ACOUSTIC CORRELATES OF STRESS SHIFT

Stefanie Shattuck-Hufnagel

Research Laboratory of Electronics, MIT, Cambridge, Mass.

ABSTRACT

The prosodic phenomenon of stress shift, in which a stronger prominence is perceived on the first syllable of a word like "Mississippi" than on its main stress syllable "-sip-" in stress-clash contexts like "Mississippi mud", has been attributed to rhythmic stress clash; the close approximation of two rhythmically prominent syllables is relieved by the left-ward shift of the first prominence to an earlier syllable. Acoustic measures suggest that intonational prominence can play a substantial role in this phenomenon.

1. INTRODUCTION

The prominence pattern known as 'stress shift' has received considerable attention in the last decade, as one of the cornerstones of metrical phonology. Speakers of a number of languages judge that, under certain circumstances, the strongest prominence in polysyllabic late-stress words like '"Mississippi" occurs not on the main stress syllable "-sip-", but on the earlier syllable "Miss-". The prosodic environments that induce this apparent stress shift, as in "Mississippi mud", have been described as 'stress clash'. That is, the close approximation of two strong prominences, on "-sip-" and "mud", is rhythmically unacceptable, and is avoided by shifting the first prominence leftward to an earlier strong syllable [2, 8, 9, 11, 13, 17, 18, 20, 21].

Some speakers demonstrate elegantly systematic intuitions about the environments in which this apparent shift will occur (e.g. in American English, the main-stress-initial word "legislator" will induce the shift in "Mississippi legislator", but the secondary-stress-initial word in "Mississippi legislation" will not.) For other speakers, the facts are less clear, and even the existence of the phenomenon may be in question. Experimental measures have not yet produced convincing evidence to support the claims of shift-producing speakers whose intuitions are, none the less, remarkably consistent [4, 7, 12].

An additional complication arises from the existence of intonation models in which a pitch marker occurs early in the utterance of a declarative sentence [3, 5, 6, 10, 14, 16, 19, 22]. To what extent might this marker, when it occurs on e.g. the first syllable of "Mississippi mud", contribute to the impression that a leftward shift of stress has occurred?

In this preliminary study, one of a series of ongoing experiments designed to disentangle these issues, we explore two acoustic measures which might be expected to reflect the perceived shift in prominence: duration and F0. We confine the investigation to spoken prose (in Abercrombie's sense, distinct from conversational speech [1]) in American English, and we omit for now any discussion of the potentially important characteristics of intensity and loudness. For a limited number of sentences, we address the following question: In utterances for which both metrical theory and perceptual evaluation indicate an apparent stress shift, is there any evidence that either the F0 or the duration of the shift-receiving syllable reflects the change?

2. METHOD

Speech materials consisted of single words spoken in the frame sentence "Say the X again" and their candidate stress-shift counterparts "Say the XY again". The three words investigated, Mississippi, Massachusetts and Maxine, begin with voiced nasal-vowel syllables which permit both f0 tracking from the preceding word, and reasonably reliable measurements of segment duration. The corresponding stress shift candidate phrases were Mississippi legislature, Massachusetts Avenue and Maxine Jones. A seventh phrase was included which was not predicted to undergo stress shift because of the lack of stress clash: Mississippi legislation.

The seven stimulus sentences were produced as part of a larger set of utterances by nine speakers, five male and four female. The utterances were recorded on cassette tape, in a partially sound-attenuated room, and digitized at 10,000 kHz. Duration measures were taken by hand from cursor readouts on waveform displays, and F0 estimates were obtained automatically by a procedure developed by Dennis Klatt that that involves finding the spacing between the harmonics in the spectrum.

Perceptual evaluation of stress shift in the resulting 63 utterances was carried out by the author. In many cases the outcome was clear: either the largest prominence was on the first syllable of the target word, (i.e. stress shift had occurred), or it remained on the syllable which would normally carry main lexical stress (i.e. no stress shift had occurred.) Interestingly, a third pattern emerged, in which the initial syllable and the main-stress syllable of the target word seemed to be of equal prominence. These cases were labelled 'unclear', and were analysed separately.

3. RESULTS & CONCLUSION

Perceptual analysis: Of the 27 target words predicted to undergo shift, 14 were judged to be shifted, while 2 had their major prominence on the mainstress syllable and thus had not shifted; both of the latter were utterances of "Say Maxine Jones again". In the remaining 11 cases the relative prominence of the first and mainstress syllables of the target word was judged unclear.

Of the 9 utterances of "Mississippi legislation", predicted not to undergo shift, 6 were shifted and 3 were unclear. Finally, of the 27 utterances of the single target words Mississippi, Massachusetts and Maxine in the frame sentence, 25 were unshifted and two were unclear. Thus, single words did not undergo shift, just over half of the candidate shift words did, and the phrase "Mississippi legislation", predicted not to shift, was perceived to shift more than half the time.

Individual speakers were somewhat consistent: five speakers shifted 4 or 3 utterances, and four shifted 1 or none. Individual sentences were also somewhat consistent, shifting for 5, 6, 5 and 4 of nine speakers. This pattern of results suggests the wisdom of perceptually evaluating candidate shift utterances to determine whether or not stress shift has occurred, before analysing its acoustic correlates.

Duration analysis: For each speaker, the duration of the first syllable of a target word produced alone in the frame sentence was compared with its duration in the shift candidate context, and the results tabulated separately for shifted, unclear and unshifted utterances. No striking differences among the 3 distributions were noted (Fig. 1a), perhaps because of variation in speaking rate from utterance to utterance. If stress shift is accompanied by systematic timing differences in the shifted-to syllable, the differences (as other investigators have reported) are not easy to demonstrate with this simple comparison between utterances.

F0 analysis: The F0 results present a somewhat clearer picture. We report here only the within-utterance measure of F0 change in the first syllable of the target words. This was defined as the size and direction of the change between the highest and lowest F0 values in the syllable. In words judged to show stress shift, the change was generally large and positive, ranging up to 71 Hz, while the unclear cases were more often small or negative. Finally, the 2 cases judged to be unshifted, with their major prominence remaining on the mainstress syllable, showed large negative changes in F0 in the first syllable: -36 and -15 Hz. The distribution of F0 changes in the initial syllable of the target words is summarized in Figure 1b.

These results suggest that utterances in which stress shift is perceived tend to have large F0 rises in the shifted-to syllable, although such a rise is apparently not sufficient to ensure the perception of stress in all cases, since a subset of

those labelled 'unclear' were also associated with large rises (49, 34 and 16 Hz). All 3 of these cases were produced by the same speaker, and were instances where both the first and the mainstress syllables were strongly and equally prominent, suggesting that speakers can place pitch markers on more than one of the strong syllables of the target word under some circumstances.

The fact that stress shift was also perceived for a few utterances with no clear F0 change in the first syllable of the target word suggests that other acoustic cues may be used. Three of the five examples of this kind were produced by the same speaker, and there was no evidence that this speaker relied on duration increases: the initial syllable of the target word was 30-50 mS shorter in the stress-shifted utterances than in the corresponding single-word utterances. Other possibilities include a change in F0 from the last syllable of the preceding word, or the relative F0 change (or relative duration) of syllable 1 and the mainstress syllable. The single-word cases for this speaker show a substantial fall in the first syllable of the target word (30-50 Hz), so that the stress shift cases always have a lesser fall in F0 in the shifted-to syllable than the single word cases, but it is unclear whether this fact is related to the perception of stress shift.

An interesting aspect of the initial-syllable F0 patterns is the pitch marker observed in utterances of single words in frame sentences. An example is shown to the left in Figure 2, where the initial syllable "Mi-" shows an F0 rise for both "Mississippi" and "Mississippi legislature". Since no stress shift was perceived in the single-word case for this speaker, the initial-syllable marker is apparently overshadowed by a more prominent marker on the mainstress syllable "-sip-". This inference is supported by the F0 pattern for the mainstress syllable in the same word, shown to the right in the figure. A possible interpretation of this pattern is that speakers have two separate options for the placement of pitch markers on a polysyllabic target word: they can mark the initial syllable or not, and they can mark the mainstress syllable or not. On this view, the combination of pitch marking on the first syllable and no pitch marking on the mainstress syllable could contribute substantially to the perception of stress shift. For a synthesis algorithm compatible with this hypothesis see Monaghan and Ladd [15].

Conclusion: The preliminary results reported here illustrate several significant points: (1) it is important to evaluate stress shift candidate utterances perceptually before measuring possible correlates of stress shift, since not all clash contexts invariably induce shift and it occurs in some non-clash contexts, (2) in some shift cases, the greater perceptual prominence of the shifted-to syllable may be a matter of intonational rather than rhythmic prominence, (3) the hypothesis that this prominence early in the word reflects in part an 'unmasking' of the prominence associated with an onset intonational marker on an earlier syllable of the word, an unmasking which results from the disappearance of phrasal prominence from the mainstress syllable (in favor of a later word), requires further testing, and (4) speakers can take different approaches to the problem addressed by stress shift models; determining the options available to speakers will be an important step toward understanding the relation between not only rhythmic and intonational aspects of prosody, but also lexical and phrasal prominence.

Future work: Clearly, an understanding of stress shift will require a comprehensive approach involving phonological, acoustic-phonetic and perceptual analyses, with more speakers, more utterances, and more listeners doing the perceptual evaluations [4]. In addition, an important control experiment remains to be run. If the longer string of syllables in the stress shift candidate sentences causes the speaker to reach a higher early F0, the results reported above would have a very different interpretation. A control experiment comparing F0 and duration changes for initial syllables in non-shiftable pairs like "manageable" vs. "manageable legislators" will test this possibility.

4. ACKNOWLEDGEMENTS

This work was supported by grants NSF IRI-8805680 and NIH 8-301-DC00075. Conversations with P.J. Price and M. Ostendorf have been invaluable.

REFERENCES

[1] Abercrombie, D. (1965), Studies in Phonetics and Linguistics, London: Oxford U. Press
[2] Beckman, M. E. (1986), Stress and non-stress accent, Dordrecht:Foris
[3] Beckman, M.E. and Pierrehumbert, J. (1986), Intonational structure in Japanese and English, *Phonology Yearbook 3*, 255-309
[4] Beckman, M.E., Swora, M.G., Rauschenberg, J. anad deJong, K. , Stress shift, stress clash and polysyllabic shortening in a prosodically annotated discourse, presented at the Conference on Speech Processing, Kobe Japan, November 1990
[5] Bolinger, D. (1981), Two kinds of vowels, two kinds of rhythm, distributed by the Indiana University Linguistics Club, Bloomington
[6] Bollinger, D. (1986), Intonation and its parts, Stanford: Stanford U. Press
[7] Cooper, W. and Eady, S.J. (1986), Metrical phonology in speech production, *J. Memory and Language 25*, 369-384
[8] Gussenhoven, C. (1983), Stress shift and the nucleus, in On the grammar and semantics of sentence accents, Dordrecht: Foris (also in *Linguistics 21*)
[9] Halle, M. and Vergnaud, J.-R., An essay on stress, Cambridge, Mass: MIT Press
[10] Hart, J. 't and Collier, R. (1975), Integrating different levels of intonation analysis, *J. Phonetics 3*, 235-255
[11] Hayes, B. (1984), The phonology of rhythm in English, *Ling. Inq. 15*, 33-74
[12] Horne, M. (1988), Empirical evidence for a nonmovement analysis of the rhythm rule in English, *Lund Univ. Dept. Ling. Working Papers 33*, 139-152.
[13] Liberman, M. and Prince, A. (1977), On stress and linguistic rhythm, *Ling. Inq. 8*, 249-336
[14] Maeda, S. (1974), A characterization of fundamental frequency contours of speech, *Quart. Prog. Rep. MIT RLE 114*, 193-211
[15] Monaghan, A. and Ladd, D.R. (1987), An outline of the intonational component of the Edinburgh text-to-speech system, *Edinburgh Univ. Dept. of Ling. Work in Prog. 20*, 90-101
[16] dePijper, J.R. (1983), Modelling British intonation, Dordrecht: Foris
[17] Prince, A. (1983), Relating to the grid, *Ling. Inq. 14*, 19-100
[18] Nespor, M. and Vogel, I. (1986), Prosodic phonology, Dordrecht: Foris
[19] Pierrehumbert, J. (1980), The phonology and phonetics of English intonation, unpublished PhD dissertation, Massachusetts Institute of Technology
[20] Selkirk, E. (1984), Phonology and syntax, Cambridge, Mass: MIT Press
[21] Vanderslice, R. and Ladefoged, P. (1971), Binary suprasegmental features, *UCLA Working Papers in Phonetics 17*, 6-24
[22] Willems, N. (1982), English intonation from a Dutch point of view, Dordrecht: Foris

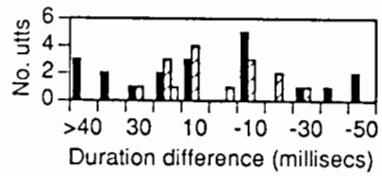
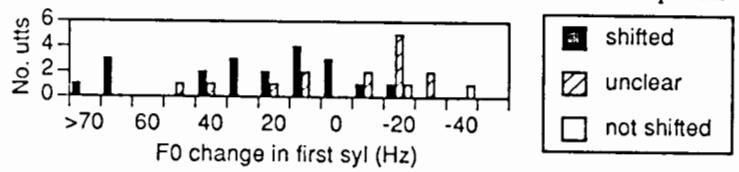


Fig. 1a (top): Difference in duration of the initial syllable for target words produced in a single-word phrase and in a corresponding stress shift candidate phrase by the same speaker
Fig 1b (bottom): F0 excursion in the initial syllable of target words produced in stress shift candidate phrases



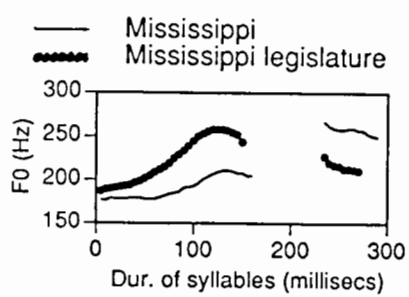


Fig. 2: Left traces are F0 values for the initial syllable Mi- in "Say the Mississippi again" (single word) and "Say the Mississippi legislature again" (stress shift candidate phrase) produced by a single speaker. Right traces are F0 values for mainstress syllable -sip- in the same utterances. Time axis reflects elapsed time for each syllable but not between syllables; onsets of voiced portions of syllables are aligned.