TEMPO AND THE RHYTHM RULE

P.F. McCormack*, J. C. Ingram~

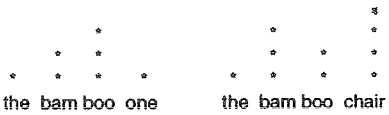
*Department of Speech Pathology
Flinders University
South Australia

~Department of English
University of Queensland

ABSTRACT - The adjustment of linguistic stress patterns under the influence of rhythm is well attested, and forms the basis of a number of models of speech organisation. An experiment is reported on the perceptual and acoustic consequences of manipulating speaking rate for such rhythmic stress shifts. The results are discussed in terms of their implications for the place of rhythm in models of speech production.

INTRODUCTION

Is the Rhythm Rule in speech production a reflection of phonological (hence pre-articulatory) planning or an automatic articulatory consequence of producing speech? It is generally acknowledged that shifts in the prominence patterns on some words in connected speech are due to a strong rhythmic constraint to prefer the alternation of stressed and unstressed elements and to avoid the juxtaposition of stresses. Hence, while *bamboo* spoken in a noun phrase such as *the bamboo one* has the main stress on the last syllable, in a noun phrase such as *the bamboo chair* there is a perception that the main stress has shifted to the first syllable in *bamboo*. Some current phonological and psycholinguistic models of speech organisation account for the Rhythm Rule by representing rhythmic principles as actively operating solely on the "phonological" pre-articulatory plan of an utterance (Prince, 1983; Selkirk, 1984; Levelt, 1989; Ferreira, 1993). Both Selkirk (1984) and Levelt (1989) have argued that a "metrical component" mediates between the syntactic-semantic representation of an utterance and its segmental phonological representation, transforming syntactic phrasing into phonological phrasing. Such operations are prior to any articulatory formulation. The Rhythm Rule proposed by Selkirk (1984) as a formal operation where stress shifts from one syllable of a word on to another in order to avoid "clashing" with an adjoining stress. Each beat of prominence in the representation is represented by an asterisk (*). These prominence beats form a layered hierarchy called a "metrical grid", with each layer representing another level of prominence within the utterance. On the first layer each syllable is represented by a beat, on the second layer, each stressed syllable, on the third each word's main stress. Beats can shift leftward within the metrical grid so as to avoid a "stress clash" of 2 beats adjoining each other. While recent grid models have been "enriched" with beats for syntactic and non-syntactic pauses (Ferreira, 1993), they are fundamentally "top down" models of speech organisation where adjustments to the phonological representation of an utterance make no direct reference to the current articulatory performance status of the speaker. A metrical grid representation of the 2 bamboo phrases could be as follows:



Yet there is an older tradition in speech science that views rhythm in speech as essentially phonetic in character and in motivation; possibly reflecting more general principles of perceptual and motor performance (Jones, 1918; Bolinger, 1962; Abercrombie, 1967; Allen, 1975). In speech production, rhythmic adjustments to the juxtaposition of stressed syllables occur at the articulatory level as an automatic consequence of the timing relations between stressed syllables. The Rhythm Rule is highly variable in its application, and it has been suggested for some time that its expression is closely tied to

a person's speaking rate (Giegerich, 1981; Hayes, 1984). At a faster speaking rate the tendency for shifts to occur is thought to increase. Conversely, at a slower speaking rate it is suggested that speakers will produce less stress shifting. If the Rhythm Rule is dependent on speech tempo, then the question arises as to whether a phonetic explanation of the phenomenon, as a "low level" rule of articulatory adjustment, may be more appropriate than an explanation that represents the rhythm rule more centrally in speech organisation as an abstract principle operating on pre-articulatory representations. Alternatively, if tempo is found to affect the application of the Rhythm Rule, then the more abstract phonological models of speech rhythm need to be more explicit in accounting for how such motorically motivated variability as is found in speaking rate can result in adjustments at a pre-articulatory level of representation.

AIM

While direct effects of variation in tempo on the Rhythm Rule have been suggested by a number of authors, it has never been empirically tested. The aim of this experiment is to systematically investigate the effects of altering speaking rate on the expression of the Rhythm Rule.

METHOD

Stimuli

Six speakers of Australian English were recorded reading a series of sentences containing noun phrases which comprised of a potential stress shift word followed by a word with varying syllable distance to its main stress. The sentences were designed to provide a phonological context where shift and non shift environments could be manipulated. Examples of the 5 contexts used are as follows:

Two contexts where no shift was predicted:

No stress following	They were japanese ones at the hotel.
Shift word focused	They were JAPANESE tourists at the hotel.

Three contexts where shift was predicted:

One syllable distance	They were japanese tourists at the hotel.
Two syllable distance	They were japanese developers at the hotel.
Three syllable distance	They were japanese politicians at the hotel.

Six potential stress shift words were used: *thirteen*, *bamboo*, *sardine*, *underdone*, *overnight*, and *japanese*. These words had been identified in a previous experiment as being particularly susceptible to stress shift in speech production. Each word consists of 2 metrical feet, 3 have feet containing one syllable each (eg | sar | dine |), while 3 have feet with the first foot containing 2 syllables and the second one syllable (eg | over | night |). Speakers produced each sentence at 3 different rates: normal, slow, and fast. The average speech rate for each tempo was 4.6, 2.4, and 6.9 syllables per second respectively.

Analysis

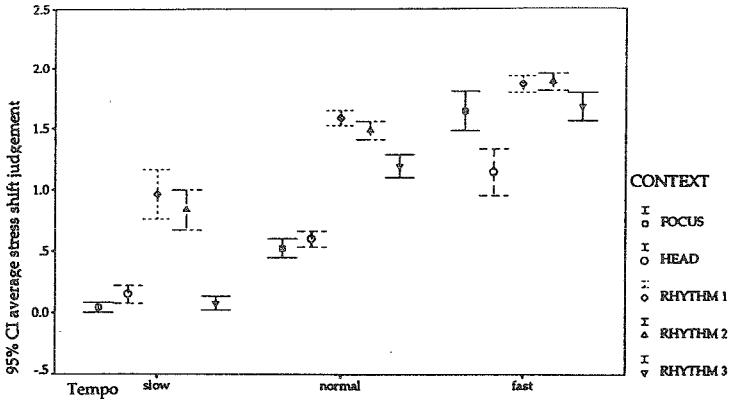
Recorded shift words, embedded in their noun phrase, were digitised at 20.8 kHz using the Soundscope speech signal processing program. The duration of the shift word, the duration of each foot, and the duration of the pause between the shift word and the following word was measured. In order to obtain a measure of variation in the duration of the first foot compared to the second foot, the duration of the first foot as a percentage of the duration of the whole word was calculated (henceforth described as percentage duration). The peak fundamental frequency for each foot was also calculated using a peak-picking algorithm within the Soundscope program. In order to obtain some measure of the relative changes in fundamental frequency pattern between the 2 feet over different contexts, the value for the second peak was subtracted from that of the first (henceforth described as fundamental

frequency shift). Rate of speech was calculated by dividing the duration of each sentence by the number of syllables it contained.

Each shift word, embedded in its noun phrase, was edited on to an audio tape in a pseudo-randomised order. Three phonetically trained linguists were asked to rate the stress levels in each shift word token as either: 1) the last stressed syllable is more prominent 2) both stressed syllables have equal prominence, or 3) the first stressed syllable is more prominent. These perceptual judgements were converted to a numeric expression of stress shift: zero for no shift, one for equal prominence, and 2 for full shift.

RESULTS

Figure 1. displays an error bar plot with a 95% confidence interval of the average stress shift judgement for each of the 5 contexts at normal, slow, and fast speaking rates.



Normal tempo

The results for speech at normal tempo are consistent with the predictions of the Rhythm Rule. Not only is there a strong perception of shift in the 3 Rhythm contexts, but the strength of the shift drops away as the syllable distance between the main stress in the shift word and the main stress in the following word increases. There is a clear pattern to the perceived stress shift judgements across the contexts. The 3 contexts in which shift is predicted (Rhythm 1, 2 & 3) demonstrate strong shift values, well above the "equal prominence" value of one. The 2 contexts in which shift is predicted not to occur (Head & Focus) demonstrate low shift values, well below the value of one. A one way analysis of variance for context against stress shift judgement with an adjusted least significant difference (Bonferroni) test at the .05 level indicated significant differences between the Rhythm and non rhythm contexts, and between the Rhythm 3 context (with 3 syllables distance) and the other 2 Rhythm contexts ($p = .000$, $F = 178.28$, d.f. = 4, 175).

There were phonetic changes to the shift words that corresponded to the pattern of stress shift judgement. In the Rhythm contexts the relative duration of the first foot was higher than in the non-rhythm contexts. There was also a positive increase in fundamental frequency shift. Inspection of the data indicated that these results were due not to absolute changes in the first foot of each word but to changes in the second. The absolute duration and fundamental frequency of the second foot decreased in the Rhythm contexts. This resulted in the relative prominence between the 2 feet shifting from the second to the first. Table 1 outlines the mean values for percentage durational change, fundamental frequency shift, and stress shift judgement for each context at normal tempo. One way analyses of variance for relative duration of the first foot against context, and for fundamental frequency shift against context were highly significant ($p = .000$, $F = 29.1$, d.f. = 4, 175; $p = .000$, $F = 34.83$, d.f. = 4, 175).

Table 1. Normal Tempo: mean values (and standard deviations) for average stress shift judgement, relative duration of first foot (percentage) and fundamental frequency shift (in Hertz) across the 5 contexts.

	Head	Focus	Rhythm 1	Rhythm 2	Rhythm 3
stress shift	.60(.19)	.52(.23)	1.59(.19)	1.49(.22)	1.20(.28)
duration	46.6(4.2)	44.9(5.2)	54.1(4.9)	54.5(4.0)	51.1(5.3)
f0 shift	-4.8(13.5)	-14.7(15.6)	14.3(13.5)	11.7(6.5)	6.5(1.7)

Slow tempo

The results for speech at the slow tempo conform to the predictions made by Giegerich (1981) and Hayes (1984) that the tendency to stress shift will decrease as the rate of speech decreases. On re-inspecting Figure 1, one can notice a similarity to the pattern at the normal tempo; there is a higher shift rating for Rhythm contexts 1 and 2 compared to the non-rhythm contexts, and the Rhythm 3 context has a lower shift rating than Rhythm 1 and 2. A one way analysis of variance for context against stress shift judgement with an adjusted least significant difference (Bonferroni) test at the .05 level indicated significant differences between the Rhythm 1 and 2 contexts and the other 3 contexts ($p = .000$, $F = 53.94$, d.f. = 4, 175). While this indicates that at a slow rate of speaking the syllable distance between main stresses still had a noticeable effect on the judges' perception of shift, it should be noted that the stress shift values for all contexts are markedly lower than those achieved at a normal rate and are below a rating of "equal prominence". This indicates a significant decline in stress shifting at the slow tempo.

There were phonetic changes in the shift words that corresponded to this marked decline in the perception of stress shift. In particular, the pattern of relative duration of the first foot across the contexts was significantly different from that obtained at normal tempo. A one way analysis of variance of relative duration of the first foot against context indicated that there were no significant differences in relative foot duration across the contexts. However, a one way analysis of variance for fundamental frequency shift against context indicated a significantly greater shift in fundamental frequency for the Rhythm 1 and 2 contexts compared to the other 3 contexts ($p = .000$, $F = 10.20$, d.f. = 4, 175). These results suggest that at a slow tempo durational adjustments to upcoming stressed syllables in the following word no longer occur. While there are quite marked fundamental frequency shifts at the slow tempo, their effect on the perception of prominence shift are slight. It appears that durational changes in the relationship between the 2 feet in a potential shift word are critical to the definite perception of stress shift. Table 2 outlines the mean values for percentage durational change, fundamental frequency shift, and stress shift judgement for each context at slow tempo.

Table 2. Slow Tempo: mean values (and standard deviations) for average stress shift judgement, relative duration of first foot (percentage) and fundamental frequency shift (in Hertz) across the 5 contexts.

	Head	Focus	Rhythm 1	Rhythm 2	Rhythm 3
stress shift	.15(.21)	.04(.12)	.96(.6)	.83(.48)	.07(.16)
duration	45.3(4.8)	45.7(4.7)	46.9(4.7)	46.0(7.3)	43.9(5.3)
f0 shift	-22.0(16.9)	-27.1(29.8)	-7.7(31.5)	1.5(16.6)	-1.4(19.4)

Fast tempo

The results for speech at the fast tempo are somewhat unexpected. Giegerich (1981) and Hayes (1984) had predicted that the tendency for the Rhythm Rule to apply would increase as the rate of speech increased. Figure 1 indicates that the most striking feature at fast tempo is, not that there is an increase in the perception of shift in the Rhythm contexts, but that there is a strong perception of stress shift in all contexts, including both Head and Focus. All contexts at the fast tempo had a shift rating above the "equal prominence" value of one, while all except the Head context had shift ratings higher than those of the 3 Rhythm contexts at the normal tempo. A one way analysis of variance of stress shift judgement against context with a modified least significant difference (Bonferroni) test

($p < .05$) indicated that the shift rating for the Head context, while higher than an "equal prominence" rating, was significantly lower than those for the other contexts ($p = .000$, $F = 21.43$, d.f. = 4, 175). On reflection, perhaps this is not surprising since one could assume a strong resistance by judges to mark stress shifts on words presented in contexts equivalent to their citation form! This very robust pattern of results suggest that words with the particular pattern of having 2 feet in the same word (hence 2 stressed syllables) permit stress to shift from the second foot to the first at fast rates of speaking even in "non-shift" environments.

This marked increase in the perception of stress shift in all contexts corresponded to phonetic changes in the shift words. In particular, the patterns of both relative duration of the first foot and fundamental frequency shift across the contexts were significantly different from those obtained at the normal rate. The relative duration values and the fundamental frequency shift values for all 5 contexts at the fast tempo were similar to those obtained only for the Rhythm contexts at the normal tempo. A one way analysis of variance of relative duration against context indicated no significant differences. A one way analysis of variance of fundamental frequency shift against context, with an adjusted least significant difference (Bonferroni) test ($p < .05$), indicated that there was a significantly lower f0 shift in the Rhythm 1 context compared to the Head and Rhythm 3 contexts. Inspection of fundamental frequency traces for the shift words in this environment indicates that this bizarre result, where there is least shift in the context supposedly most conducive to it, is an artefact of the fundamental frequency peak on the second foot being higher because, through durational compression, it is associated with the commencement of the pitch accent on the following word. The overall results for fast tempo suggest that duration and fundamental frequency adjustments associated with stress shifting are generally heightened and occur even when an upcoming word has no stress. These phonetic changes most likely reflect the effects of durational compression by the speakers. Table 3 outlines the mean values for percentage durational change, fundamental frequency shift, and stress shift judgement for each context at the fast tempo.

Table 3. Fast Tempo: mean values (and standard deviations) for average stress shift judgement, relative duration of first foot (percentage) and fundamental frequency shift (in Hertz) across the 5 contexts.

	Head	Focus	Rhythm 1	Rhythm 2	Rhythm 3
stress shift	1.15(.57)	1.65(.48)	1.87(.20)	1.89(.21)	1.69(.35)
duration	52.2(5.0)	52.5(2.6)	53.7(4.4)	51.7(6.3)	52.5(7.4)
f0 shift	11.4(9.3)	7.9(8.9)	3.3(11.8)	10.0(8.6)	11.9(12.2)

As a whole, the results of the stress shift perception ratings, and their associated changes in duration and fundamental frequency, indicate that the phenomenon of stress shifting is not only related to the phonological structure of the phrase in which a shift word occurs, but is intimately associated with a person's rate of speaking. The correlation between the rate of speech and stress shift judgement was high (0.65 ; $p = .000$, 2-tailed). A stepwise linear regression analysis, with a forward entering of variables, was carried out in order to determine which identified phonetic characteristics of the speech signal most contributed to the judges' perception of stress shift. Relative duration of the first foot, fundamental frequency shift, and rate of speech were entered as the predictors and average stress shift judgement as the dependent variable. The analysis achieved an R^2 value of 0.567 (0.565 adjusted) which was highly significant ($p = .000$). The regression equation was dominated by rate of speech which was extracted at the first step (R^2 value of 0.421 adjusted). At the second step fundamental frequency contributed to an R^2 value of 0..545 adjusted, while at the third step relative duration contributed to an R^2 value of 0.565 adjusted.

As well as the strong relationship between speaking rate and stress shift, the phonetic mechanism underlying the perception of a shift in prominence from the second foot to the first foot adds weight to the argument that the rhythm rule arises as a consequence of articulatory adjustments between stressed syllables. The "shift" did not involve an "action at a distance" with positive phonetic changes in the first foot, but involved changes in the second foot that adjoins the first syllable of the following word. In this sense the changes were quite localised responses, making timing and fundamental frequency adjustments to the following words stress structure.

DISCUSSION

The results from this experiment suggest that the usual formulation of the Rhythm Rule may be too restrictive. For these particular words that contain 2 feet, shifts in prominence from the second to the first foot can occur, or not occur, in any context, depending on the speaker's tempo. Current formulations of the rule appear to be an artefact of only observing the phenomenon at normal rates of speech. This has tended to obscure the direct and localised articulatory motivation for the shifts. The problem for pre-articulatory models of rhythm that only allow for metrical adjustments at that level of representation is to account for how an obvious performance variable such as tempo can have such a profound effect on the realisation of stress. If the underlying pre-articulatory representations of *the bamboo one* and *the bamboo chair* are the same for each tempo, even if they are "enriched" with phonological pauses, then, of themselves, they cannot account for the systematic variations observed in this experiment , especially stress shift in "non shift" phonological environments. Even if the formalisms can be modified to account for tempo as a variable, there is a more important conceptual issue at stake. While rhythm certainly operates at all levels of language organisation, and rhythmic adjustments may occur at a pre-articulatory level of representation, the direct articulatory motivation for synchronic stress shifting cannot be ignored. In the first metrical formulation of the Rhythm Rule by Liberman & Prince (1977), there were 2 levels of representation. The hierarchical organisation of linguistic stress within the utterance was represented by a phonological metrical "tree", but the rhythmic constraints, such as "stress clash", which motivated adjustments within the tree, were viewed as performance factors formulated within a seperate metrical grid representation. As Hayes (1984) has pointed out , the difficulty with uni- dimensional models of phonological rhythm such as those proposed by Prince (1983) and Selkirk (1984) is that they do not acknowledge that, while constrained by linguistic structure, the fundamental motivation for rhythmic adjustments are perceptual and motoric.

REFERENCES

Abercrombie, D. (1965) *Studies in phonetics and linguistics*, (O.U.P.: London).

Allen, G. D. (1975) *Speech rhythm: Its relation to performance universals and articulatory timing*, Journal of Phonetics, 3, 75-86.

Bolinger, D. L. (1965) *Pitch accent and sentence rhythm.* in Forms of English: Accent, morpheme, order, 163-169, (Harvard University Press, Cambridge, Mass.).

Ferreira, F. (1993) *The creation of prosody during sentence production*, Psychological Review, 150, 233-253.

Giegerich, H. J. (1981) *On stress-timing in English phonology*, Lingua, 51, 187-221.

Hayes, B. (1984) *The phonology of rhythm in English*, Linguistic Inquiry, 15, 33-74.

Jones, D. (1918) *An outline of English phonetics*, (C.U.P.: Cambridge).

Levelt, W. J. (1989) *Speaking: From intention to articulation*, (M.I.T. Press: Cambridge, Mass.).

Liberman, M. & Prince, A.(1977) *On stress and linguistic rhythm*, Linguistic Inquiry, 8, 249-336.

Lindblom, B. (1990) *Explaining phonetic variation: A sketch of the H & H theory*, in W.J. Hardcastle & A. Marchal (Eds.) Speech production and speech modelling, 403-441, (Kluwer: Dordrecht).

Prince, A. (1983) *Relating to the grid*, Linguistic Inquiry, 11, 19-100.

Selkirk, E. (1984) *Phonology and syntax: The relationship between sound and structure*, (M.I.T. Press: Cambridge, Mass.).