Lund University, Dept. of Linguistics
Working Papers 33 (1988), 139-152

Empirical Evidence for a Nonmovement Analysis of the Rhythm Rule in English

Merle Horne

Abstract
Empirical data from American English are presented which shed light on the phonetic reality behind the phenomenon variously termed the 'Rhythm Rule', 'Iambic Reversal', 'Beat Movement'. The data argue against the prevalent movement analysis of prominence from the stressed syllable to the preceding syllable in Rhythm Rule words like *Dundee* when occurring in a phrase such as *Dundee tartan*. Rather, it is seen that the reason more prominence is perceived on the first syllable is related to the fact that the strong pitch movement that occurs on the stressed syllable when the word occurs focussed at the head of its phrase is absent when the word occurs prefocally or unfocussed. In this situation, the underlying prominence on the initial syllable becomes salient and outweighs that on the stressed syllable. This salience is realized phonetically by greater relative duration combined with a fundamental frequency obtrusion that is equal to or greater than that on the stressed syllable.

Among the many research topics that have become the focus of Bengt Sigurd's attention, that of text generation and text-to-speech have constituted an area of particular interest (Sigurd 1981, 1982, 1983, 1984). The goal of the present paper is to present some empirical data that have implications for how the prosodic component of a text-to-speech system for English should be structured.

INTRODUCTION

One of the most widely discussed phenomena in metrical phonological theory has been the Rhythm Rule. In e.g. Liberman and Prince 1977 or Selkirk 1984, the Rhythm Rule is analyzed as involving a movement of 'stress'. For example, when the well-known phrase *thirteen men* is uttered out of context, one gets the impression that in the word *thirteen*, stress shifts from the lexically stressed syllable, *-teen* to the syllable *thir-* when followed by another word with lexical stress on the first syllable, e.g. *men*. That is to say, the potential 'clash' that is assumed to arise when two stressed syllables lie next to each other is avoided by moving stress to the left. Various formulations of the rule have appeared, but until recently, no empirical data have been presented that show what the phonetic reality is behind the impressionistic change in prominence relations reflected in the Rhythm Rule.

COOPER & EADY'S 1986 TEST OF THE RHYTHM RULE

In 1986, Cooper and Eady, however, attempted to test the Rhythm Rule as formulated by Selkirk in her 'grid-only' 1984 analysis. They compared pairs of phrases containing the word *thirteen*, such as *thirteen colleges/thirteen universities*. According to Selkirk's analysis, one would expect the Rhythm Rule to apply in the case of *thirteen colleges* where *thirteen* is followed by a word beginning with a stressed syllable, but not in the case of *thirteen universities* where no such stress-clash arises (see Figure 1). The third-level word-stress prominences are here separated by a place-holder one level down.

Cooper and Eady measured the difference in Fo (fundamental frequency) peaks on the two syllables of the word *thirteen* in both contexts as well as the difference in absolute duration of the syllable *thir-* in both contexts. They expected that the syllable *thir-* would have increased absolute values of duration and a higher Fo peak in the case where it should have undergone the Rhythm Rule (or 'Beat Movement') than in the second case where no movement of prominence is expected. Moreover, they also expected that the Fo peak on *-teen* should be lower in the second case. The predicted patterns were not, however, observed in the production data collected by Cooper and Eady. Instead, they found that for the first syllable of *thirteen*, the mean duration is actually **less** in the case where *teen* was followed by a stressed syllable than for the case where *-teen* was followed by a word with non-initial stress. Moreover, there was only a nonsignificant pitch difference of 1 Hz on both syllables of *thirteen*. These results led Cooper and Eady to question the empirical adequacy of the metrical analysis.

In all fairness to metrical phonology, however, it should be pointed out that according to Liberman and Prince's ('tree-grid') 1977 formulation of the Rhythm Rule, one would not have expected any difference between the two cases of *thirteen* that Cooper and Eady discuss. Both should have been subject to the Rhythm Rule or Iambic Reversal in their respective metrical trees due to the clashes in the associated grids (see Figure 2). (Note that, unlike Selkirk, Liberman and Prince allow stress clashes at any level of the grid to trigger the Rhythm Rule.) What Cooper and Eady have shown is that *thirteen* has the same rhythmic structure in both contexts, and according to one version of the Rhythm Rule, both instances of *thirteen* **should** have undergone the Rhythm Rule.

MOTIVATION FOR RETESTING OF RHYTHM RULE

In order to test a movement analysis of the Rhythm Rule, what we decided to do was to compare a word like *thirteen* with a phrase like *thirteen colleges*. All versions of metrical theory would agree that there should, in some sense, be more prominence on *thir-* in the second instance than in the first. Moreover,

Figure 1. Metrical grids for the phrases *thirteen colleges* and *thirteen universities* according to Selkirk's 1984 analysis.

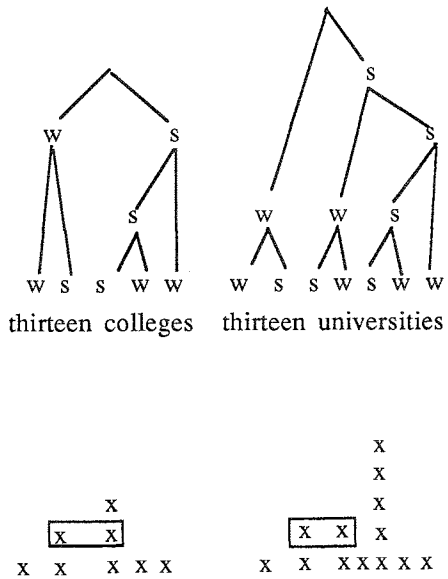

Figure 2. Metrical trees and corresponding metrical grids for the phrases *thirteen colleges* and *thirteen universities* according to Liberman and Prince 1977.

since the Rhythm Rule is meant to capture the notion of relative prominence, it seemed to us that it would be more appropriate to measure **relative** values of phonetic parameters when testing the Rhythm Rule. That is to say, rather than just regard the Rhythm Rule as involving a movement of **absolute** strength from one syllable to another as Cooper and Eady did, one should instead interpret the rule as expressing an adjustment in relative strength. So that, instead of looking at the duration of just the syllable *thir-* as Cooper and Eady did, one should

compare the relative durations of **both** *thir-* and *-teen* in phrases like *thirteen colleges* and in the one-word phrase *thirteen.*

Bolinger 1981, following Bruce's 1981 work on rhythm in Swedish, suggests that one must separate the syllabic rhythmic patterning from the accentual (or tonal) rhythmic patterning. Following this idea, one could hypothesize, for instance, that the reason why *thir-* in *thirteen colleges* is more prominent than *-teen* is because, in the absence of the strong focal prominence on *-teen* which is present when *thirteen* is focussed or phrase-final, the underlying syllabic prominence on the first syllable (expressed in terms of duration, Fo and intensity) becomes salient and overrides that of the second syllable.

DATA

In order to test this non-movement interpretation, we decided to compare the relative values of these phonetic parameters in the first and second syllables of a number of Rhythm Rule words (*Dundee, canteen, maintain*) in focal and prefocal position (or phrase-final and prefinal position). The isolated words and the two-word, right-branching phrases were placed in the carrier sentence: *He said* ____, *I think* which was uttered in response to the question: *What did he say?* The following words and two-word phrases were used:

Dundee	Focal
Dundee tartan	Prefocal
Canteen	Focal
Canteen cook	Prefocal
Maintain	Focal
Maintain roads	Prefocal

PROCEDURES

Three tokens of each utterance were recorded in the sound studio at the Dept. of Linguistics, Lund University. The informant was an American male from Minnesota in his early 30's who has not had any training in linguistics. Acoustic analysis of the test sentences was performed using the ILS program package implemented on a VAX 11/730 computer. The speech was first digitized at a sampling rate of 10 kHz. Duration measurements were made with a computer-controlled cursor using computer oscillograms and an interactive wave form editor. For syllables beginning with a voiceless plosive, the syllable onset was measured from the beginning of the plosive burst. The period of silence

containing the plosive closure was thus included in the duration measurement of the preceding syllable. Fo and intensity measurements were made from curves plotted using pitch editing and intensity plotting commands, respectively.

RESULTS

Dundee

If we examine the results for *Dundee* (Figure 3), it can be seen that as far as duration goes, the first syllable's mean relative duration is 60% of the total duration of the word in prefocal position as opposed to 49% in focal position. However, the absolute duration of *dun-* in prefocal position is not greater than it was in focal position. One would not expect it to be either, since it occurs in a longer utterance and absolute syllable duration is dependent on utterance length (see Strangert 1985).

As far as the parameter of fundamental frequency is concerned, we measured both the relative difference in the size of the Fo obtrusion on both syllables as well as the difference in the height of the tops of the Fo obtrusions. If one compares the range of the Fo obtrusion on both syllables of the test word (see Figure 4), it is seen that in the focal cases, the mean range of the Fo obtrusion on the first syllable is 20 Hz less than that on the second. In prefocal position, however, the mean Fo obtrusion on the first syllable is 8 Hz greater than that on the second. Moreover, whereas the Fo peak on the first syllable was 1 Hz lower than that on the second in the focal case, in prefocal position, the mean height of the Fo peak on the first syllable was 12 Hz higher than that on the second (Figure 5). In short, in the phrase *Dundee tartan*, the first syllable of *Dundee* shows both a relatively greater pitch obtrusion than that on the second syllable as well as a higher Fo top than that on the second. It has also greater duration than the second.

As far as intensity is concerned, the mean relative intensity (measured in terms of peak height) on the first syllable was in both cases greater than that on the second syllable (6 dB greater in focal position and 8 dB greater in prefocal position; see Figure 6). The role played by intensity is thus less clear than that of duration and Fo relations between the two syllables in creating the impression of prominence.

Canteen

If one compares the results for *Dundee* with those for the word *canteen* in the same environments, a somewhat different picture arises. In focal position, the second syllable was always perceived (in informal listening tests) as more prominent as in the case of *Dundee*, but, of the three tokens of the word uttered in prefocal position, only the first token was perceived as having more

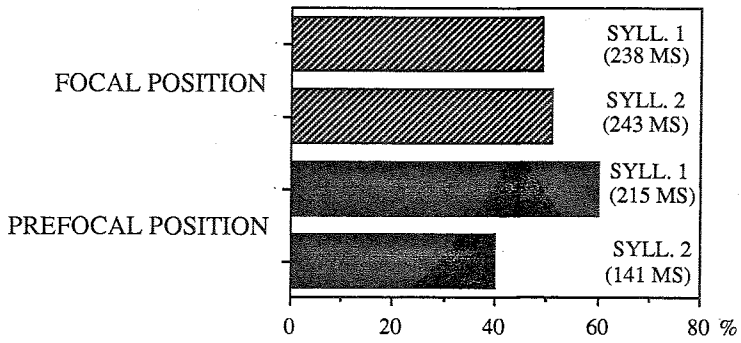

Figure 3. Mean relative duration of syllables with respect to that of whole word (absolute values in parentheses).

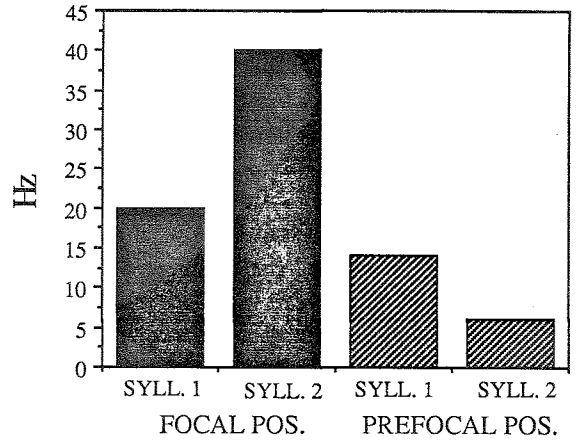

Figure 4. Mean range of Fo obtrusion in focal and prefocal position.

prominence on the first syllable. The second and third tokens seemed to have about an equal amount of prominence. That is to say, only the first token seemed to have undergone the Rhythm Rule. Consequently, we have presented separately the results for the first token of *canteen* in prefocal position.

In focal position, the first syllable has a shorter duration than the second (Figure 7), as well as a smaller Fo obtrusion (Figure 8) and a lower Fo peak (Figure 9).

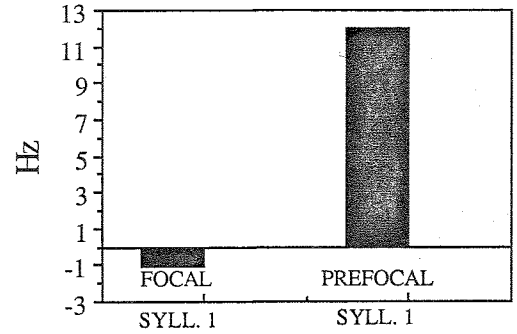

Figure 5. Difference between 1st syllable and 2nd syllable Fo tops (horizontal line passing through 0 represents height of top on 2nd syllable).

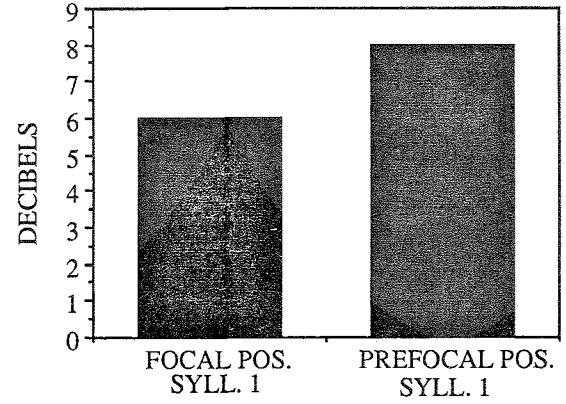

Figure 6. Difference between 1st syllable and 2nd syllable peak intensity (horizontal line passing through 0 represents intensity peak on 2nd syllable).

For the first token of *canteen* in prefocal position, where the first syllable was perceived as more prominent than the second, it can be seen (Figure 7) that the first syllable has both a relatively greater duration than the second (6%), and has an Fo obtrusion which is 10 Hz larger than the second (Figure 8). Moreover, it has an Fo peak which is 10 Hz higher than that on the second syllable (Figure 9).

For the second and third tokens in prefocal position, where both syllables were perceived as having about the same amount of prominence, one can see

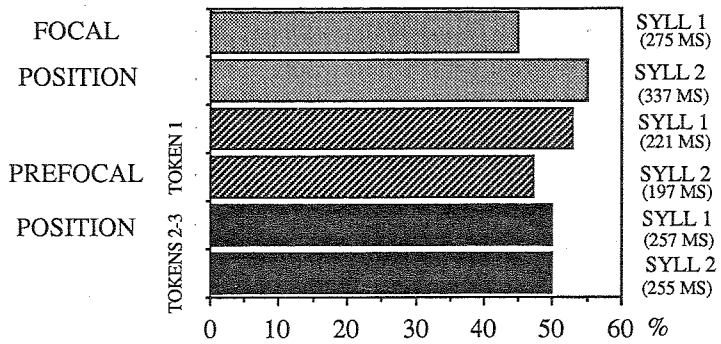

Figure 7. Relative duration of syllables with respect to that of whole word.

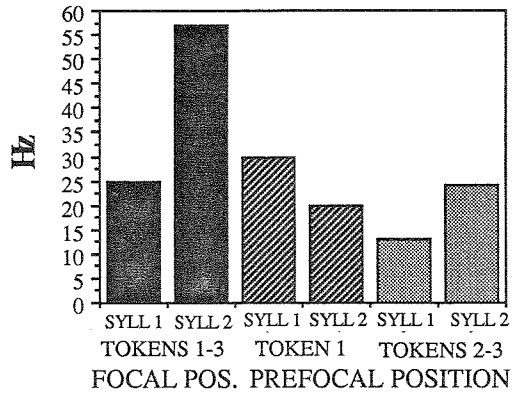

Figure 8. Difference between 1st syllable and 2nd syllable Fo range.

(Figure 7) that both syllables have the same relative duration. Moreover, the Fo obtrusion on the first syllable is 11 Hz less than that on the second syllable (Figure 8). The Fo peak on the first syllable is somewhat higher than that on the second, but only by 3 Hz (Figure 9).

As in the case of *Dundee,* the peak intensity on the first syllable is in all cases greater than that on the second (see Figure 10).

In order to explain what it was that made the first syllable of *canteen* more prominent in the first token of the prefocal occurrences of the word as opposed to the second and third tokens, one can point to the fact that, as in the case of

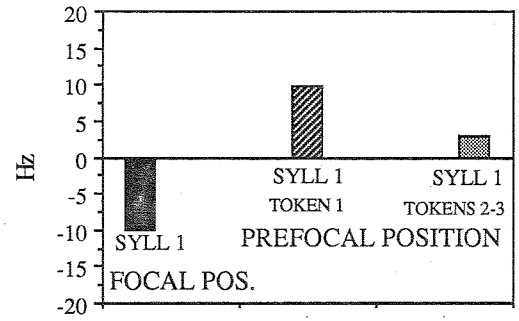

Figure 9. Difference between 1st syllable and 2nd syllable Fo tops (horizontal line passing through 0 represents height of top on 2nd syllable).

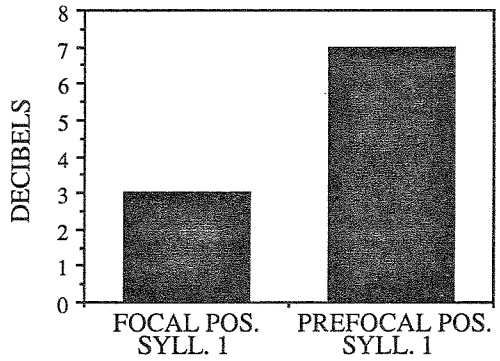

Figure 10. Difference between 1st syllable and 2nd syllable peak intensity (horizontal line passing through 0 represents intensity peak on 2nd syllable).

Dundee, the first syllable had both a greater duration than the second as well as a larger Fo obtrusion.

All three prefocal tokens had a higher Fo and intensity peak on the first syllable. However, these two factors are not sufficient by themselves to give the impression of greater prominence. It is crucial that the Fo obtrusion on the first syllable not be less than that on the second and the duration of the first syllable must be greater than the second.

Maintain

In order to test whether the Rhythm Rule could apply across phonological phrase boundaries as defined by Selkirk 1980 or Nespor and Vogel 1982, we examined the relative prominence of the syllables in the word *maintain* in focal position as well as in the phrase *maintain roads* which consists of two phonological phrases. As in the other words examined so far, in focal position, the first syllable has shorter duration than the second as well as a smaller Fo obtrusion and a lower Fo peak. The intensity peak was also lower than that on the second.

In prefocal position, however, the first token of *maintain* exhibited a different rhythmic structure than it did in the second and third tokens. Both syllables of the first token in prefocal position were judged to have about the same level of prominence in informal listening tests, whereas in the second and third tokens, the first syllable was definitely heard as being more prominent than the second. In other words, the Rhythm Rule did not apply in the first token, but did in the second and third instances. If one compares the values of the phonetic parameters in prefocal position, it is seen that in all three tokens, the first syllable had a greater duration than the second (see Figure 11). However, in the first token, where the first syllable was not perceived as having greater prominence than the second, it can be seen that the Fo obtrusion on this syllable was 15 Hz smaller than that on the second (Figure 12).

In the second and third tokens, where the first syllable was perceived as having more prominence than the second, the Fo obtrusions on both syllables were the same (Figure 12).

In all prefocal tokens, the height of the Fo top on the first syllable was lower than that on the second but only by at most 2 Hz (Figure 13). Thus, it would appear to be the greater duration of the first syllable coupled with an Fo obtrusion that is not less than that on the second syllable which creates the impression of prominence on the first syllable here as in the other cases.

One could perhaps interpret these results as showing that the speaker interpreted the second and third tokens of *maintain* in the phrase *maintain roads* as contextually given and consequently did not assign them any level of Fo prominence associated with focus or highlighting.

The data show that the rhythmic adjustment associated with the Rhythm Rule is not just restricted to the phonological phrase, but is rather conditioned by prominence relations between adjacent words. The affected word must be unfocussed (i.e. does not have an Fo movement associated with new information or semantic highlighting) and, at least in the data we have looked at, the following word begins with a strong syllable. According to Vogel (personal communication), the predicate-object phrase *maintain roads* could instead be

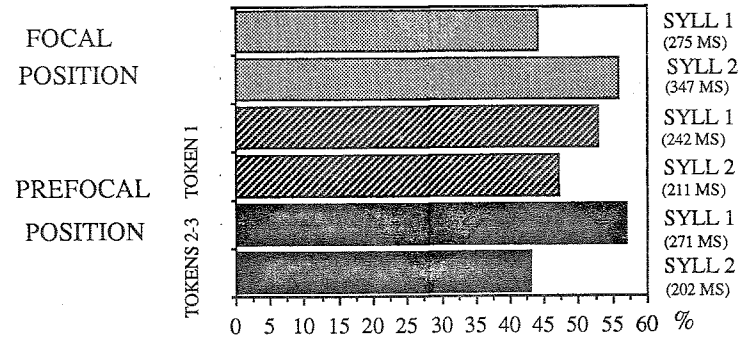

Figure 11. Relative duration of syllables with respect to that of whole word.

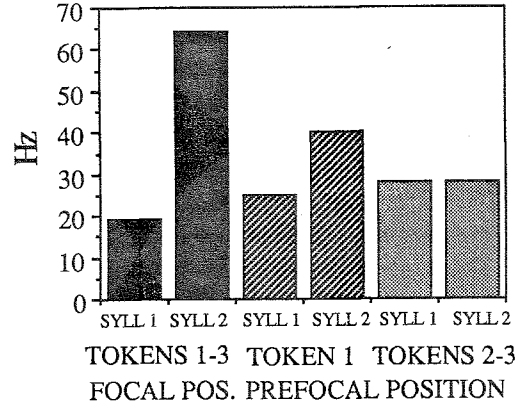

Figure 12. Difference between 1st syllable and 2nd syllable Fo range.

regarded as constituting a 'restructured' phonological phrase, the condition for restructuring then being that the predicate is given information.

CONCLUSION

The results presented here do not support an interpretation of the Rhythm Rule as involving a movement of absolute strength from the stressed syllable to a preceding one, as a pitch-first theory such as Selkirk's would imply. What we found when comparing focal and prefocal forms of words which are subject to change in rhythmic structure was that they in all cases exhibit both a relative and

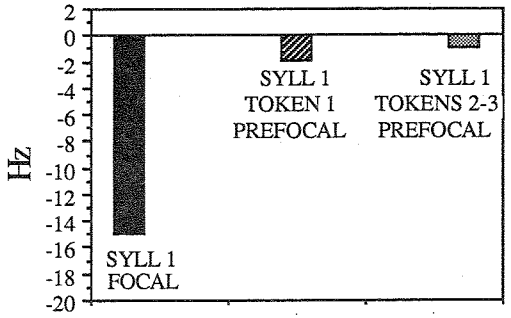

Figure 13. Difference between 1st syllable and 2nd syllable Fo tops.

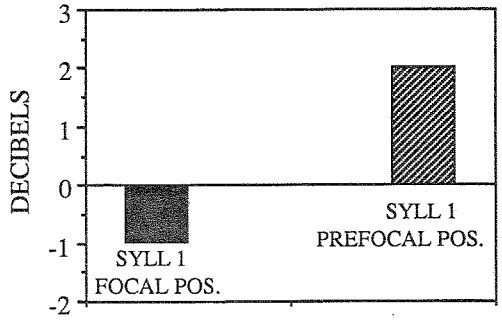

Figure 14. Difference between 1st syllable and 2nd syllable peak intensity.

an absolute **decrease** in prominence on the stressed syllable between focal and prefocal forms. However, the first syllable did not show any significant increase in absolute prominence between the focal and prefocal forms. It did, however, show an increase in **relative** prominence with respect to the stressed syllable in prefocal position. The results support a view of rhythm which separates syllabic from accentual rhythm such as Bolinger and Bruce propose. Alternations in relative duration would appear to be the most consistent phonetic pattern associated with the realization of syllabic rhythmic prominence. This is in agreement with Bruce's 1981 work on Swedish rhythm. In order for these alternations in duration to become salient, however, it is crucial that the stressed syllable of the word undergoing the Rhythm Rule be devoid of any strong pitch movement that could be associated with focus (this view is adopted in Horne 1986). This observation is, furthermore, in line with a more recent 'pitch accent' deletion

analysis of the Rhythm Rule presented in Gussenhoven 1986 and developed in 1988 (where pitch accent refers to "an abstract place marker for the association of (intonational) tonal morphemes").

An interesting question that remains to be answered is whether, for example, the greater duration on the first syllable of the test words is due to a rhythmic process or whether it is an automatic result of the underlying durations and Fo values of the segments making up the syllables. For example, the first vowels in *Dundee* and *canteen* are phonologically stronger than the second and have greater intrinsic duration than the second. This could perhaps be tested, however, if the same words were put in a prefocal context where the following word began with a weak syllable. Then one could hypothesize that if it was a rhythmic process that accounted for the alternation in syllable duration, then if the word *Dundee* were followed by a word beginning with a weak syllable, e.g. *delight*, then the second syllable of *Dundee* would have relatively greater duration than the first. The fact that there is no difference in the size of the Fo obtrusions on the two identical vowels of *maintain* in prefocal position, however, would lead one to suspect that it is perhaps underlying (intrinsic) values of phonetic parameters that surface in this environment in the absence of focal prominence. If such is the case, then one would not predict that *Dundee* would behave differently when followed by a weak word-initial syllable.

Our data also support the interpretation of the Rhythm Rule that was presented in Horne 1986, where it was argued that the domain of the rule was not the 'phonological phrase' as defined by Selkirk 1980 and Nespor and Vogel 1982, but rather that it was determined by prominence levels holding within a given sentence. According to the analysis that limits the application of the Rhythm Rule to the phonological phrase, the process should not apply to e.g. Verb-Object concatenations like *maintain roads*. However, we have seen that these constructions do exhibit the same rhythmic alternation as do simple NPs provided the first word does not have any Fo prominence associated with focus.

The finding that there is no movement of absolute prominence in the Rhythm Rule data has implications for the development of the prosodic component of text-to-speech systems. Assuming that lexical forms are stored in their focussed (citation) form, no transferring of absolute values of Fo, duration and intensity need be done when the word occurs in prefocal position. What is required is a deletion of absolute phonetic strength from the stressed syllable associated with focus and an adjustment of the relative syllable duration depending on utterance length.

REFERENCES

Bolinger, Dwight. 1981. *Two kinds of vowels, two kinds of rhythm.* Bloomington: IULC.

Bruce, Gösta. 1981. 'Tonal and temporal interplay'. *Working Papers* 21, 49-60. Lund: Dept. of Linguistics.

Cooper, William and Stephen Eady. 1986. 'Metrical phonology in speech production'. *Journal of Memory and Language* 25, 369-84.

Gussenhoven, Carlos. 1986. Review of Selkirk 1984. *Journal of Linguistics* 22, 455-74.

Gussenhoven, Carlos. 1988. 'Lexical accent rules in English'. Unpublished manuscript, Instituut Engels-Amerikaans, Nijmegen University.

Horne, Merle. 1986. 'Focal prominence and the 'phonological phrase' within some recent theories'. *Studia Linguistica* 40, 101-21. (Also published in *Towards a discourse-based model of English sentence intonation, Working Papers* 32, 1987. Lund: Dept. of Linguistics.)

Liberman, Mark and Alan Prince. 1977. 'On stress and linguistic rhythm'. *Linguistic Inquiry* 8, 249-336.

Nespor, Marina and Irene Vogel. 1982. 'Prosodic domains of external sandhi rules'. *The structure of phonological representations,* ed. Harry van der Hulst and Norval Smith, 225-55. Dordrecht: Foris.

Selkirk, Elisabeth. 1980. *On prosodic structure and its relation to syntactic structure.* Bloomington: Indiana University Linguistics Club.

Selkirk, Elisabeth. 1984. *Phonology and syntax: the relation between sound and structure.* Cambridge, Mass.: MIT Press.

Sigurd, Bengt. 1981. 'Commentator. A computer system simulating verbal behaviour'. *Working Papers* 20, 67-89. Lund: Dept. of Linguistics.

Sigurd, Bengt. 1982. 'Text representation in a text production model'. *Text processing. Proceedings of the Nobel Symposium*, ed. Sture Allén, 135-52. Stockholm: Almqvist & Wiksell.

Sigurd, Bengt. 1983. 'How to make a text production system work'. *Working Papers* 25, 179-94. Lund: Dept. of Linguistics.

Sigurd, Bengt. 1984. 'Computer simulation of spontaneous speech production'. *Proceedings of Coling 84*, 79-83. Association for Computational Linguistics.

Strangert, Eva. 1985. *Swedish speech rhythm in a cross-language perspective.* Stockholm: Almqvist & Wiksell.

Lund University, Dept. of Linguistics
Working Papers 33 (1988), 153-161

Recognition of Prosodic Categories in Swedish: Rule Implementation

David House, Gösta Bruce, Lars Eriksson and Francisco Lacerda*

Abstract
Descriptive rules for recognition of prosodic categories in Swedish are currently being implemented in an automatic prosody recognition scheme. An algorithm is described in which the speech signal is segmented into syllables (tonal segments) using intensity measurements and fundamental frequency. Each syllable is then given six values related to fundamental frequency and duration. The values for each syllable are tested against conditions which describe the prosodic categories. The category attaining the highest score is assigned to the syllable. Preliminary results for two sets of rule conditions for ten test sentences are presented.

INTRODUCTION

This paper represents a status report from an ongoing joint research project shared by the Phonetics Departments at the Universities of Lund and Stockholm. The project, "Prosodic Parsing for Swedish Speech Recognition", is sponsored by the National Swedish Board for Technical Development and is part of the National Swedish Speech Recognition Effort in Speech Technology. The primary goal of the project is to develop a method for extracting relevant prosodic information from a speech signal. We hope to devise a system which from a speech signal input will provide us with a transcription showing syllabification of the utterance, categorization of the syllables into STRESSED and UNSTRESSED, categorization of the stressed syllables into WORD ACCENTS (ACUTE and GRAVE) and categorization of the word accents into FOCAL and NON-FOCAL accents. We also hope to be able to identify JUNCTURE (connective and boundary signals for phrases). We are currently working with 20 prosodically varied sentences spoken by two speakers of Stockholm Swedish.

The type and structure of the information to be presented to the recognizer has been based on a series of mingogram reading experiments (see House et al. 1987a, 1987b). In the first experiment an expert in Swedish prosody (Gösta Bruce) was presented with mingogram representations of ten unknown sentences showing a duplex oscillogram, fundamental frequency contour and intensity curve. On the basis of this information, he was able to identify 85% of all

* At Stockholm University, Department of Linguistics and Phonetics