

## Timing Patterns in New Zealand English Rhythm<sup>1</sup>

Paul Warren

*Victoria University of Wellington*

### Introduction

In the discussion of linguistic rhythm, a distinction is often drawn between languages in which intervals between *stresses* are regular and languages in which *syllables* are regularly timed. While English is generally assumed to belong to the former category, recent research has suggested that New Zealand English (NZE) might be less strictly stress-timed than some other varieties, showing instead a tendency towards syllabic rhythm (Ainsworth 1993, Ainsworth and Holmes 1996, Holmes and Ainsworth 1996, 1997). These previous studies have based their conclusions on a qualitative measure of rhythm, i.e. the proportion of full-vowel realisations of potentially reduced vowels, such as the realisation of *the* as [ði] or [ðə]. The research presented in this paper examines the timing patterns that underlie these claims for syllable rhythm, using acoustic measures of vowel, syllable and foot durations.

In referring to stress timing and syllable timing, we are really describing tendencies towards stress- or syllable-based rhythms, since research has shown that 'there is no language which is totally syllable-timed or totally stress-timed — all languages display both sorts of timing ... different types of timing will be exhibited by the same speaker on different occasions and in different contexts' (Roach 1982: 78). Our consideration of NZE and other varieties is therefore a consideration of their relative position on a cline from strictly stress-timed to strictly syllable-timed rhythms. In line with Roach's comment above, we must also consider potential situational or stylistic influences. Amongst such factors listed by Crystal (1995) are child-directed speech, the expression of irritation or sarcasm, and styles such as *Airspeak* (a variety used in air-traffic control situations). The current study includes only read speech, but even here we will see that stylistic factors can play an important part in determining the rhythmic patterns of speech.

In addition, Crystal notes that the tendency towards syllable timing is influenced by varietal differences, and argues that syllabic rhythm is in many cases attributable to 'contact with a range of languages of diverse

---

<sup>1</sup> My thanks go to Helen Ainsworth for providing her tape recordings, to the Marsden Fund (grant number VUW604) for financial support, and to an anonymous reviewer for constructive comments on a previous version of this manuscript.

structural types, most of which have rhythms of a broadly isosyllabic character' (1995: 176). Amongst these varieties he lists Indian English, African Englishes including the Afrikaans-influenced South African variety, Italian- and Spanish-influenced English in the Americas, as well as pidgins and creoles such as Krio in West Africa and Tok Pisin in Papua New Guinea. Low (1998) suggests that the intonation and rhythm of Singapore English is likely to have been influenced by contact with a range of substratum languages (including a number of Chinese languages as well as Malay, Tamil and others).

It is not surprising that NZE rhythm has also frequently been discussed in the context of language contact, the relevant issue being whether there are distinctive rhythmic properties of Maori English that may arise from the mora-based timing structures of the Maori language, and which may affect the rhythm of other NZE varieties. A (relatively) early indication of a distinctive rhythm in Maori English comes in a study of Maori school-children speaking English, amongst whom Benton (1966) observed a tendency to use full vowels where other English speakers would most likely use reduced vowels. He comments that this results in a 'jerky rhythm'. The recent studies by Ainsworth (1993) and Holmes and Ainsworth (1996, 1997) provide a more thorough analysis that supports the greater use of full for reduced vowels in Maori English than in other NZEs. At the same time, these studies show that NZE in general shows a greater propensity for full vowels in the place of reduced than is found for a standard southern variety of British English.

### **Syllable-timing in NZE – previous evidence**

In her first study, Ainsworth (1993) considered the incidence of full-vowel realisations of potentially reduced vowels in samples of Maori English (ME), Pakeha New Zealand English (PE) and British English (BE). The vowels of interest occurred in words for which both reduced and full vowel forms are possible, taken from a list published by Gimson (1980), including the articles (*a*: /ə/ or /eɪ/; *the*: /ðɪ, ðə or /ði:/) as well as certain modal and auxiliary verbs, conjunctions, prepositions, etc. Only non-contrastive (unstressed) uses of these words were considered. If English has a stress-based rhythm, then the expectation is that these words will be produced with reduced vowels, since stress-based timing generally results in the reduction of unaccented syllables as these get compressed to fit into the rhythmic interval between stresses. Conversely, if the rhythm is based on a more equal interval between syllables, then there should be a greater tendency towards using full vowels, since these productions will reduce the timing contrast between stressed and unstressed syllables.

Radio broadcasts from a range of English-language sources available in New Zealand were analysed, representing the BBC World News (BBC), two commercial NZ stations (Windy and ZMFM), the more conservative NZ National Radio News (National) and the Maori news service, Mana News (Mana). The BBC sample could be characterised as "near RP".

Intended for the overseas market, it may be more conservative than many British English samples, but probably has a similar target audience to that of National Radio, i.e. "older people and the higher education and professional levels" (Bell 1991:111). Both the National Radio News and Mana News samples were broadcast on National Radio, and are characterised by Ainsworth as being directed at educated audiences and as having a clear, measured and deliberate style (Ainsworth 1993: 6). Of the commercial stations, Windy carried newscasts distributed by Independent News, aimed at middle-class middle-aged audiences. The style is more familiar than that of National Radio, but is still aimed at an older audience than ZMFM, which is also less formal than the other styles.

Ainsworth's auditory analysis of roughly 100 tokens of Gimson's words for each sample showed that Mana, with nearly 60% full-vowel forms, departs clearly from BBC and the commercial NZ stations, with around 20%. But, interestingly, National, the 'prestige' network, lies between these, with around 40%. Ainsworth conjectured that two forces are at work here — firstly there is a dimension running from BE through PE to ME, with increasing likelihood of full-vowel forms; secondly, there are differences in formality, such that within the PE varieties, the slower and more deliberate style of National results in an increase in full vowels relative to the faster Windy and ZMFM. These dimensions correspond to the variational and situational factors identified by Crystal (1995) in his observations on rhythmic change in English, and are summarised in Figure 1, where the difference between National and the commercial stations is shown as a parametric shift on a stylistic dimension.

In a follow-up study, Holmes and Ainsworth (1996) analysed recordings of young (20-35 year old) female and male Maori and Pakeha speakers of NZE, and again found more full vowels in ME than in PE, and also a greater tendency for full vowels amongst Pakeha women than men. Their research (see also Ainsworth and Holmes 1996, Holmes and Ainsworth 1997) also included a small-scale task comparison, in which they

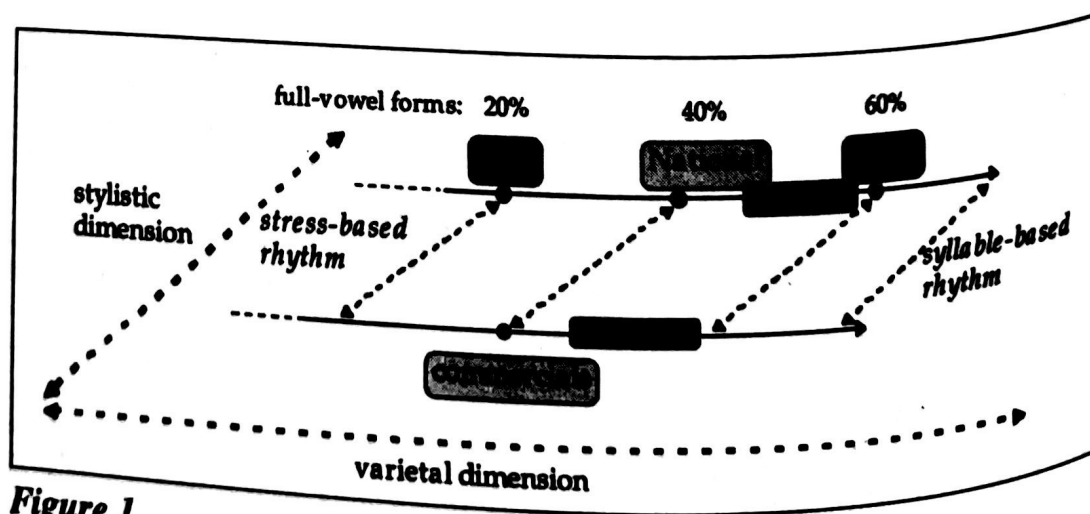


Figure 1

found differences between read and conversational speech for ME and BE speakers which fit the general pattern shown in Figure 1, with a decrease in the full vowel counts in conversational speech compared to the levels found in read speech.

Ainsworth and Holmes (1996, Holmes and Ainsworth 1997) also investigated age factors, with data from middle-age Maori males, Pakeha males and females, and from older Pakeha females. The PE data show no effect of age, all male groups having 15-16% full vowel scores, and all female groups in the 21-24% range. Although the middle-age Maori males use more full vowels than the younger Maori males (39% vs 30%), Ainsworth and Holmes attribute this to the degree of contact with the Maori language, which is higher for the middle-age group. The incidence of full vowel forms thus seems to depend on ethnicity and gender, and less directly on speaker age.

### **Acoustic analysis of the newsreader data**

While the stronger tendency for potentially reduced vowels to be realised in their full vowel form in NZE varieties is compatible with the assumption of a more syllable-based rhythm in NZE, it is not clear whether this pattern of vowel qualities is a *result* of a rhythmic difference or a *cause* of one. That is, the vowel quality difference may result from the nature of the vowel system of NZE (and particularly ME) compared to other varieties, and may in turn lead to differences in perceived rhythmic structures. Looking to the structure of Maori does not help much either. On the one hand, Maori is a mora-timed language, and so might be expected to exhibit a timing pattern that is more like syllable- than stress-timing, with less variation in syllable lengths. Contact with the Maori language, and with Maori speakers using such rhythmic patterns in English, could have led to the changing patterns of NZE, and there is some evidence for this in the data cited above. But on the other hand, Maori does not use reduced or centralised vowels to the same extent as English, although contact with English may be changing this. So a greater tendency towards using full vowels instead of reduced in NZE could be linked to influences from the Maori vowel system, rather than from rhythmic factors per se.

Ultimately it may not be possible to distinguish unequivocally between these two explanations, not least because the two phenomena — i.e. the use of full vowels instead of reduced, and of syllable-based rather than stress-based timing — may be opposite sides of the same coin, and may both be present in NZE and in Maori. Nevertheless, the current study aims to extend the research of Ainsworth and Holmes and to shed more light on this issue through an acoustic analysis of Ainsworth's newsreader data that includes measures of the timing properties of the entire texts, and not just of the potentially reduced vowels. Because the analysis is of newsreader texts, any conclusions we draw will initially have to be limited to the newsreader style, and in particular of course to read speech. However, as Ainsworth points out in her original study (1993: 2), each station attempts to cultivate

a certain style according to the perceived audience, and it is anticipated that any consequent style differences will include phonological features such as the type of rhythm typical of the variety. Since Ainsworth and Holmes' studies of further samples, including additional speech styles, have confirmed many of the initial findings of Ainsworth's study, we can be confident that the newsreader samples provide an appropriate starting point for the acoustic investigation of rhythmic patterns in NZE, though future research will need to extend the analysis to other samples.

The questions that are addressed in the course of this analysis are:

- (i) since the stylistic difference between National and commercial PE stations has been assumed to involve differences in rate of speech, are there measurable rate differences between the samples, and are they related to the differences in full vowel counts?
- (ii) is there evidence that vowel quality differences between the other samples may similarly be attributed to speech rate differences?

If the higher full vowel counts in NZE are indicative of a variety in which the rhythmic pattern is less stress-based and more syllable-based, then:

- (iii) is there consequently less variation in syllable durations across the entire NZE texts, compared with the BE sample?
- (iv) do the NZE samples conversely show greater variation in foot durations (inter-stress intervals) than BE?
- (v) is foot duration in NZE more dependent on the number of syllables in the foot than is the case in BE?
- (vi) is syllable duration in NZE more independent of the number of syllables in the foot than is the case in BE?

To address these questions, Ainsworth's tape recordings of newsreader speech were digitised, and durational measurements were made of each syllable. Syllabification was based on an assumption of maximal onset, i.e. that consonants between vowels are preferably analysed as belonging to the onset of the right-hand syllable, as long as this does not contradict the phonotactic constraints of English (thus /fʌ.u.nə.tæk.tɪk/). Exceptions were when the speaker obviously included consonants leftwards into the coda of the preceding syllable, such as before a prosodic phrase boundary (i.e. in situations where connected speech processes across word boundaries would also be unlikely). Syllables were identified as accented (i.e. both rhythmically and tonally prominent), stressed (rhythmically prominent but unaccented) or unstressed; they were also labelled according to whether they were the last syllable before a prosodic boundary (and as such susceptible to pre-boundary lengthening effects). A total of 3058 syllables was analysed from the five texts.

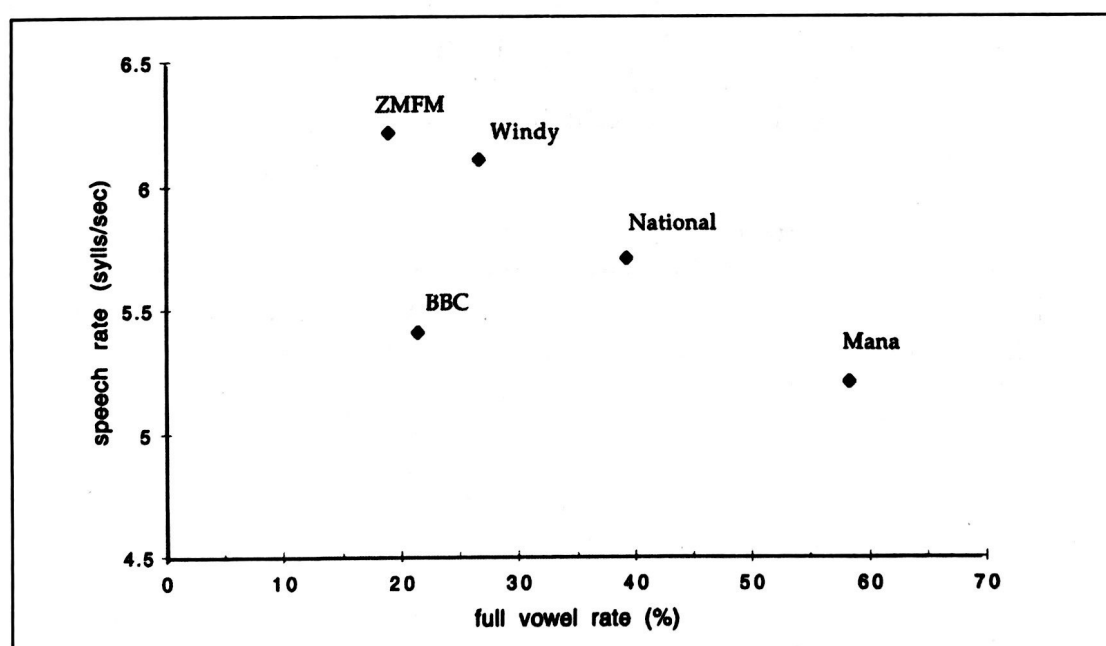
#### **Full vowels and speech rate**

To examine further the relationship between full vowel incidence and speech rate (which was assumed to be involved in the stylistic distinction between

National and the commercial stations), and to investigate whether rate was also a factor in the full vowel counts in the other samples, speech rate (in syllables per second) was calculated for each sample. In fact two rates were calculated, one including all non-pre-boundary syllables, and the other including only stressed non-pre-boundary syllables. The second calculation provides a measure of speech rate more clearly independent of the speakers' realisations of the word set with potentially reduced vowels, which were all unstressed. Since the same patterns for rate discussed below were found for both calculations, only the more inclusive one is presented here.

The rate data show that the commercial stations have a faster speech rate than National (Windy 6.10 syllables/second, ZMFM 6.21, National 5.75), and each of these is in turn faster than both the ME (5.18) and BE samples (5.35). The difference between National and the commercial stations confirms Ainsworth's observation that the newsreaders on the commercial stations are using a faster pace, and it is highly probable that this in turn leads to 'reduction and obscuration of unaccented words' (Gimson 1980:264), and hence the lower full vowel count. But note that the rate difference between National and Mana could similarly be taken as an explanation of the difference in full vowel count between these two samples, as is suggested by the plot of rate and full vowel data for the five samples in Figure 2.

Conversely, the BE sample has a high rate of vowel reduction (i.e. low full vowel count), despite having a low speech rate. One interpretation of these data is that rate determines the full vowel count across all of the NZE samples, and that the parametric difference is not the stylistic one



*Figure 2: Speech rate and full vowel counts in the five speech samples*

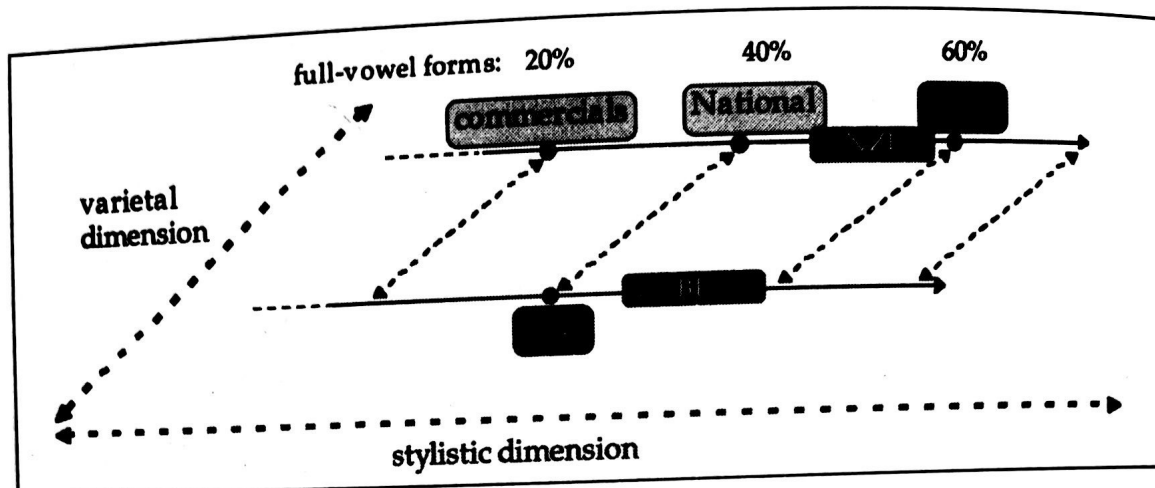


Figure 3

between commercial and National suggested by Figure 1, but a varietal one between NZE and BE, as in Figure 3.

Such a conclusion is clearly very different from Ainsworth's (1993) for the same data. It is also not in accord with the widespread observation of a more syllabic rhythm in NZE, and would require us to expect speech rate differences between Maori and Pakeha and between female and male Pakeha speakers as an explanation for the additional data reported for larger groups of speakers by Ainsworth and Holmes (1996, Holmes and Ainsworth 1996, 1997).

What is clear from the data displayed in Figure 2 is that the difference in full vowel realisations between NZE and BE cannot be accounted for simply in terms of speech rate. In order to investigate this difference further, we turn now to a more complete analysis of the five newsreader samples in terms of the syllable and foot timing patterns of the entire texts. Our basic hypothesis is that NZE will exhibit stronger tendencies towards syllable timing than BE, which will have patterns more compatible with stress timing.

### Syllable and foot timing patterns

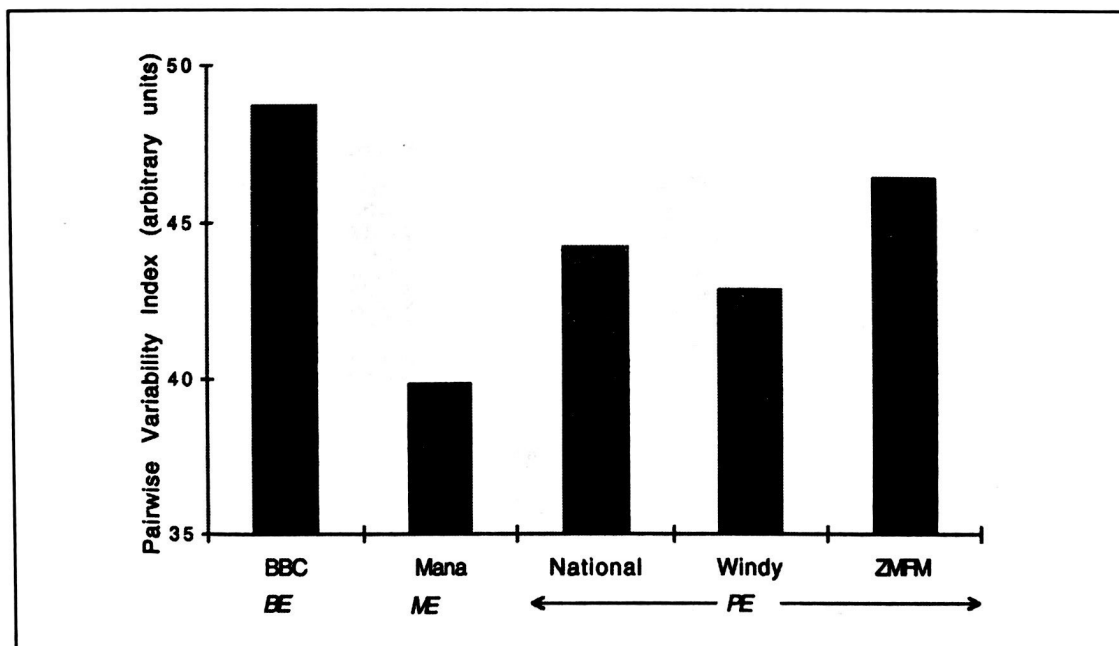
Since languages with syllable-based timing will have syllables of more equal length than stress-timed languages, in which unstressed syllables are typically shorter than stressed, a further measure of the position of the samples on the stress-syllable timing continuum comes from an investigation of variance in syllable durations. However, Roach (1982) reported standard deviations of mean syllable durations, and failed to find a consistent difference between stress-timed English and Russian on the one hand and syllable-timed French and Yoruba on the other. Low (1998) points out that the standard deviation is probably not a reliable measure of the variation in durations. To take an extreme example, a sequence of durations of 100-100-100-50-50-50 will give the same standard deviation (27.4) as a pattern of 100-50-100-50-100-50, and yet the rhythmic pattern of the two is

very different, with the former more compatible with two sequences of evenly timed syllables (e.g. syllable timing with a dramatic change in rate), and the latter more similar to the alternation of strong and weak syllables found in stress-timed languages. Instead, Low suggests a Pairwise Variability Index (PVI) based on the relative difference in duration of adjacent syllables, normalised for local rate variations, and calculated according to the following formula:

$$PVI = 100 * \sum_{i=1}^{n-1} ( |d_{i+1} - d_i| / ( (d_{i+1} + d_i) / 2 ) )$$

Applying this formula to the hypothetical examples above results in values of 13.3 and 66.7 respectively, reflecting the difference in rhythmic patterns, with a higher value showing more variability in durations.

Syllable PVI values were calculated for Ainsworth's five samples of newsreader speech, excluding pre-pausal syllables (see Figure 4). Analysis of variance reveals an overall effect of sample ( $F[4,2720] = 3.70, p < 0.01$ ), and post-hoc comparisons (Fisher's Protected LSD) show that the PVI for Mana is significantly lower than those for BBC, ZMFM and National (all at  $p < 0.05$  or better), and that BBC and Windy also differ from one another (at  $p < 0.05$ ). No other differences are significant. Clearly, the ME sample shows less variation in syllable durations than the other samples. Again this cannot be due simply to the slower rate of ME, since the BE sample, which also has a slow rate of delivery, has a significantly higher PVI. In addition, the Windy sample, with a higher speech rate, does not have a significantly higher PVI than ME.



*Figure 4: Syllable PVI across the five samples*



These syllable PVI values largely confirm the pattern of differences found by Ainsworth for potentially reduced vowels — most variation in BE, least in ME, with PE samples between these, although a PVI difference between National and the commercial PE stations is not apparent. Since the calculation of PVI normalises for speech rate, these data confirm that the full vowel count differences between PE samples, but not those between PE and ME, are due to rate differences.

The lower PVI found for ME means that there is less inter-syllable variation in duration in that sample, which in turn is compatible with the conjecture that ME rhythm is closer to syllable timing. Conversely, a more stress-timed language should show less variability in the durations of stress intervals, or feet. To investigate this expectation, a PVI analysis was also carried out on foot durations. Feet were defined here in terms of rhythmic prominence, i.e. they included a stressed (or accented) syllable and all following unstressed syllables up to but not including the following stressed syllable, or up to a pause. The approximate average length of a foot in all of the samples was 2.7 syllables, and average foot duration ranged from 443 to 521 msec, reflecting differences in speech rate. PVI values were calculated for each sample according to the formula given above. They are presented in Figure 5.

The foot PVI data do not support the difference between stress- and syllable-timed samples as clearly as the syllable PVI data. There is no overall effect of sample on the foot PVI values ( $F(4,923)=1.90, p>0.10$ ). However, note that the three samples which exhibited the largest syllable PVIs (BBC, National, ZMFM) have the lowest foot PVIs, while the low

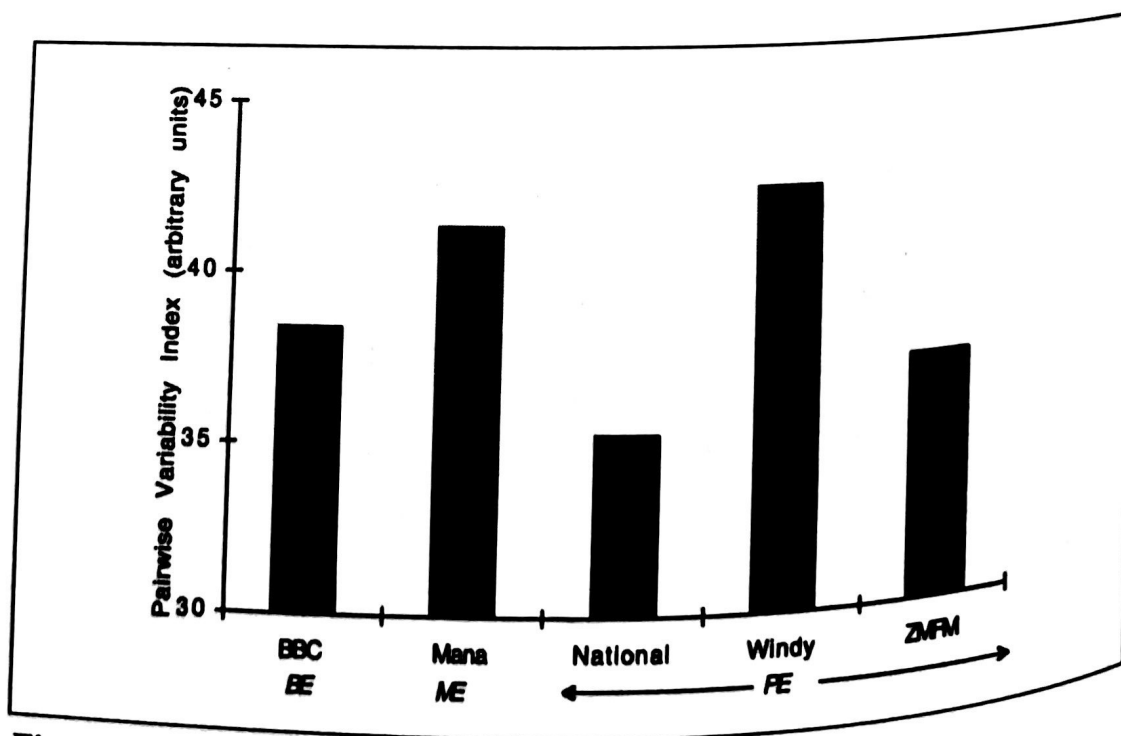
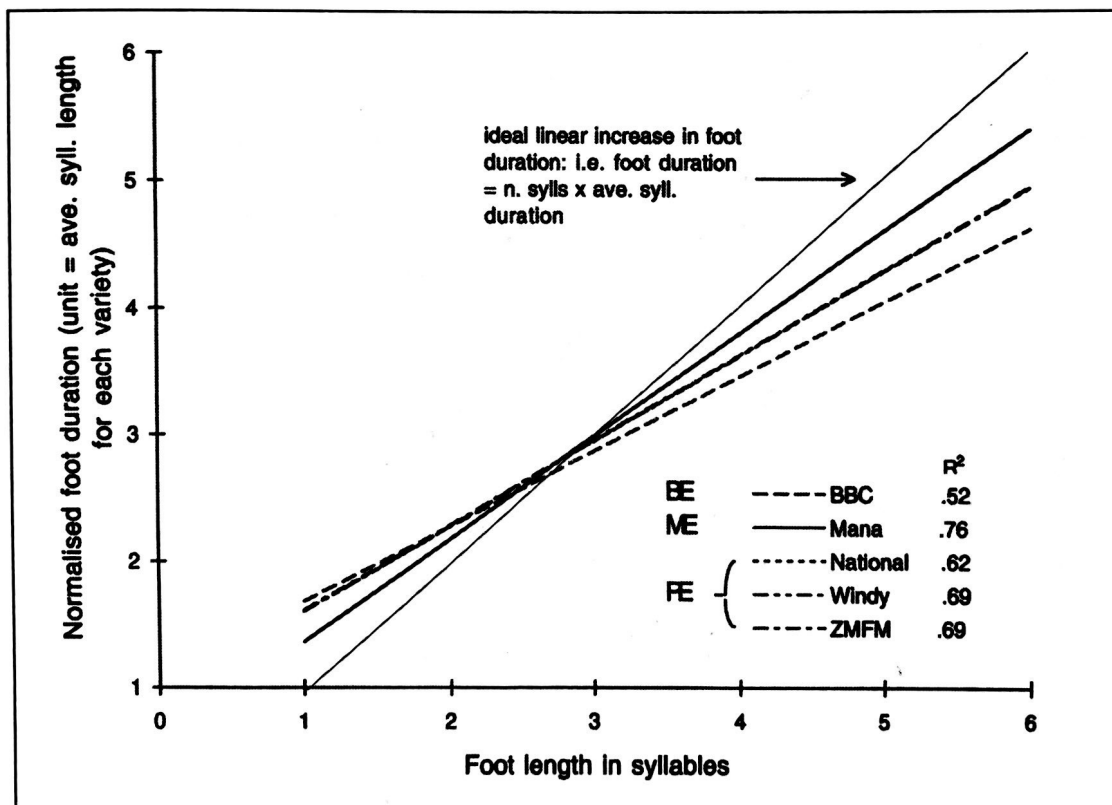


Figure 5: Foot PVI across the five samples

syllable PVIs of Mana and Windy correspond to high foot PVIs. This pattern may well be indicative of the relative importance of the foot and the syllable as the unit of rhythmic organisation.

Our final pair of measures explores the relationship between syllables and feet in the samples. Specifically, given that the PVI data have shown that there is variation in foot and syllable duration, what are the constraints on such variation, and do these in turn reflect the different rhythmical tendencies of the samples? First, is foot duration more dependent on the number of syllables in the foot in NZE than in BE because of the greater tendency in NZE towards a syllable-based rhythm? Second, as feet increase in size in terms of numbers of syllables, are these syllables more likely to be compressed in BE than in NZE, because of the former's stronger tendency to stress-based timing?

The first of these measures involves a regression of foot duration (in msec.) on foot length (in syllables) for each sample. The duration measure was re-expressed as a function of the average syllable duration for each sample (to normalise for speech rate differences between the samples). If the rhythm is strictly syllable-based, then the duration of a foot will be the product of the number of syllables and the average syllable duration, and the regression function should be perfect. In fact, regression functions were strong, and the syllable count accounted for a significant proportion of the



**Figure 6: Regression functions of foot duration and length**

variance in foot duration in all cases (compare the  $R^2$  values given in Figure 6). The functions were stronger for ME than for the PE samples, and weakest for BE. In addition, comparisons of the regression functions (Cohen and Cohen 1983) showed that the slope (0.81) is significantly steeper for ME than for all other varieties, which do not differ from one another (average 0.62).

Although the slope for ME does not reach unity (i.e. foot duration is not perfectly determined by syllable count and average syllable duration), it is clear that the duration of feet is much more dependent on the size of feet in syllables for this sample than for the others. Note also that even the BE sample is far from being strictly stress-timed, which would correspond to a horizontal line on Figure 6.

The second measure, assessing the degree of compression of syllables as feet get larger, required a regression of syllable duration in a foot (again normalised for speech rate differences across the samples) and the length of the foot in syllables. The data were, for each sample, modelled better by a logarithmic function than a linear one, reflecting the fact that syllable compression becomes increasingly difficult as the syllable becomes shorter. The curves of the regression functions, and the variance in syllable duration accounted for by the regression ( $R^2$  values) are given in Figure 7.

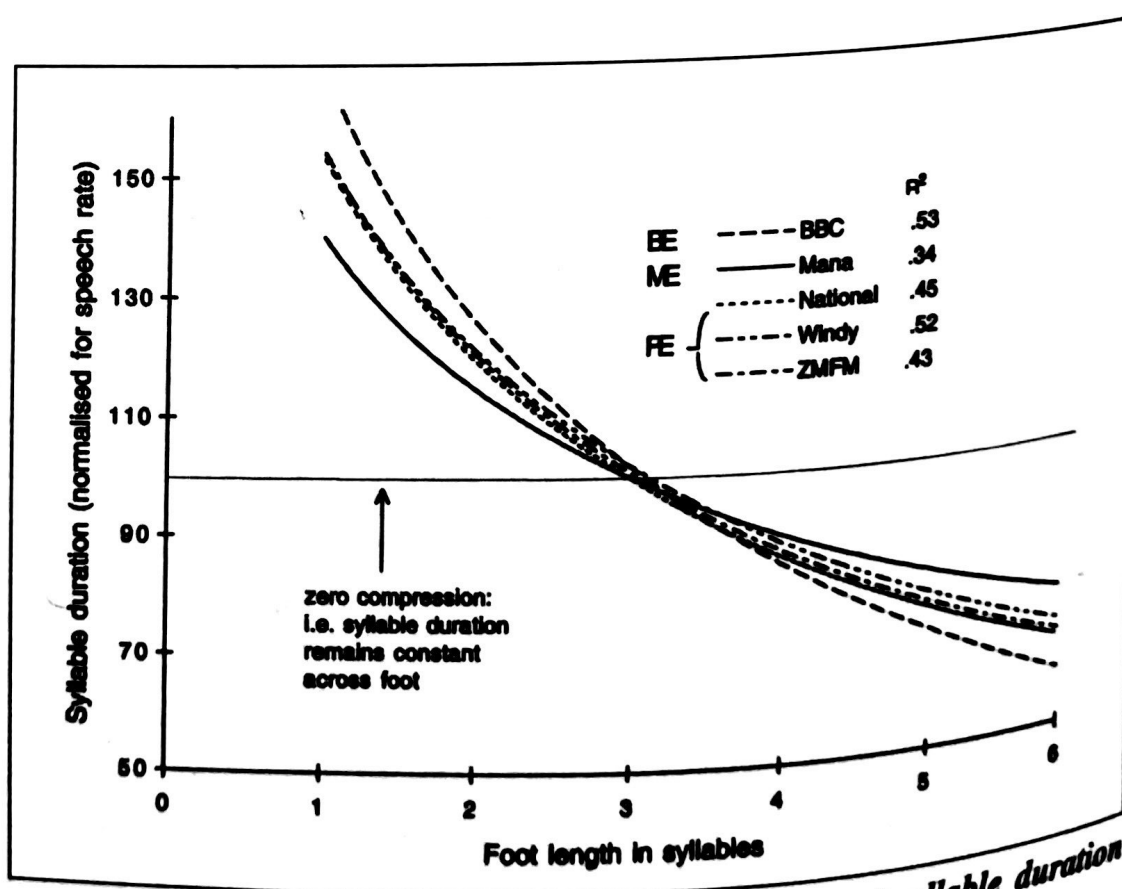


Figure 7 Compression effects: regression function of syllable duration and foot length

The regression function (and therefore extent of compression) is stronger for BE than for PE, and weakest for ME. A pairwise comparison of the slopes of the functions confirms that ME shows significantly less compression than BE, that both BE and ME differ from each of the PE varieties, and that the PE varieties do not differ from one another. Clearly the constraining influence of stress-based rhythm is stronger in BE than NZE, and stronger for PE than for ME.

### **Conclusion**

The acoustic analysis of the newsreader samples presented above has shown that no sample is entirely stress-timed, since they each show significant increases in foot duration as the number of syllables in a foot increases. Nor is any sample entirely syllable-timed, as all samples show significant compression of syllable durations as the number of syllables in a foot increases. However, there are clear differences between the timing features of the samples. These differences relate to the research questions posed earlier in this paper.

- (i) The acoustic analysis provided empirical support for the speech rate difference between National and commercial PE stations assumed by Ainsworth (1993) to be a factor in determining full vowel incidence.
- (ii) In addition, the data suggest that speech rate may account for differences in full vowel realisations across the complete set of NZE samples, and not just the PE samples. BBC does not fit in the same rate-vowel quality series, having a similar speech rate to Mana, but very different full vowel scores (as shown in Figure 2).

However, the analyses carried out to address the remaining questions, which included analysis of syllable and foot patterns across the entire texts of each sample, showed that there is more to the differences in timing of the samples than a simple difference in speech rate:

- (iii) There is less variation in syllable durations in the NZE texts than in BE, but within the NZE samples, we find significantly less variation for ME than PE. This suggests that NZE is more syllable-timed than BE, and ME more syllable-timed than PE.
- (iv) However, foot duration did not vary more in the NZE samples than in BE, contrary to expectations based on BE being more stress-timed than NZE.

The constraints on foot duration appear to differ between the varieties:

- (v) In particular, ME shows a significantly stronger relationship between foot duration and syllable count in the foot than any other sample.
- (vi) Finally, syllables in BE are more likely to be compressed as the number of syllables in a foot increases, and this compression is also stronger in PE than in ME.

Importantly, the analyses relating to questions (iii) to (vi) show that the PE samples do not differ in terms of foot and syllable timing, suggesting that these measures are relatively independent of speech rate, and that it is rate alone that determines the differences in full vowel realisations in the National and commercial PE samples.

Overall, the results of the acoustic analyses presented in this paper show clear differences between British, Pakeha and Maori English timing structures. BE and ME have similar rates of speech but foot duration in ME is more dependent on foot length, and there is less syllable compression as the foot gets longer. On a continuum from stress-based to syllable-based rhythm, ME is nearer syllable-based than BE.

PE shows less variability in syllable duration than BE, and has a lower degree of syllable compression. This is consistent with a tendency towards syllable-based timing in PE. This tendency is not as strong as that found in ME, where there is significantly less variability in syllable durations, a stronger relationship of foot length and syllable count, and less compression of syllables as feet get longer.

This more detailed analysis of Ainsworth's recordings of newsreader speech provides additional support for the claim that there are rhythmic differences between the samples collected, representing a greater extent of syllable-based timing in New Zealand English than in British English, and in Maori English than in Pakeha English. These differences are not limited to the patterning of full vs reduced vowel forms of the words in Gimson's list, but affect the basic timing structures of the varieties. Clearly further work is necessary before we are able confidently to generalise beyond these five samples of newsreader speech. Research in progress, for instance, is applying the PVI measure to reading lists in which sentences with only full vowels are paired with sentences with a mix of full and potentially reduced vowels. The data are being collected from male and female Maori and Pakeha New Zealanders, and will be directly comparable with Low's data for British and Singapore English (Low 1998). The same data collection exercise is also providing story texts as well as monologues and conversational data from the same speakers, allowing a comparison of styles. Finally, research in this area will need to assess the exact nature of the timing differences between varieties — for instance, is the tendency in NZE towards syllable-timing, or is it (reflecting the Maori influence) towards something that might be more appropriately described as mora-timing?

### References

- Ainsworth, Helen. 1993. 'Rhythm in New Zealand English'. Manuscript, Victoria University of Wellington.
- Ainsworth, Helen and Janet Holmes. 1996. 'Syllable timing in New Zealand English'. Paper presented at the Fifth New Zealand Language and Society Conference. Christchurch, June 21-23, 1996.

*Timing Patterns in New Zealand English Rhythm*

- Bell, Allan. 1991. *The Language of the Media*. Oxford: Basil Blackwell.
- Benton, Richard. 1966. *Research into the English Language Difficulties of Maori School Children 1963-1964*. Wellington: Maori Education Foundation
- Cohen, Jacob and Patricia Cohen. 1983. *Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
- Crystal, David. 1995. 'Documenting rhythmical change'. *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor*. Ed. by Jack Windsor Lewis. London: Routledge. 174-179.
- Gimson, A.C. 1980. *An Introduction to the Pronunciation of English*. 3rd ed. London: Edward Arnold.
- Holmes, Janet and Helen Ainsworth. 1996. 'Syllable timing and Maori English'. *Te Reo*, 39: 75-84.
- Holmes, Janet and Helen Ainsworth. 1997. 'Unpacking the research process: investigating syllable-timing in New Zealand English'. *Language Awareness*, 6,1: 32-47.
- Low, Ee Ling. 1998. *Prosodic Prominence in Singapore English*. Unpublished PhD dissertation, University of Cambridge, UK.
- Roach, Peter 1982. 'On the distinction between stress-timed and syllable-timed languages'. *Linguistic Controversies*. Ed. by David Crystal. London: Edward Arnold.