A COMPUTATIONAL APPROACH TO STRESS PATTERNS IN PENRHYN

Ayako Yasuda-Graefe and Volker Graefe

1. PURPOSE

'Stress' or 'accent' in the Polynesian languages has been described as predictable, dependent on penultimate-and-alternate-preceding-vowels rules and first-vowel-and-long-vowel-precedence rules (Biggs 1961; Hohepa 1967; Newbrand 1951). However, many exceptions to such rules have also been reported. This is also the case in Penrhyn, a Polynesian language spoken in one of the Northern Cook Islands (Yasuda 1968).

The first difficulty when dealing with the phenomenon of stress is to identify its nature objectively. In this study we use the term 'stress' to indicate the prominence of a syllable relative to other syllables in a stretch of speech (Folkins et al 1975). It is one of our goals to find an objective, quantitative measure for the perceived stress. Such a measure would be valuable in two respects: It would facilitate the formulation and the testing of theories or rules related to stress patterns, and it would be the first step towards the utilisation of stress to improve the performance of speech understanding machines.

It is also our goal to demonstrate that all the necessary experiments for such studies can efficiently be conducted using no other tool than an ordinary digital computer with a few simple peripheral devices to input and output speech signals.

A third goal is to find out to what extent modelling the psychoacoustic properties of the human auditory sense can contribute to a better understanding of the nature of human speech and to the development of better speech-understanding machines.

2. METHOD

What we perceive as emphasised or 'prominent' is related to other perceptual qualities such as pitch, loudness and tempo of speech segments, whose acoustic correlates in many languages have been said to be changes in fundamental frequency, intensity and syllable duration respectively (e.g. Lehiste 1970; Folkins et al 1975). It is therefore necessary to extract these acoustic parameters from the speech signal.

A great variety of instruments exists for the analysis of speech signals. Among them are sound spectrographs, level meters, filters, fundamental frequency meters and many more. Using these instruments correctly is not always easy.

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 357-376. *Pacific Linguistics*, C-93, 1986.

Yasuda-Graefe, A. and Graefe, V. "A computational approach to stress patterns in Penrhyn". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*. **C-93:357-376**. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.357*

Some of them are quite complicated and if the operator does not fully understand their internal working and their inherent limitations, gross errors may occur. All these instruments tend to age and to give unreliable results if they are not properly and diligently maintained. Moreover, if the same recording of a speech signal is analysed with different instruments it can be difficult to establish the exact temporal correspondence between their outputs. (Does the peak in fundamental frequency precede or follow the peak in intensity?)

We have avoided all these difficulties by using only one piece of equipment, a digital computer. A suitably programmed digital computer can perform all the functions of the various instruments mentioned above and many more. Therefore our speech samples were stored in digital form in the computer and all analysis was performed using the methods of digital signal processing.

There are great advantages to this technique, especially in respect of precision, reproducibility and ease of use. A digital filter, for instance, will never drift nor age; if so designed it will not introduce any phase lag, nor noise, nor distortion; its passband will be flat and its rolloff will be steep. All of its characteristics can easily and predictably be modified just by changing a few lines of program code. The same holds true for all other methods and tools of digital signal processing.

2.1 Equipment

Figure 1 shows the equipment we used for our experiments. The speech samples were available as tape recordings and an ordinary tape recorder was used for playback.

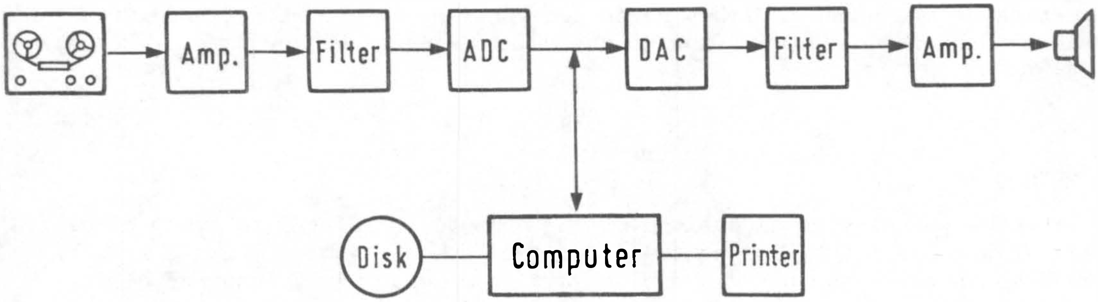

Figure 1: The equipment used for the analysis of the speech data

To store speech samples in the computer, the output from the tape recorder was passed through a pre-amplifier and a bandpass filter (70 Hz - 10 kHz) to an analog-to-digital converter (ADC) with 12 bits of resolution and a sample rate of 25 kHz. The digitised speech signal was fed into the computer, and simultaneously into a digital-to-analog converter (DAC), also with 12 bits of resolution. The reconstructed analog signal was again bandpass-filtered (70 Hz - 10 kHz), amplified, and fed into a loudspeaker to permit the continuous monitoring of the digitised signal.

The same setup, with the ADC switched off, was used to listen to speech samples after they had been stored in the computer.

The computer is a Perkin Elmer 3210 with 2 MB of memory and 160 MB of mass storage capacity. A matrix printer (Facit 4542) with limited graphics capabilities was used to produce all hard copy output, including curves and spectra.

As figure 1 shows, the hardware we have been using is very simple, as all the complexity is hidden within the computer and its software. Because of its simplicity it is reliable and easy to operate; since the equipment involved is always the same, regardless of the type of analysis being performed at any given moment, nothing ever has to be moved around or reconfigured.

The equipment did not introduce any audible noise or distortion; therefore we did not find it necessary to conduct specific measurements to evaluate its performance quantitatively.

2.2 Data

The speech samples were recorded during a field study in 1966-67 in Auckland, New Zealand. A small consumer-type battery operated tape recorder with a primitive microphone was used to record the utterances of the informants. Although several informants from Penrhyn supplied data, the main informant was an elderly woman, who spoke only Penrhyn and Rarotongan. The data were analysed grammatically in detail by Yasuda (1968).

Unfortunately the recordings are of poor technical quality in several respects:

(a) Noise

The dynamic range is only about 35 dB. Not only tape hiss is present but also motor noise from the tape recorder, traffic and household noise (the recording was done in the informant's living room) and sometimes even voices in the background.

(b) Frequency range

Practically no signal is present above 3 kHz.

(c) Nonlinear distortions

Occasionally the tape recorder was overloaded in spite of its automatic volume control.

These deficiencies are severe. The high noise level limits our studies to the louder parts of each utterance, while the limited frequency range limits us mostly to the voiced parts.

On the other hand, for many practical applications it is important to develop the ability to analyse speech signals which are band limited and corrupted by noise. Also, we are interested in stress patterns, and these stress patterns are clearly audible even in our low-quality tape recordings.

2.3 Programs

2.3.1 Speech input/output

All input and output of speech signals is performed by one interactive program. It includes the following functions:

- Transfer of selected speech samples from the tape recorder into the computer's memory;
- Transfer of speech samples, together with written comments, between storage disk and main memory;
- Playback of stored speech samples (either the entire sample or only selected parts of it).

In playback mode the operator can control the exact start time and end time (relative to the beginning of the speech sample) of the segment he wants to hear. Start time, end time, or both can also be incremented or decremented automatically in operator selected steps. This makes it fairly easy to determine with great accuracy the borders of each audible segment within the speech sample.

2.3.2 Sound pressure level

A typical plot of sound pressure level versus time is shown in figure 2. The level is computed for each 10-ms-interval. Since the actual sound pressure level at the time of the recording is not known, it is assumed that the strongest signals which were recorded on the tape correspond to a sound pressure level of 80 dB.

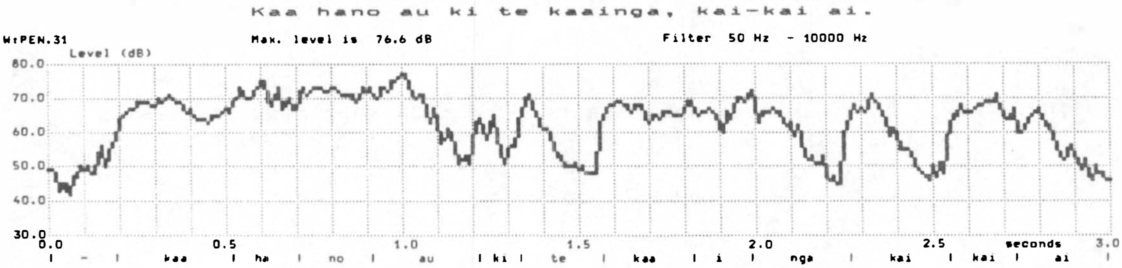

Figure 2: Typical plot of computed sound pressure level versus time

(As in all of the following computer-printed diagrams, the width of the original diagram is about 30cm, and the time scale is relative to the beginning of the digital speech sample.)

It is possible to apply various numerical filters to the data before computing the level. The important advantages of numerical filters compared to physical filters have already been pointed out: all their characteristics are freely selectable and perfectly stable, and they do not introduce any noticeable phase lag, distortion or noise.

2.3.3 Spectrogram

Sound spectrograms have been an important means to analyse speech signals since they were introduced by Koenig et al (1946). Figure 3 shows a sound spectrogram as it has been produced by the computer. Compared to using a conventional sound spectrograph, computing the spectrogram has several advantages.

Temporal and spectral resolution can easily be varied, there are no errors caused by phase lag, noise, or distortion and all parameter settings are perfectly reproducible.

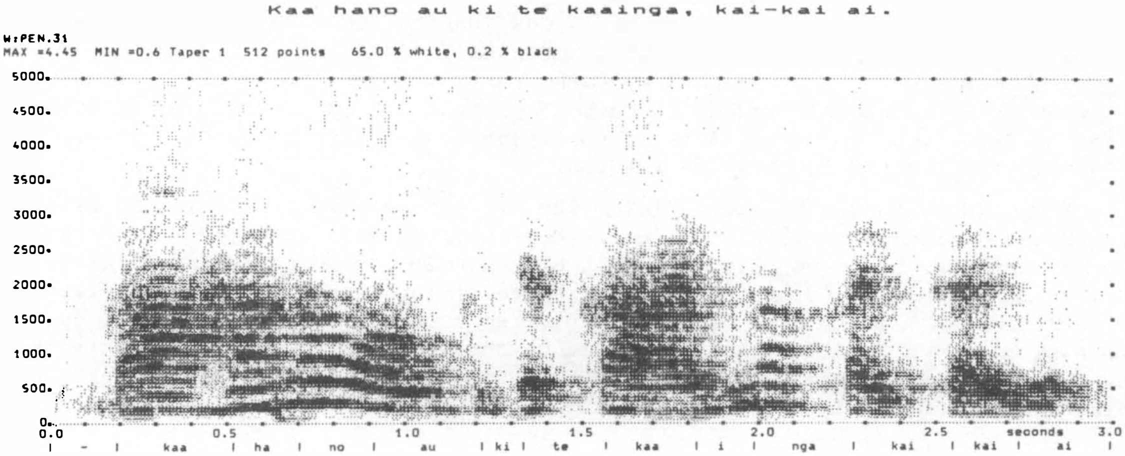

Figure 3: Typical computed sound spectrogram

(The frequency range in this example is 0 to 5000 Hz.)

In figure 3 one spectrum was computed for each 10-ms-interval. Because in speech signals the high frequency components tend to be weaker than the low frequency components, the data were differenced, which is equivalent to using a 6 dB per octave high-pass filter. For each spectrum 512 data points were used (0.02s) and a cosine window was applied before performing the fast Fourier transformation. This results in a spectral resolution of about 100 Hz.

2.3.4 Loudness

All speech signals consist of louder and softer segments, and it is reasonable to expect that the fluctuations of loudness are related to the syntactic or semantic structure of the utterance. Loudness is not a physical quantity but a perceptual or subjective one; two sounds are of equal loudness if the 'average' listener perceives them to be equally loud. Loudness is, of course, related to physical quantities like sound pressure and frequency, but in a very complicated way which is not fully understood yet, in spite of many years of effort by numerous researchers. The basic facts are (Zwicker and Feldtkeller 1967; Zwicker 1982):

- If the frequency of a sound signal is varied while its intensity and other characteristics are kept constant, the loudness varies in a characteristic way. The loudness is greatest in a frequency range around 1 kHz to 4 kHz.
- If the intensity of a sound signal is varied while all other characteristics are kept constant, the loudness doubles whenever the sound pressure level is increased by 10 dB. (Strictly speaking, this is true only if the sound pressure level is at least 40 dB.)

- The auditory system forms critical (frequency-) bands. Below 500 Hz the width of each critical band is approximately 100 Hz, above 500 Hz it is approximately 20% of the centre frequency. If a sound signal has several components within the same critical band, they contribute to the loudness according to the sum of their intensities. If it has several components in different critical bands, components in one critical band can mask components in other bands. The masking is not symmetrical, essentially only components at higher frequencies are masked by components at lower frequencies. If a component is completely masked it cannot be heard and it does not contribute to the perceived loudness; if it is partly masked, it can be heard, but its contribution to the loudness is reduced.

These facts were established in series of psychoacoustic experiments using static sound signals which did not carry any information. Probably the perception of loudness of speech signals which are dynamic and do carry information is even more complicated, but we are not aware of any experimental data related to such signals. Therefore, we wrote a program to compute the loudness of speech signals, based on an algorithm that would be correct for static noise. This is not a completely satisfactory method, but we consider it a first step towards a more adequate model of the auditory system.

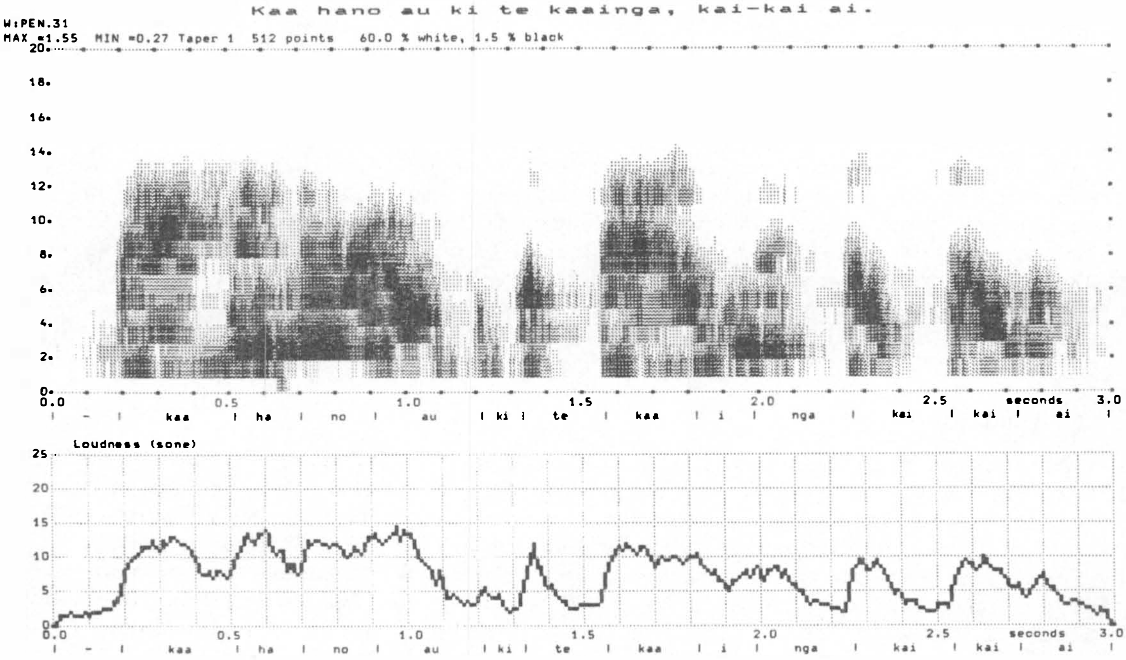

Figure 4: Typical output of the loudness program

(Top: specific loudness versus time and pitch;
pitch is given in units of Bark = 100 mel.
Bottom: computed loudness versus time.)

Figure 4 shows typical results of the program to compute loudness. The program includes the following steps:

- Computation of the power spectrum as described in section 2.3.3;
- Combining adjacent spectral lines into critical bands;
- Computation of loudness and specific loudness, using a subroutine by Paulus and Zwicker (1972). Specific loudness indicates how strongly components of different pitch contribute to the total loudness. Pitch is a perceptual quantity, corresponding closely to the physical quantity frequency.

2.3.5 Fundamental frequency

Several methods are known to determine the fundamental frequency of a speech signal (for an overview see, e.g. Markel and Gray 1976). They all utilise one of two basic ideas: they either evaluate a periodicity in the frequency domain (which also manifests itself as a pattern of horizontal stripes in the sound spectrogram), or they evaluate a periodicity in the time domain directly. Finding the fundamental frequency automatically is difficult, and apparently no method is known which always works reliably.

We have used two methods simultaneously: a sound spectrogram to indicate the general trend of the fundamental frequency within an utterance, and the frequency of zero-crossings to get quantitative data (figure 5). Where the zero-crossing-data are too irregular we use the spectrogram to correct them manually.

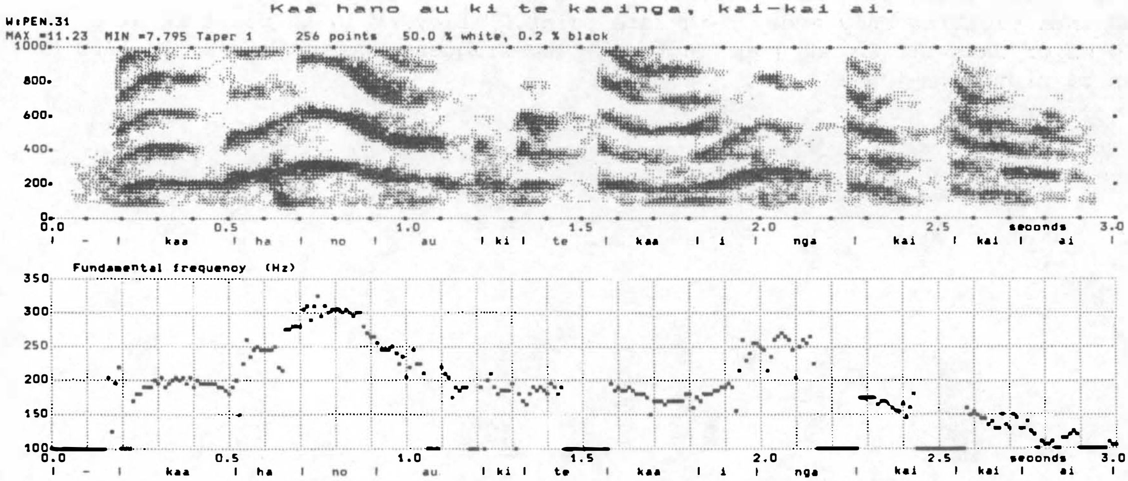

Figure 5: Output of the program for finding the fundamental frequency

(Top: power spectrum, 0 to 1000 Hz.
Bottom: fundamental frequency from zero-crossings.)

Before the zero-crossing rate can be used to compute the fundamental frequency, all components of the speech signal which are outside the fundamental frequency range should be filtered out. In our speech samples the fundamental frequency is almost always between 100 Hz and 300 Hz, therefore we used a filter with a passband from 100 Hz to 300 Hz for most of our speech samples.

A threshold is used to detect pauses and unvoiced segments; ideally a value of zero should in such cases be returned for the fundamental frequency, but because of the poor signal-to-noise-ratio of our speech samples it is difficult to select the correct value for the threshold.

For a fully automatic system it would probably be necessary to use adaptive filters and to adjust the threshold automatically according to the momentary signal level, but for our purpose it was acceptable to use constant settings of the filter and of the threshold, and to correct irregular data manually.

2.3.6 Waveform plot

Sometimes it is useful to scrutinise the data in very great detail, for instance, to analyse some unexpected phenomenon that can be heard, to determine the exact location of syllable borders in situations where even careful listening yields no clear result, or to find out why the fundamental frequency program behaves erratically at a certain point within the speech sample.

For these and similar purposes a program exists which plots the waveform data on paper (figure 6).

The difficulty with such a program is the mass of data, 25 000 values for each second of speech. If every single data point is plotted, 36 ms of data fill one standard page. The plot quickly becomes unreasonably large, unless only very short sections of the signal are plotted. Therefore, in another mode of operation, the program compresses the data by first applying a low-pass filter and then plotting only every n-th data point. If every 10-th point is used, 600 ms of data fit in one page, but only low-frequency components up to 1200 Hz can be plotted.

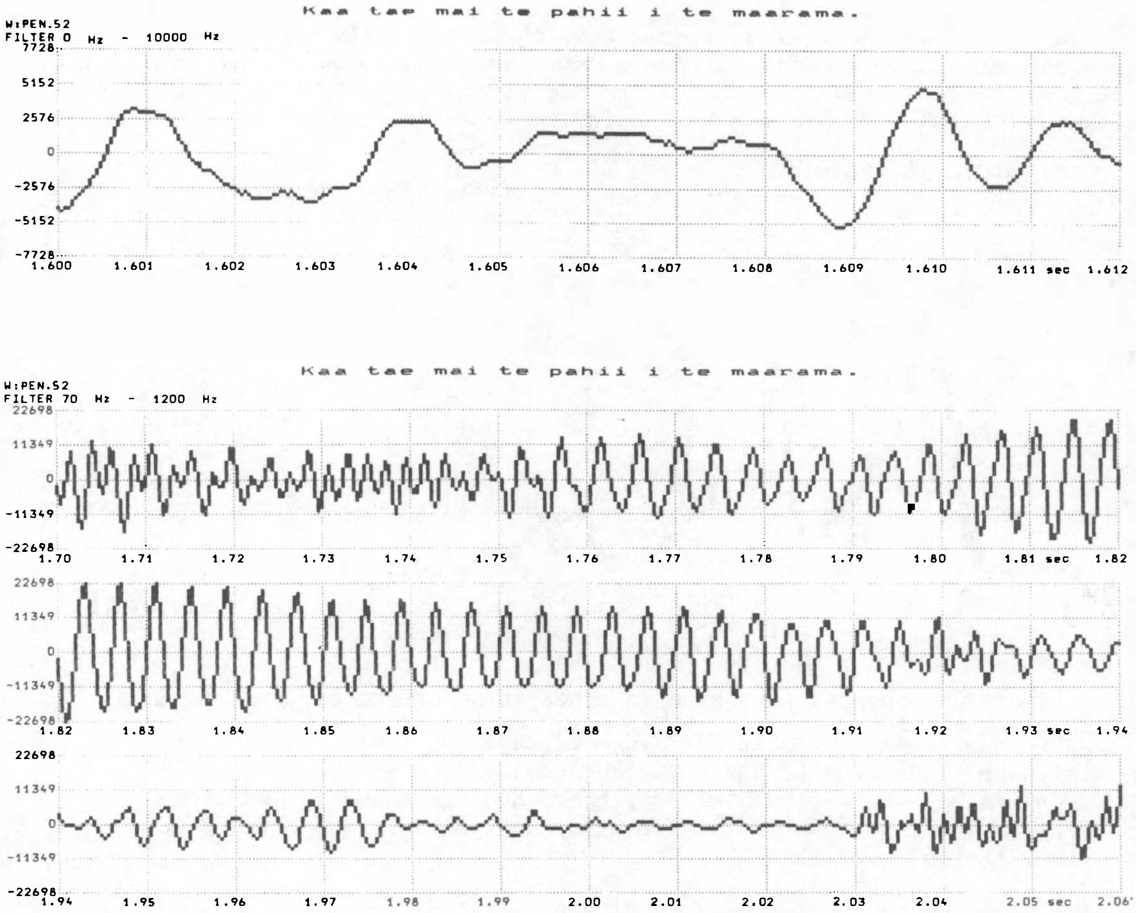

Figure 6: Waveform plots

(In the upper diagram every data point of a 12-ms-section of an utterance has been plotted, in the lower diagram every tenth point of a longer section has been plotted after removing all high frequency components above 1200 Hz.)

3. THE NATURE OF STRESS

In this section we discuss the relationship between perceived stress and the acoustic parameters which we have extracted from the data. In particular, we study loudness, syllable duration and fundamental frequency as possible counterparts to perceived stress. Since stress is primarily an attribute of syllables, we also study the correspondence between acoustic parameters and the syllable structure of utterances.

3.1 Loudness and intensity

In declarative sentences the maximum loudness usually occurs in one of the first few syllables; then the loudness decreases slowly towards the end of the sentence. There are utterances that do not fit into this pattern, but we have no explanation for it. Stress has no influence on the deviation.

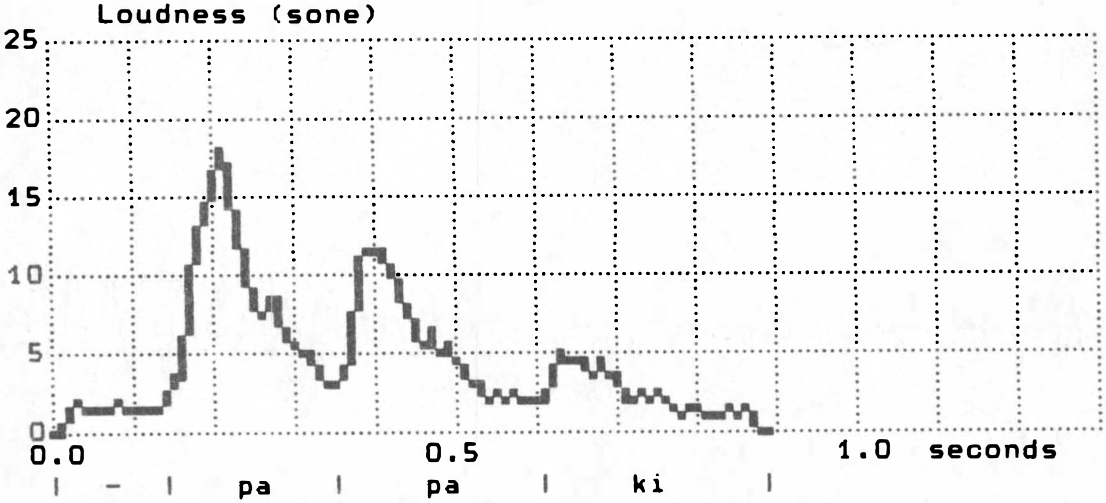

Figure 7: Loudness (measured in sone) in an utterance: papAki *push*

(The unstressed syllable /pa/ is more than 50% louder than the stressed counterpart /pA/. The duration of the stressed syllable /pA/ is approximately 10% longer than the unstressed counterpart /pa/. (cf. the fundamental frequency of /pA/ is, however, higher than that of /pa/.))

At this point we have not found any direct relationship between stress and loudness. It seems, however, that between certain vowels there are intrinsic differences in loudness. [a] is normally higher than [i] in loudness scale, but when [i] is stressed, it is heard a little louder, sometimes almost as loud as [a] in the same stretch of speech.[1] Other vowels do not show any relevant data in terms of stress. As a matter of fact, there are many cases where an unstressed vowel is even significantly louder than its stressed counterpart (figure 7).

The plots of loudness (figure 4) and sound level or intensity (figure 2) are, however, quite useful as a guide for the segmentation of Penrhyn utterances into syllables. Both types of data can be used, but the loudness plot usually has a better structure, clearer appearance and is easier to work with. Very often the nucleus of a syllable coincides with a maximum of loudness and the syllable boundaries with minima. The reason is that vowels tend to be louder than consonants (with the exception of nasals). This is particularly true for voiceless stop consonants, which consist of a short period of silence, followed by a sharp rise in sound level (often 30 dB in 10 ms). Since many Penrhyn syllables begin with a voiceless consonant,[2] a large share of all syllable boundaries can be detected by visual inspection of the loudness diagram; figure 9 shows an example where most syllable boundaries coincide with loudness minima. The final segmentation has to be done by careful listening to sections of the utterance. This time-consuming task is much easier if good candidates for syllable boundaries are already known.

3.2 Syllable duration

The duration of syllables has often been cited to have a high correlation to stress. In many languages, including English, an increase in duration of a given syllable gives an indication that the syllable is more prominent in a word or utterance (Lehiste 1970).

The measurement of syllable durations is not always an easy task. The difficulty is in the determination of the exact location of the syllable boundaries. If a syllable is enclosed within stop consonants the ends of the stops can be taken as very distinct syllable boundaries (see figure 7). The duration of the syllable can then be determined with an uncertainty of less than 10 ms. Examples are:

/papAki/ *push*: [pa] and [pA]
/kOpe/ *hurry* and /koopUU/ *guts*: [kO] and [koo]

In other cases syllable boundaries are not marked precisely by any acoustic parameter and cannot easily be determined by careful listening either. In such cases the duration of a syllable can be uncertain by more than 100 ms. Sometimes a careful examination of the waveform plots (figure 6) can help in finding syllable boundaries.

We have not found any significant durational differences between stressed and unstressed syllables. At most, a stressed syllable may be perhaps 10% longer than its unstressed counterpart. On the other hand, there are cases where an unstressed syllable is even longer than a stressed one. There are, however, great differences between long and short syllables. A long syllable is often two or three times longer than a short syllable. There are no cases in which the duration of a long syllable is less than that of a short syllable. Even an unstressed long syllable is always significantly longer (1.5 to 2 times) than a stressed short syllable. This result is consistent with the fact that long vowels are phonemic in Penrhyn; an increase in duration gives a phonemic difference rather than a stressed impression.

Therefore, we conclude that an increase in duration may be coincidental to stress, but has no direct relationship to it in Penrhyn.

3.3 Fundamental frequency

A typical and idealised pattern of our informant's fundamental frequency movement in a neutral declarative sentence is shown in figure 8. This general pattern of fundamental frequency is more distinct in a longer utterance (e.g. 5-sec-utterance) than in a shorter one (e.g. 1-sec-utterance). This is a slightly modified pattern of what Vaissière (1983) calls a language independent fundamental frequency contour.

The utterance usually begins with an intermediate frequency (about 150 Hz - 200 Hz) and after a few syllables it reaches the highest frequency (about 250 Hz - 300 Hz). From a communications point of view this appears plausible because it lets the hearer find out quickly which frequency range and tempo the speaker is going to use for the following utterance. The fundamental frequency rises and falls almost periodically as a function of time, but with decreasing amplitude and an overall falling tendency. It reaches the lowest frequency when it ceases voicing near the end of the utterance.

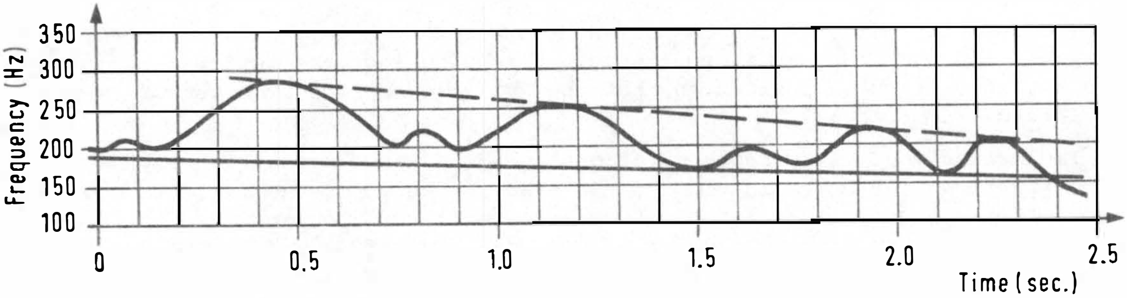

Figure 8: Typical modulation of fundamental frequency in a Penrhyn declarative sentence, forming major waves (touching the upper broken line) and minor waves

(The number of waves is not fixed.)

A very strong relationship exists between perceived stress and the modulations of fundamental frequency. Especially whenever a maximum exceeds the upper broken line, one perceives phonetic prominence very clearly. Usually the stress perceived as 'strongest' in an utterance coincides with such a high peak of the fundamental frequency contour. Besides, along each wave one perceives a stress, too. The closer the peak comes to the upper broken line in figure 8, the more prominent the corresponding syllable appears to be.

In both cases, however, it is not necessarily the syllable which coincides exactly with the peak of the wave but the syllable with the greatest change in fundamental frequency that is perceived as more prominent than others; when the fundamental frequency rises or first rises and then falls immediately or before it falls rapidly, one perceives a phonetic prominence on the corresponding syllable. On the other hand, in an unstressed syllable the fundamental frequency stays near the base line or in the valley of a wave.

From these facts we conclude that stress in Penrhyn is closely related to changes in fundamental frequency. It is not the absolute value of the frequency, but rather its relative value within an utterance that is perceived as stress.

4. STRESS PATTERNS

In general we must take the following points into consideration when we analyse the data:

(a) Stress is an indication of a physical effort of a speaker to make a particular part of an utterance more prominent than others.

(b) Like other suprasegmental features, stress alone does not convey any linguistic information. Stress must always be considered together with segmental and other suprasegmental features.

(c) Stress in carefully spoken slow speech often behaves differently from stress in normal fast speech. Therefore these two styles of speech should be distinguished when dealing with stress patterns (see section 4.2).

(d) A one-word utterance must be considered as an independent short utterance. The stress pattern within such an utterance does not necessarily correspond to 'word stress' but to stress in a short, perhaps carefully spoken utterance.

(e) There are many different types of stress in each language: an intrinsic stress, which all speakers of the language use as part of a natural intonation, a pay-attention stress, which a speaker uses to emphasise a particular information, a stress that expresses emotions, and many others. Although these different types of stress belong to different categories of language analysis, they are encoded using the same set of acoustic parameters.

(f) There is no 'word stress' in Penrhyn that gives a minimal pair such as *pErmit* and *permIt* in English.

(g) In Penrhyn, an intrinsic sentence stress distinguishes a declarative sentence from an interrogative sentence and perhaps from an imperative sentence too. In this paper we only deal with declarative sentences. A superficial look at the data suggests, however, that the stress pattern of a Penrhyn declarative sentence is similar, if not identical to the pattern of interrogative and imperative sentences containing interrogative morphemes and imperative morphemes respectively. Only yes/no questions (interrogative sentences without an interrogative morpheme) differ in stress patterns. However, this has not been tested because of lack of data.

4.1 Syllables

Stress is primarily an attribute of syllables. However, what we mean by 'syllable' is not easy to define. There is no exact universal definition of syllables, although it is a basic building block of many languages.

It has been pointed out in section 3 that syllable boundaries are not marked precisely by any single acoustic parameter. When we use all the parameters (figure 2 to figure 6) and when we listen to the speech data carefully, we may be able to find out a great amount of phonetic change that takes place in a stretch of utterance. However, it is very difficult to segment it, because it is always continuous. It is also extremely difficult to distinguish between consonants and vowels; a certain quality of vowels depends upon the preceding and the following consonant and vice versa.

On the other hand we often discover some significant changes that occur in a number of acoustic parameters. When we use these data, we can determine a minimum phonetic unit and we call it 'syllable'. We define the Penrhyn syllable structure as follows:

> Penrhyn syllables are either long or short. A short syllable has the form [C]V; a long syllable has the form [C]VV. Syllable boundaries occur at every possible breath pause, before every consonant and after every second vowel in a sequence of vowels. In Penrhyn there are no phonemic consonant clusters.

A sequence of two identical vowels does not take any special position in the syllable structure of Penrhyn, because it behaves just like a sequence of two different vowels:

/pooro/	*ball*	[pOoro] or [poOro]
/puaka/	*pig*	[pUaka] or [puAka]

The stress on the first vowel or the second varies freely from speaker to speaker and from one occasion to another. In slow careful speech the stress occurs more often on the second vowel in both words, and in normal fast speech the stress is likely to occur on the first vowel in both cases.

4.2 Domains of stress

As we mentioned in section 3.3, Penrhyn stress is almost exclusively related to changes in fundamental frequency. In this section we investigate if these changes of fundamental frequency show systematic stress patterns in several domains: sentence, phrase and a certain smaller domain which we call measure.

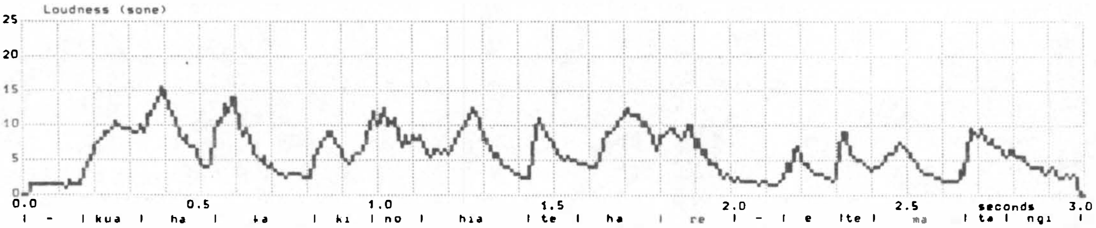

Figure 9: Syllable boundaries and minima in loudness

(#kua haka-kIno-hia te hAre e te matAngi#
The house was damaged by the wind.

The morphologically expected syllables are:
kua ha ka ki no hia te ha re e te ma ta ngi

Most of these syllables are clearly separated by a minimum in loudness.)

4.2.1 Sentence

When we examine the changes of fundamental frequency, we notice that there is often a deviation from the standard contour; a maximum exceeds the upper broken line of figure 8. Here one perceives a very strong stress which usually indicates the most important information of a given sentence. So, we call it 'sentence stress' (figure 11).

The position of a sentence stress is normally semantically determined (see figure 11):

Kaa hano au ki te kaaInga, kai-kai ai.
I am going HOME to have a meal there.

It is interesting to note, however, that a sentence stress occurs very frequently towards the end of an utterance. Because of the structure of Penrhyn sentences, a semantically important morpheme is often placed at the end of a sentence. Also by raising the fundamental frequency at the end, Penrhyn speakers seem to indicate that the utterance is ending. We call this 'sentence-final stress' (see figure 10). A sentence-final stress is usually accompanied by devoicing of the final vowels and decreasing loudness. This may suggest a typical Penrhyn declarative sentence structure, because this 'sentence-final stress' is perceived as 'stress' by a hearer who does not know the language, while it is perceived as 'a natural intonation' by a hearer who speaks or knows the language well.

Some sentences contain no sentence stress (the prosodically neutral sentences), others have one or more, depending upon which information is emphasised by the speaker. The degree of emphasis does not depend on the absolute frequency value but rather on the relative amount of the deviation: how far the peak exceeds the upper broken line or how much is the difference between the maximum and the preceding or the following minimum.

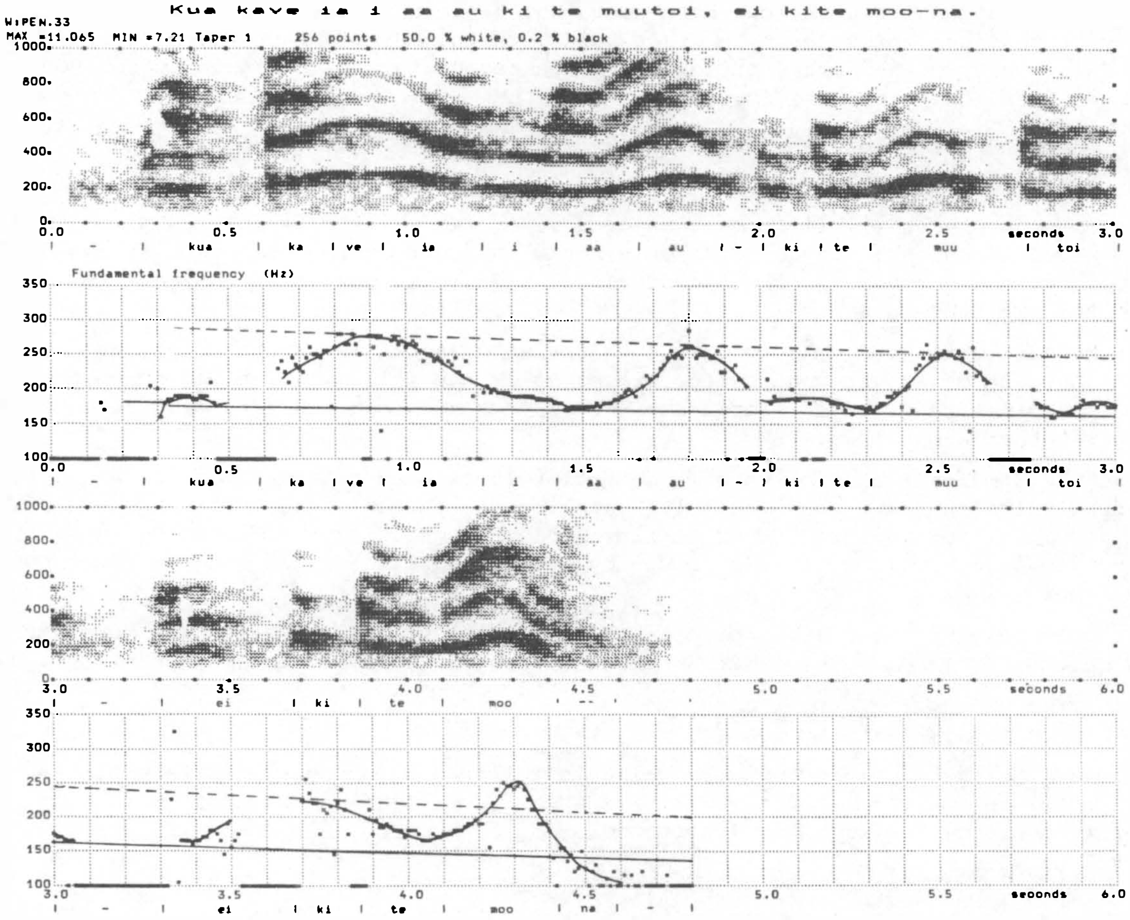

Figure 10: Fundamental frequency of a Penrhyn declarative sentence

(A sentence-final stress occurs at the morpheme mOO-na *for him* where a maximum exceeds the upper broken line, and phrase stress occurs along the major waves, which contain lexical morphemes.)

#kua kAve ia i aa AU ki te mUUtoi ei kIte mOO-na#
take I police appear
He took me to the police as his witness (to appear FOR HIM).

4.2.2 Phrase

Besides sentence stress we further perceive phonetic prominence along the waves (see figure 10).

There are essentially two types of waves: waves whose peaks reach or are close to the upper broken line and waves whose peaks are far away from it. The former we call 'major waves' and the latter 'minor waves'. When we examine the domains of these waves, each major wave usually contains one lexical morpheme and a few optional grammatical morphemes, while a minor wave mostly consists of only grammatical morphemes. In other words, the number of major waves corresponds roughly to the number of lexical morphemes contained in a given utterance with the following exceptions:

(a) A sequence of lexical morphemes such as a noun and an adjective:

moni fiitii *Fijian money*

In slow speech fiitii forms a major wave or two while moni forms a minor wave. In such a case a major wave seems to occur on the semantically more important morpheme in that particular utterance. In fast speech they form together one major wave.

(b) Monosyllabic personal pronouns such as ia *he* and au *I*. In fast speech they are often attached to the preceding or the following lexical morpheme to form a major wave. For example kave *take* and ia *he* in figure 10 form together a major wave.

The domain of a major wave is most closely related to the domain of a morphological phrase, so we call it 'phonological phrase', and the stress of this domain 'phrase stress'. A similar notion is called 'prosodic word', 'phonological word' or 'stress group' by Cutler and Ladd (1983) and 'Akzentgruppe' by Zinglé (1982).

A phrase stress occurs normally on the first syllable of a lexical morpheme, or on the final syllable, if it is the only long syllable in the morpheme.

//ka hInangaro//	*will want*
//te pahII//	*the ship*

However, before a phrase boundary which is followed by a pause, a phrase stress can occur on the final short vowel. In such cases the final short vowel is never devoiced, but slightly lengthened.

//te tangatA//	*the man*
//te tamaiti inA//	*the child there*

4.2.3 Measure

Besides the stress along the major wave, we also perceive stress along the minor wave, particularly in carefully spoken slow speech.

The domain of each minor wave corresponds most closely to the notion 'measure' introduced by Scott (1948) for Fijian 'words' and expanded by Schütz as a phonological unit for Fijian (1976), Hawaiian (1978, 1981) and Maori (1984).

The minimum unit of the domain of this minor wave in Penrhyn is one long syllable. The longest wave includes a lexical morpheme. While one lexical morpheme sometimes extends over two waves, a few grammatical morphemes can also form one wave (figure 11). Each wave has one stressed syllable. In other words, this is the minimal stress domain, and we also call it 'measure'. It is a phonological notion, which is placed between phonological phrases and syllables.

¦rAA ¦ UA ¦	*they* (dual)
¦pOO ¦ pOngi ¦	*morning*
¦kI te ¦	*to the*
¦ei ¦ kIte ¦ mOOna ¦	*to appear for him*

In careful slow speech, like in a one-word utterance, a measure stress occurs on the penultimate syllable or on the final long syllable.

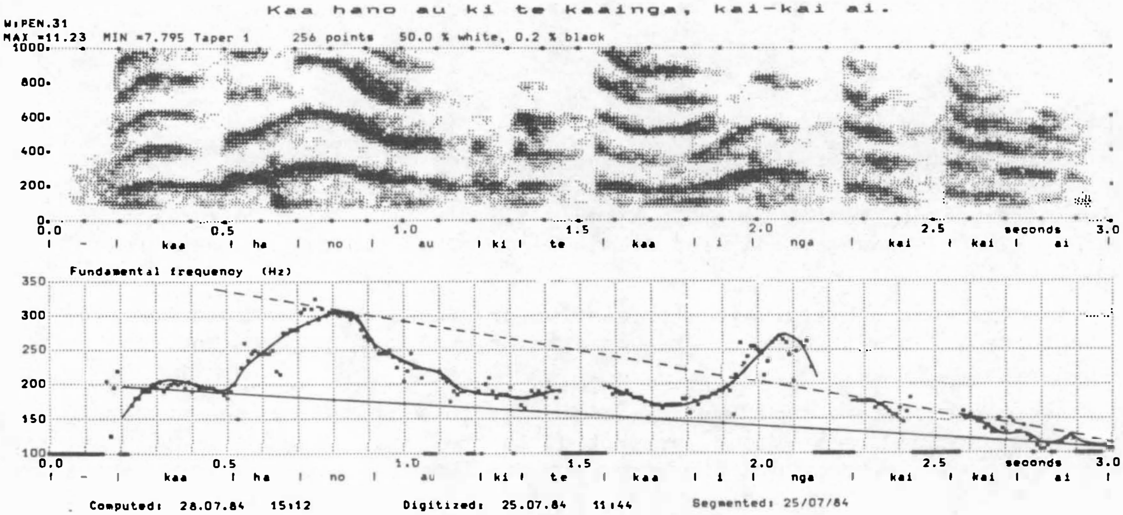

Figure 11: A Penrhyn declarative sentence, divided into measures

(Each measure corresponds to one wave.)

¦kaa ¦ hano ¦ au ¦ ki te ¦ kaa ¦ inga ¦ kai ¦ kai ¦ ai ¦

#kaa hAno au ki te kaaInga KAI-kai ai#
go I home eat
I am going HOME to have a meal there.

(Phrase stress is marked by capital letters. A sentence stress occurs on kaaInga *home*, where the fundamental frequency exceeds the upper broken line.)

We have not found any successful rules yet as to when a grammatical morpheme is attached to a preceding or a following lexical morpheme or when two or three grammatical morphemes can be combined to form a measure.

4.3 Interaction of stress in different levels

It is not yet clear how measure stress, phrase stress and sentence stress influence each other in an utterance. It seems that in carefully spoken slow sentences one perceives a measure stress throughout the utterance.

In normal fast speech, however, some measures are attached to a preceding or a following measure and a regrouping occurs. The new group is likely to act like a phrase and accordingly gets a phrase stress. Therefore a potential measure is suppressed (see kave ia and i aa au in figure 10). Both, phrase stress and measure stress, are possible candidates for receiving sentence stress, depending upon how important the corresponding morphemes are in a given sentence.

This is why 'words' based on a morphological analysis sometimes receive stress on the penultimate syllable and sometimes on the first syllable. It is now very clear that there is no 'intrinsic word stress' in Penrhyn. A stress in a word (= lexical morpheme) can be shifted from one syllable to another according to its position within a measure, a phrase or a sentence, and depending upon the style or the tempo of the speech.

4.4 Function of stress

One perceives a certain rhythm along the waves of fundamental frequency, which appears to be a physiological necessity for language production as well as for perception. People become extremely tired of listening to synthetic speech, if suprasegmental features are lacking. This indicates that stress simplifies the understanding of speech.

It seems that a hierarchical tree structure based on the stress pattern can be formed. This tree structure consists of measures, phrases and sentences in phonological terms.

The tree structure is in principle similar to a syntactic tree structure of for instance a Chomsky-model, but not quite the same. Suprasegmental features are superimposed in the syntactic structure and function when they are necessary. For instance, where a syntactic ambiguity occurs, a phonological phrase boundary helps a speaker and a hearer decide on the right choice.

5. NOTATIONS

Illustrative texts are cited in phonemic transcription, modified in the following respects:

: sentence boundary
// : phrase boundary
¦ : measure boundary
/ / : phonemic transcription
[ ] : phonetic transcription
upper case : stressed syllable
hyphen : morphologically determined complex word

NOTES

1. See section 5 for details of the notation.
2. Penrhyn has 10 consonant phonemes: p t k m n ng r v s h, and five vowel phonemes: i e a o u.

REFERENCES

BIGGS, B.G.

1961 The structure of New Zealand Maori. *Anthropological Linguistics* 3:1-54.

CUTLER, A. and D.R. LADD, eds

1983 *Prosody models and measurements*. Berlin: Springer.

DELMONTE, R.

1982 An automatic text-to-speech prosodic translator for the synthesis of Italian. *Fortschritte der Akustik-DAGA* '82:1021-1026.

FOLKINS, J.W., C.J. MILLER and F.D. MINIFIE

1975 Rhythm and syllable timing in phrase level stress patterning. *Journal of Speech and Hearing Research* 18:739-753.

HIROSE, K. and H. FUJISAKI

1983 Analysis and synthesis of voice fundamental frequency contours of complex sentences. *Proceedings of the 11th International Congress of Acoustics* 4.

HOHEPA, P.

1967 A profile-generative grammar of Maori. Ph.D. dissertation, Indiana University, Bloomington.

KOENIG, W., H.K. DUNN and L.Y. LACY

1946 The sound spectrograph. *The Journal of the Acoustical Society of America* 17:19-49.

LEHISTE, I.

1970 *Suprasegmentals*. Cambridge, Mass.: M.I.T. Press.

MARKEL, J.D. and A. GRAY

1976 *Linear prediction of speech*. Berlin: Springer.

NEWBRAND, H.L.

1951 A phonemic analysis in Hawaiian. M.A. thesis, University of Hawaii.

PAULUS, E. and E. ZWICKER

1972 Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln. *Acustica* 27:253-266.

SCHÜTZ, A.J.

1976 Fijian prosody I: syllables and groups. *Working Papers in Linguistics, University of Hawaii* 8/2:75-100.

1978 Accent in two Oceanic languages. *Anthropological Linguistics* 20/4: 141-149.

1981 A reanalysis of the Hawaiian vowel system. *Oceanic Linguistics* 20/1: 1-44.

1984 Accent and accent units in Maori: the evidence from English borrowings. *Journal of the Polynesian Society* (in press).

SCOTT, N.C.

1948 A study in the phonetics of Fijian. *Bulletin of the School of Oriental and African Studies* 12:737-752.

VAISSIÈRE, J.

1983 Language-independent prosodic features. In A. Cutler and D.R. Ladd, eds *Prosody models and measurements*, 53-66. Berlin: Springer.

WOLF, H.E.

1982 Intonationsmotive zur Prosodiesteuerung nach Regeln für die Sprachsynthese. *Fortschritte der Akustik-DAGA* '82:1031-1034.

YASUDA, A.

1968 The structure of the Penrhyn phrase. M.A. thesis, University of Hawaii.

ZINGLÉ, H.

1982 Automatische Prosodiesteuerung in der Synthese der deutschen Sprache. *Fortschritte der Akustik-DAGA* '82:1027-1030.

ZWICKER, E.

1982 *Psychoakustik*. Heidelberg: Springer.

ZWICKER, E. and R. FELDTKELLER

1967 *Das Ohr als Nachrichtenempfänger*. Stuttgart: S. Hirzel.

Yasuda-Graefe, A. and Graefe, V. "A computational approach to stress patterns in Penrhyn". In Geraghty, P., Carrington, L. and Wurm, S.A. editors, *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*.
C-93:357-376. Pacific Linguistics, The Australian National University, 1986. DOI:*10.15144/PL-C93.357*