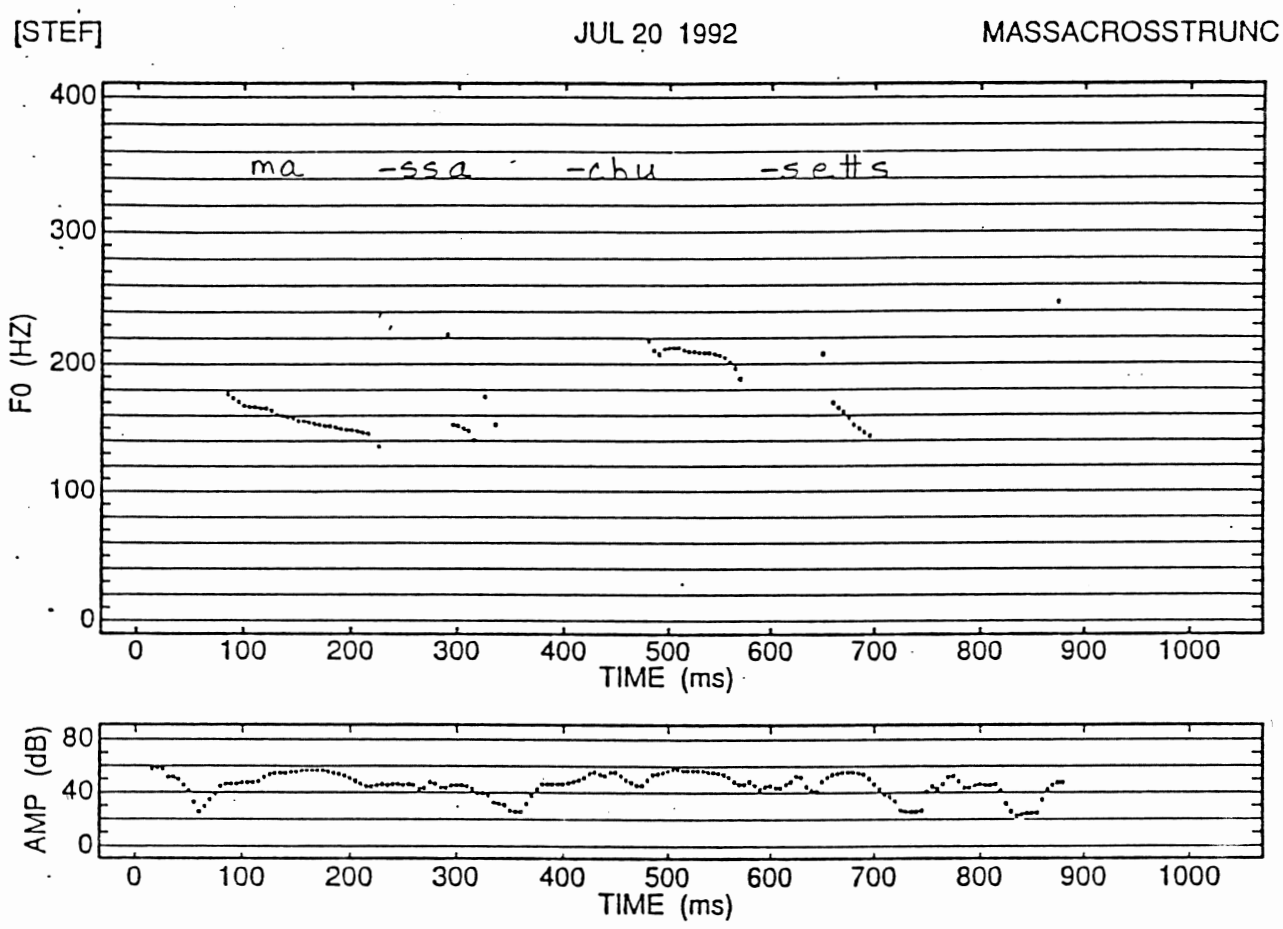

Figure 2: F0 contour for an example of the word *Massachusetts* labeled with an accent on the third (main-stress) syllable. The F0 fall on the non-accented *Ma-* is moderate in size. The F0 fall on the accented vowel *-u-* is also moderate but lies well above the F0 values for *Ma-*.

Acknowledgments

This research was funded by NSF under grant number IRI-8805680 with additional support coming from NIH, award number NIH 8-RO1-DCO-0075, and the Department of Education under the Graduate Assistance Applied to National Needs Program, award number P200A90080.

References

[1] C.Gussenhoven, "The English Rhythm Rule as an Accent Deletion Rule," *Phonology* 8, 1-35, 1991

[2] D. Jones, *An Outline of English Phonetics,* Cambridge: Heffer, 1917, 9th edition 1964

[3] R. Vanderslice and P. Ladefoged, "Binary Suprasegmental Features and Transformational Word-Accentuation Rules," *Language,* 48, 819-838, 1972

[4] M. Liberman, *The Intonational System of English,* Massachusetts Institute of Technology Ph.D. Dissertation, 1975

[5] M. Liberman and A. Prince, "On Stress and Linguistic Rhythm," *Linguistic Inquiry,* 8, 249–336, 1977.

[6] A. Prince, "Relating to the Grid," *Linguistic Inquiry,* 14, 19-100, 1983

[7] E. Selkirk, *Phonology and Syntax: The relation between sound and structure,* Cambridge, Mass.: MIT Press, 1984

[8] B. Hayes, "The Phonology of Rhythm in English," *Linguistic Inquiry,* 15, 33-74, 1984

[9] M. Nespor and I. Vogel, *Prosodic Phonology,* Dordrecht: Foris, 1986

[10] D. Bolinger, "Pitch Accent and Sentence Rhythm," *Forms of English: Accent, Morpheme, Order,* ed. I. Abe and T. Kanekiyo, Tokyo: Hokuou, 1965

[11] J. 't Hart and R. Collier, "Integrating different levels of intonation analysis," *Journal of Phonetics,* 3, 235-255, 1975

[12] J.R. de Pijper, *Modeling British Intonation,* Dordrecht: Foris, 1983

[13] N. Willems, *English Intonation from a Dutch Point of View,* Dordrecht: Foris, 1982

[14] S. Maeda, "A Characterization of Fundamental Frequency Contours of Speech," *MIT RLE Quarterly Progress Report 114,* 193-211, 1974

[15] S. Shattuck-Hufnagel, "Acoustic-Phonetic Correlates of Stress Shift," *Journal of the Acoustical Society of America,* 84, S98, 1988

[16] S. Shattuck-Hufnagel, "Stress Shift as Pitch Accent Placement," *Proceedings of the 1992 International Conference on Spoken Language Processing,* Banff, 747-750, 1992

[17] Shattuck-Hufnagel, S. "Stress shift as early pitch accent placement: a comment on Beckman and Edwards," *Proceedings of Labphon III,* ed. P. Keating *et al.,* forthcoming

[18] J. Pierrehumbert, *The Phonology and Phonetics of English Intonation*, Massachusetts Institute of Technology Ph.D. Dissertation, 1980

[19] M. Beckman & J. Pierrehumbert, "Intonational Structure in Japanese and English," *Phonology Yearbook 3,* ed. J. Ohala, 255-309, 1986

[20] M. Beckman and J. Edwards, "Articulatory Evidence for Differentiating Stress Categories," *Proceedings of Labphon III,*, ed. P. Keating *et al.,* forthcoming

[21] A. Monaghan, "Rhythm and stress-shift in speech synthesis," *Computer Speech and Language,* 4, 71-78, 1990

[22] M. Beckman, M.G. Swora, J. Rauschenberg and K. De Jong, "Stress Shift, Stress Clash and Polysyllabic Shortening in a Prosodically Annotated Discourse," *Proceedings of the 1990 International Conference on Spoken Language Processing,*, Kobe, Japan, 1990

[23] W. Cooper and S.E. Eady, "Metrical Phonology in Speech Production," *Journal of Memory and Language,* 25, 369-384, 1986

[24] M. Horne, "Empirical Evidence for a Deletion Formulation of the Rhythm Rule in English," *Linguistics* 28, 959-981, 1990

[25] S. Shattuck-Hufnagel, "Acoustic Correlates of Stress Shift," *Proceedings of the XII International Congress of Phonetic Sciences,* Aix-en-Provence, 1991

[26] P. Price, M. Ostendorf, S. Shattuck-Hufnagel and C. Fong, "The Use of Prosody in Syntactic Disambiguation," *Journal of the Acoustical Society of America,* 90, 2956-2970, 1991

[27] K. Silverman *et al.,* "A Standard Scheme for Labeling Prosody," *Proceedings of the International Conference on Spoken Language Processing,* Banff, 867-870, 1992.

[28] K. Ross, M. Ostendorf, S. Shattuck-Hufnagel, "Factors affecting pitch accent placement," *Proceedings of the International Conference on Spoken Language Processing,* Banff, 365-368 1992