Acoustic correlates of stress and their use in diagnosing syllable fusion in Tongan

Marc Garellek and Jamie White
marcgarellek@ucla.edu; jameswhite@ucla.edu

Abstract

The goals of this study were to determine the acoustic correlates of primary and secondary stress in Tongan, and to use these correlates in diagnosing syllable fusion, an alleged phonological process in the language by which sequences of vowels can fuse into a single syllable. Using recordings of one female native speaker, we found that pitch, duration, and vowel quality appear to be strong cues for primary stress, but intensity and voice quality can differentiate stressed from unstressed tokens for certain vowels. For secondary stress, only F0 was found to differentiate stressed from unstressed vowels, but the effect was smaller than for primary stress. Using these correlates of stress, we found evidence for syllable fusion in Tongan based on pitch and voice quality contours as well as differences in vowel height.*

1 Introduction

Cross-linguistics studies of the acoustic correlates of stress have shown that there are multiple cues to stressed vowels, and that these cues may differ across languages. Typically, stressed vowels have a higher fundamental frequency or pitch (Lieberman 1960; Adisasmito-Smith & Cohn 1996; Gordon & Applebaum 2010), greater intensity (Lieberman 1960; Everett 1998; Kochanski et al. 2005; Gordon & Applebaum 2010), longer duration (Lieberman 1960; Everett 1998; Gordon & Applebaum 2010), differences in vowel quality (Cho & Keating 2010), and differences in voice quality or phonation (Sluijter & van Heuven 1996; Campbell & Beckman 1997), although not all of these acoustic features need to correlate with stress for a given language. The correlates of secondary stress may differ from those of primary stress (Adisasmito-Smith & Cohn 1996), or in some languages may be non-existent (Gordon & Applebaum 2010).

There has been little acoustic work on stress in Polynesian languages, and very little for Tongan in particular (but see Anderson and Otsuka 2003, 2006 for pitch and duration as cues to stress). The goal of this paper is to determine what acoustic measures correlate with both primary and secondary stress in Tongan, and to see whether stress can be used as a diagnostic for a phonological process referred to as syllable fusion (Poser 1985). In Tongan it has been claimed that certain vowel sequences may become one syllable, resulting in a diphthong (Churchward 1953; Feldman 1978; Poser 1985; Schütz 2001), but there is disagreement as to whether Tongan has such a process (cf. Taumoefolau 2002). Despite this disagreement, there has been no phonetic study of

* We'd like to thank our consultant Piula Tonga as well as Kie Zuraw, Hilda Koopman, the 2010 UCLA Field Methods class, and the audience of the Tongan-fest mini-conference for their comments and feedback.

Tongan to address the issue, so the second part of this study is designed to fill this gap. If fusion does occur, then we expect that correlates of stress will be present on both vowels of a fused sequence.

In the following section, we will describe the Tongan phoneme inventory, stress patterning, and syllable fusion. In Section 3 we outline the experiment conducted to determine the correlates of primary and secondary stress in Tongan. Section 4 will examine whether syllable fusion can be assessed using the measures from Section 3. We will argue based on these stress correlates that syllable fusion does indeed occur in certain vowel sequences.

2 Background

Tongan is an Austronesian language of the Malayo-Polynesian branch spoken by 126,390 speakers predominantly in the Kingdom of Tonga. Its closest relative is the only other language of the Tongic subfamily, Niuean (Ethnologue 2009).

2.1 Phoneme inventory

Tongan has 11 consonant phonemes, shown in Table 1. The only voicing contrast is in the labial fricatives, where /f/ and /v/ contrast. Tongan has a standard five-vowel system, shown in Table 2. The language has a length distinction for all the vowels (orthographically represented with a macron), though some claim that long vowels are phonologically a sequence of two identical vowels in separate syllables. This claim is motivated by the fact that stress can fall on the second mora. Additonally, to make a nominal phrase definite, the phrase-final vowel is suffixed by an identical vowel, resulting in VV sequence equivalent in duration to monomorphemic “long” vowels (Anderson & Otsuka 2006).

Table 1: Tongan consonant inventory

	Labial	Dento-alveolar	Velar	Laryngeal
Stop	p	t	k	ʔ
Fricative	f v	s		h
Nasal	m	n	ŋ	
Lateral		l		

Table 2: Tongan vowel inventory

	Front	Central	Back
Close	i		u
Mid	e		o
Open		a	

2.2 Stress in Tongan

Primary stress predictably falls on the penultimate mora in Tongan words. If the penultimate mora falls on the second half of a sequence of identical vowels, then a breaking process occurs, whereby the two vowels become perceptually distinct due to the greater prominence of the second vowel. This difference is also represented in Tongan orthography, such that long vowels are typically written with a macron (e.g., *māsima* [ma.a.ˈsi.ma] ‘salt’) whereas broken long vowels are written as two vowels (e.g., *maama* [ma.ˈa.ma] ‘light’).

The nature of secondary stress is much less apparent. Cross-linguistically, secondary stress often has weaker cues than primary stress (Cho & Keating 2010). A number of researches have noted that Tongan may have multiple stresses in a word, at least in certain conditions (Churchward 1953; Feldman 1978; Schütz 2001; Taumoefolau 2002). Some have said that secondary stress occurs in alternating syllables and may be subject to morpheme boundaries (Feldman 1978), whereas others have noticed that secondary stress may not be predictable, especially in loans (Schütz 2001; Zuraw, O’Flynn, & Ward 2010). Nevertheless, in all the words used to examine secondary stress in this study, the secondary stress falls consistently on the leftmost vowel of the word.

2.3 Vowel sequences

Tongan allows every possible sequence of two vowels, although some are more frequent than others. As with identical vowels, sequences of two different vowels are analyzed as being part of separate syllables (Taumoefolau 2002; Anderson & Otsuka 2006). For example, the words *lā* and *lai* would be disyllabic, where [ai] is not a diphthong but a sequence of phonemes.

However, it has been observed that certain vowel sequences, notably those for which the second vowel is more close than the first, may become one syllable, resulting in a diphthong (Churchward 1953; Feldman 1978; Poser 1985; Schütz 2001). Possible target sequences for syllable fusion are said to include {ai, ae, ao, au, ei, eu, oi, ou}. Syllable fusion is said to be blocked when primary stress falls on the second vowel. Thus, syllable fusion is blocked in a word like [ma.ˈi.na] ‘to be gaping (of a wound)’, given that stress is placed on the [i]. This is similar to the breaking of sequences of identical vowels, which occurs when stress falls on the second vowel, as in [ma.ˈa.ma] ‘light’. Stress cues may provide acoustic evidence for syllable fusion. If a sequence of two vowels have fused, then stress correlates should be apparent on both vowels instead of on just one. For example, if *mai* ‘toward’ has a fused vowel sequence, then the /i/ would have some cues to stress, despite the stress being assigned to only /a/. This hypothesis will be tested in Section 4.

3 Experiment 1: Correlates of primary and secondary stress

The goal of Experiment 1 was to determine which acoustic measures correlate with primary and secondary stress. In case these correlates differ, the experiment was divided in to two parts, the first being devoted to measures of primary stress, the latter to measures of secondary stress

3.1 Experiment 1A: Correlates of primary stress

3.1.1 Introduction

As mentioned earlier, the acoustic correlates of stress across-languages may be higher pitch, greater intensity, longer duration, and differences in vowel quality (Gordon & Applebaum 2010), as well as differences in voice quality or phonation (Sluijter & van Heuven 1996; Campbell & Beckman 1997). In this experiment, we sought to determine which of these measures are good correlates of primary stress in Tongan.

3.1.2 Method

The target words were uttered by a female native speaker living in the greater Los Angeles area for approximately 40 years. To evaluate primary stress, ten words were used for each of the five Tongan vowels. The words were of the form CVˈCVCV where the first and second vowels were identical in quality, for example *nenenu* [ne.ˈne.nu] 'to hesitate persistently.' Each word was placed in the carrier phrase [aŋiˈmui ˈʔae foʔiˈlea ˈkoe _____ kiateˈau] 'Repeat the word _____ for me.' Three tokens were collected for each word, yielding a total of 30 tokens for each vowel. The recording was done in the UCLA phonetics laboratory's sound booth using a Shure SM10A head-mounted microphone, whose signal ran through an XAudioBox pre-amplifier and A-D device. The recording was done using PcQuirerX at a sampling rate of 22,000 Hz.

The targets words were labeled in Praat (Boersma & Weenink 2009) for initial and second vowel, which had no stress and primary stress, respectively. The vowel boundaries corresponded to the onset and offset of a clear second formant. The labeled sound files were then run through VoiceSauce (Shue, Keating, & Vicenik 2009) to obtain the acoustic measures. VoiceSauce calculates F0 using the STRAIGHT algorithm (Kawahara, Masuda-Katsuse, & de Cheveigné 1999). VoiceSauce also outputs the duration of the labeled segment. In addition, values for F1, F2, Root Mean Square (RMS) energy, and cepstral peak prominence (CPP) were calculated for every millisecond. The formants were measured using the Snack SoundToolkit (Sjölander 2004). CPP was calculated using the formula from Hillenbrand et al 1994. CPP values are obtained by taking the fast Fourier transform (FFT) of the log magnitude values of a power spectrum, and then measuring the height of peaks in the transformed signal, which correspond to the duration of the F0 cycle. Lower peaks are an indication of breathy voice. CPP was used as the measure of phonation for the analysis because in measuring the peaks above a noise level, it is less sensitive to formant differences. In our case, using a spectral tilt measure like H1-H2, even if corrected for formant values, would run the risk of having the results skewed by the formants of the five different vowels. Although CPP does not measure tilt per se, the breathiness effect found for unstressed vowels by means of spectral tilt in Sluijter and van Heuven 1996 and Campbell and Beckman 1997 should still be borne out in CPP. Lower CPP values are an indication of breathiness, in that they indicate a noisier signal.

3.1.3 Results and discussion

For each of the measures below, the three tokens for each item were averaged together to get an item mean. Paired samples t-tests were then run in SPSS comparing the means for stressed and unstressed vowels of each quality.

Figure 1 shows the mean F0 values for each of the Tongan vowels, separated according to stress. For each vowel quality, stressed vowels have a higher F0 than unstressed vowels. This F0 difference is usually in the range of 45-65 Hz. Paired t-tests reveal that these observed differences are all significant. The relevant statistics are shown in Table 1.

Table 1. Mean and SD of F0 for vowels with primary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Primary stress	No stress	Primary stress	No stress	
/i/	207.09	145.92	6.69	6.72	p < 0.001***
/e/	190.79	138.25	3.91	2.84	p < 0.001***
/a/	182.88	135.41	4.34	3.68	p < 0.001***
/o/	197.24	137.20	3.06	9.21	p < 0.001***
/u/	208.39	144.20	11.26	6.92	p < 0.001***

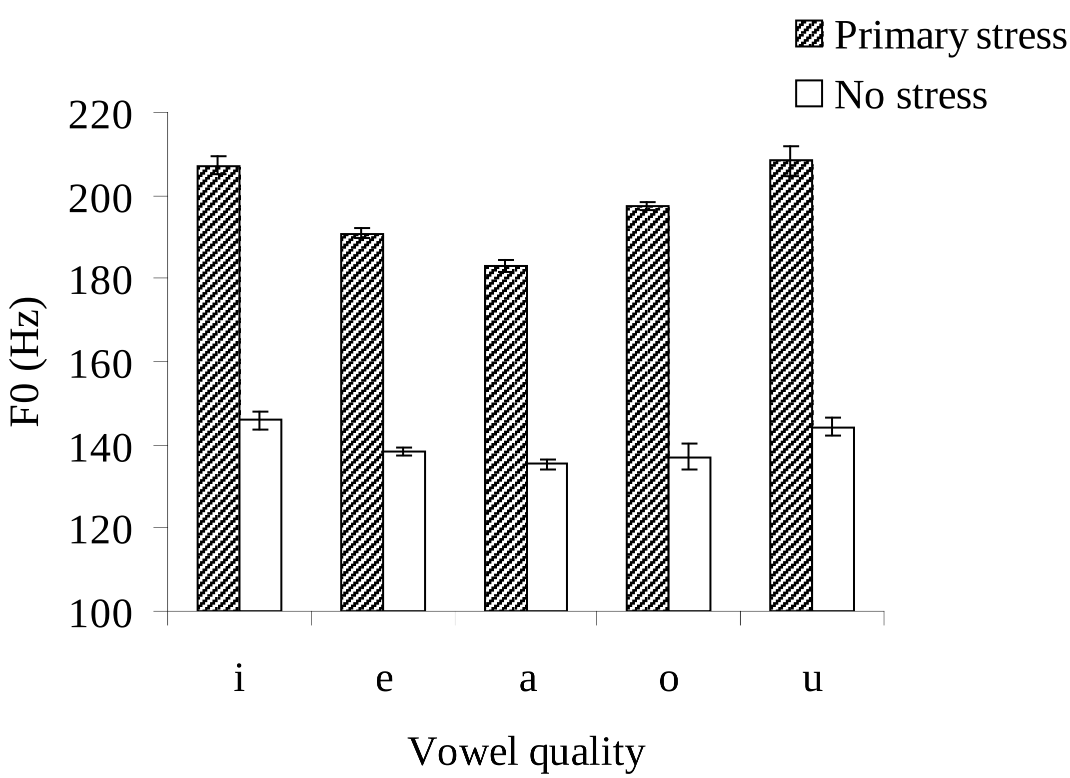

Figure 1. Mean F0 for primary stress and unstressed vowels.

Figure 2 shows the mean duration values for each of the Tongan vowels, separated according to stress. For each vowel quality, stressed vowels have a longer duration than unstressed vowels, by roughly 20-35 ms. Duration was not normalized because both the stressed and unstressed vowels were derived from the same word, and thus the measure was controlled for rate of speech. Paired t-tests reveal that these observed differences are all significant. The relevant statistics are shown in Table 2.

Table 2. Mean and SD of duration for vowels with primary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Primary stress	No stress	Primary stress	No stress	
/i/	101.54	84.20	17.70	11.84	$p < 0.001$***
/e/	105.74	84.56	12.81	11.59	$p < 0.001$***
/a/	93.77	71.90	9.39	6.34	$p = 0.013$*
/o/	105.33	70.75	13.72	9.03	$p < 0.001$***
/u/	100.59	80.95	13.51	13.07	$p < 0.001$***

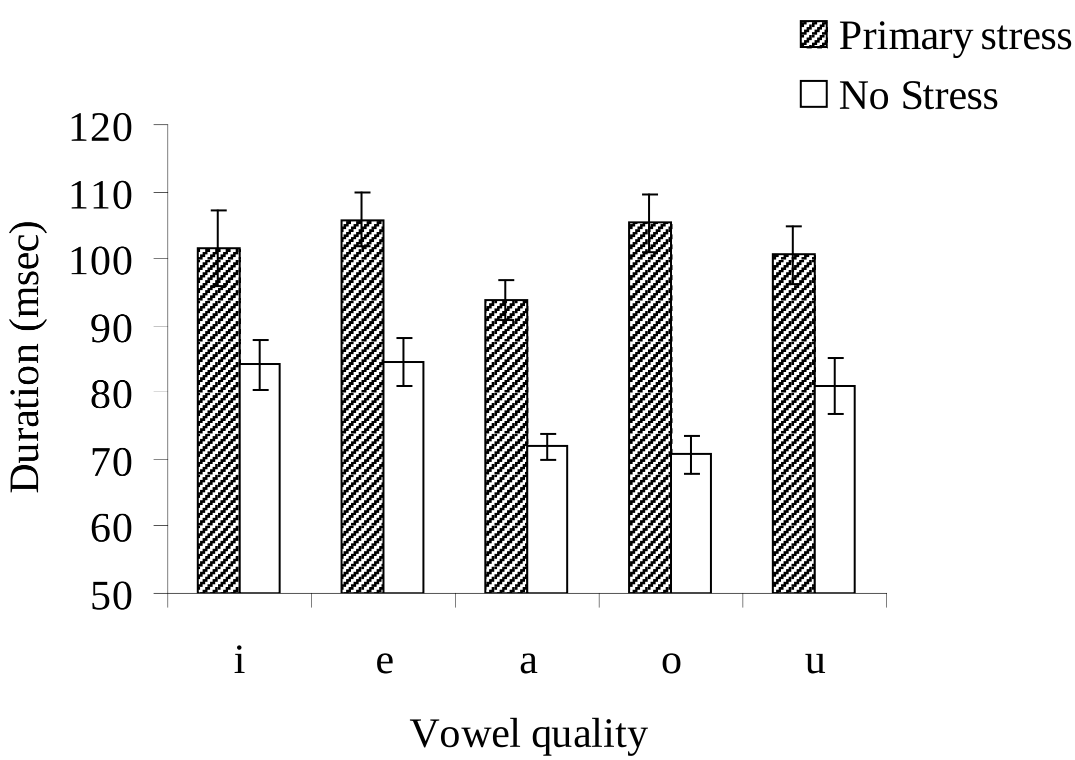

Figure 2. Mean duration for primary stress and unstressed vowels.

Figures 3 and 4 show the results for vowel quality, in terms of vowel height (measured by the first formant F1) and vowel frontness (measured by the second formant or F2) for each of the Tongan vowels, separated according to stress. Figure 3 shows that for each vowel quality, stressed vowels are lower than unstressed vowels, by roughly 50-100 Hz. Paired t-tests reveal that these observed differences are all significant. The relevant statistics are shown in Table 3.

Table 3. Mean and SD of F1 for vowels with primary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Primary stress	No stress	Primary stress	No stress	
/i/	364.89	287.21	23.72	12.63	p = 0.023*
/e/	500.01	409.42	32.90	40.42	p < 0.001***
/a/	875.54	821.58	72.67	47.78	p < 0.001***
/o/	565.54	501.46	33.94	61.92	p = 0.004**
/u/	419.30	310.58	53.71	25.08	p < 0.001***

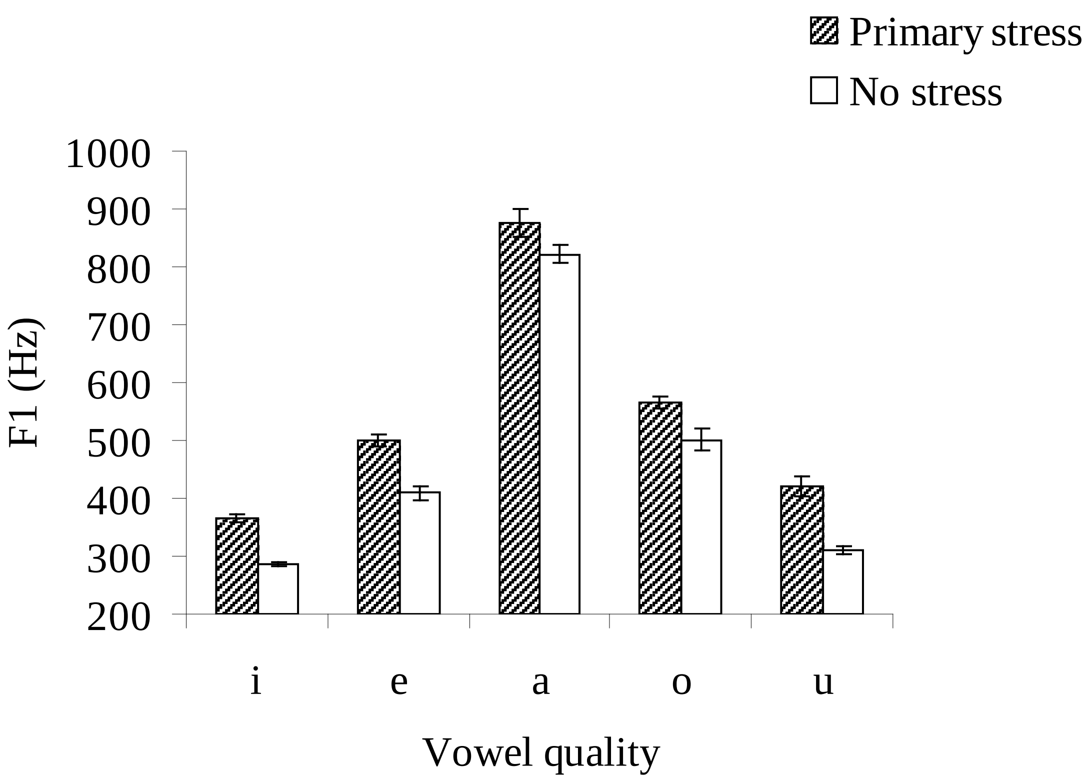

Figure 3. Mean F1 for primary stress and unstressed vowels.

Figure 4 shows that stressed vowels overall do not differ in frontness from unstressed vowels. The one exception is for /a/, which was found to be more front when stressed. The relevant statistics are shown in Table 4.

Table 4. Mean and SD of F2 for vowels with primary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Primary stress	No stress	Primary stress	No stress	
/i/	2448.20	2468.55	95.96	154.85	p = 0.028*
/e/	2228.69	2288.65	80.31	130.87	p = 0.084
/a/	1755.02	1604.15	91.98	173.27	p = 0.667
/o/	1147.28	1200.37	161.15	281.20	p = 0.356
/u/	1132.55	1238.06	249.67	402.85	p = 0.516

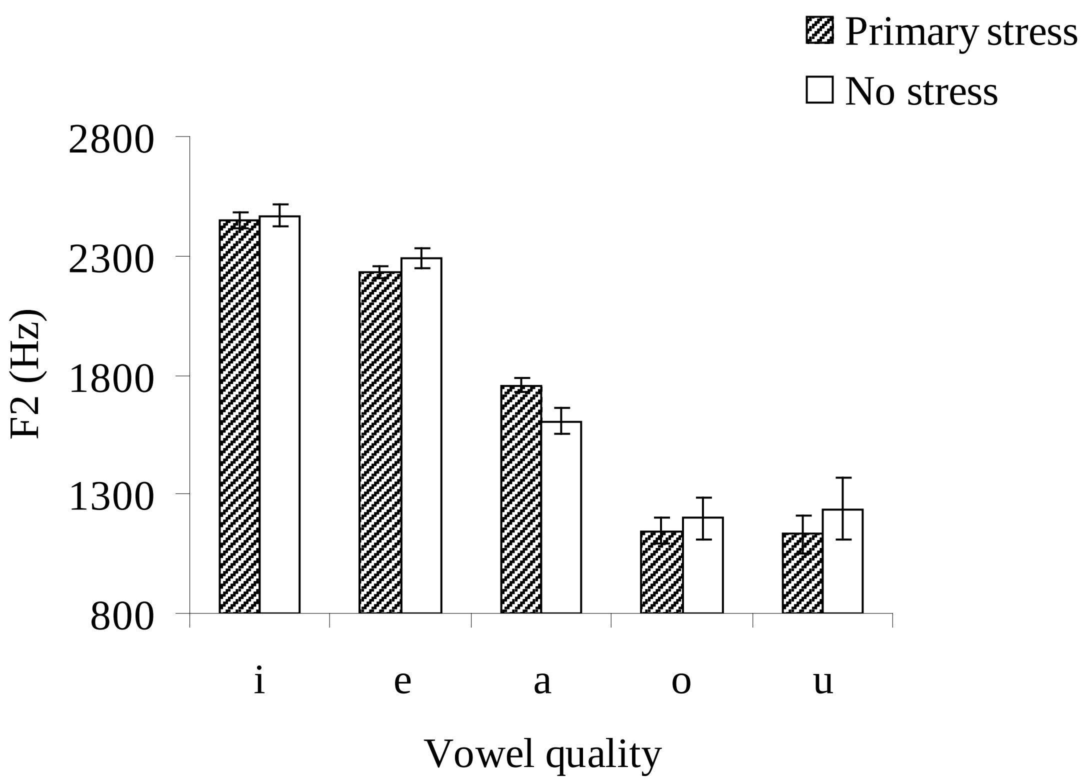

Figure 4. Mean F2 for primary stress and unstressed vowels.

Figure 5 shows that stressed /e, o, u/ have more energy than their unstressed counterparts. The relevant statistics are shown in Table 5.

Table 5. Mean and SD of intensity for vowels with primary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Primary stress	No stress	Primary stress	No stress	
/i/	0.32	0.32	0.08	0.14	$p = 0.16$
/e/	0.95	0.33	0.26	0.09	$p < 0.001$***
/a/	0.69	0.57	0.20	0.25	$p = 0.97$
/o/	1.53	0.55	0.44	0.40	$p < 0.001$***
/u/	0.43	0.23	0.12	0.08	$p = 0.005$**

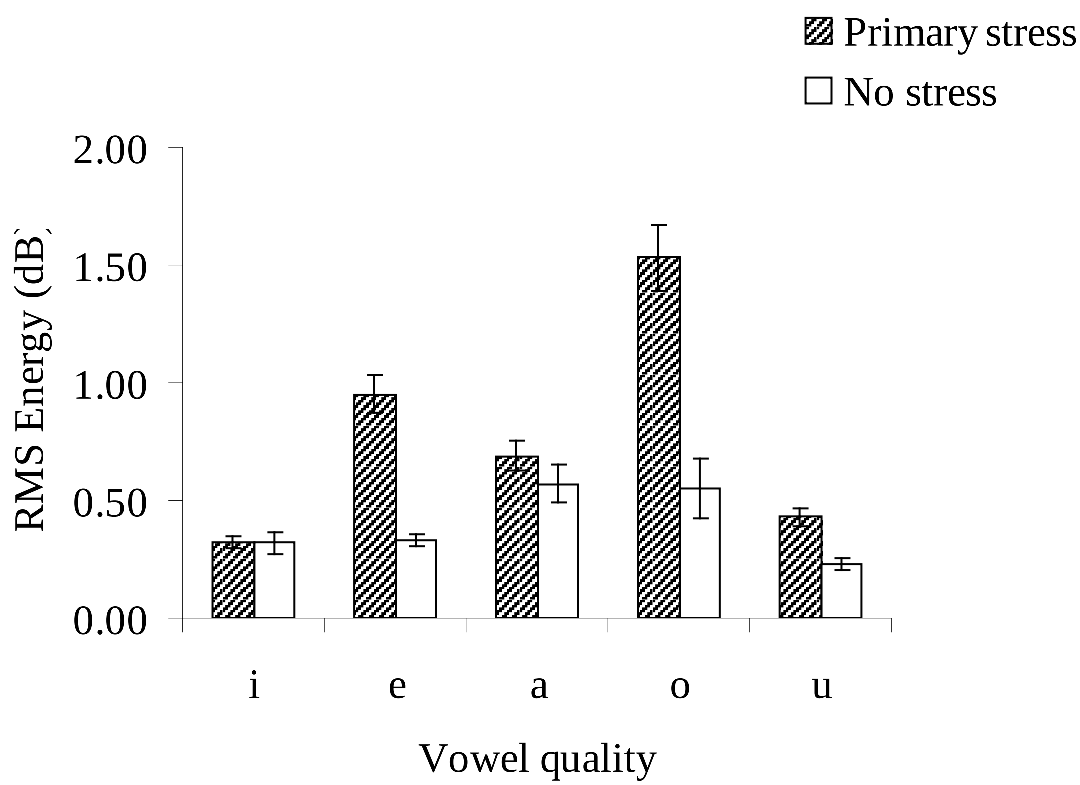

Figure 5. Mean intensity for primary stress and unstressed vowels.

Figure 6 shows that stressed /i, e, a/ have greater CPP than their unstressed counterparts. This means that they are less breathy and less noisy, which is expected. The remaining vowels /o, u/ also have a lower value, but these were not found to be statistically significant. The relevant statistics are shown in Table 6.

Table 6. Mean and SD of CPP for vowels with primary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Primary stress	No stress	Primary stress	No stress	
/i/	24.78	22.14	2.03	2.36	p = 0.011*
/e/	25.86	23.42	1.31	1.80	p = 0.006**
/a/	23.75	21.67	1.45	2.07	p = 0.047*
/o/	23.66	22.89	1.33	2.62	p = 0.448
/u/	22.38	21.11	1.04	2.92	p = 0.134

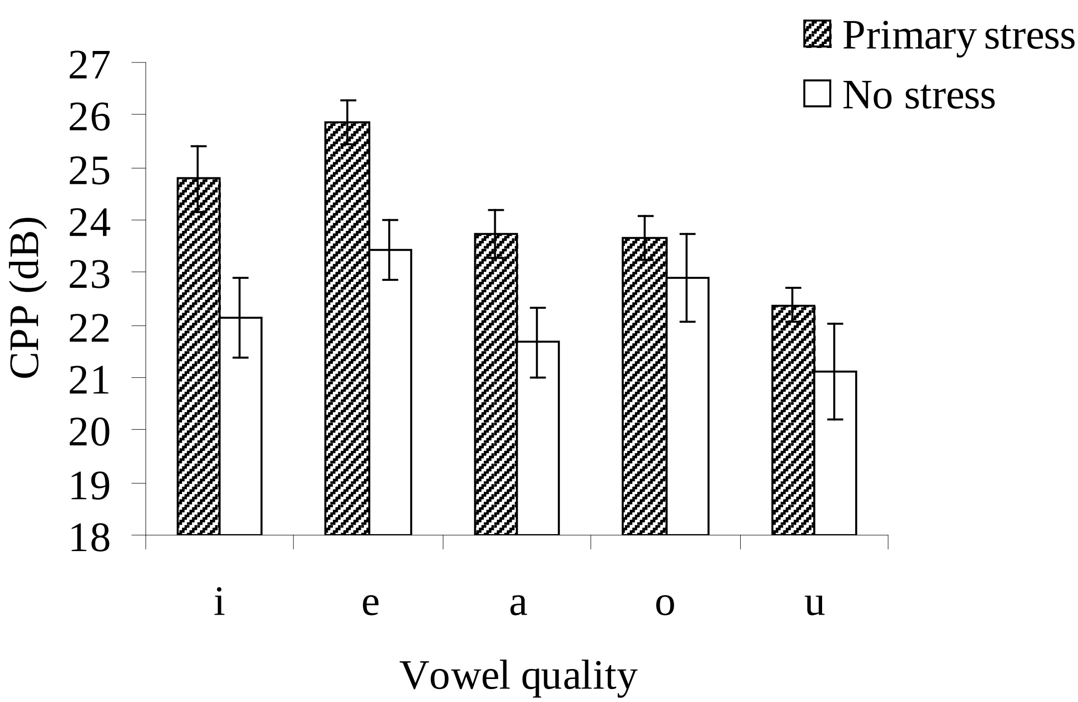

Figure 6. Mean CPP for primary stress and unstressed vowels.

In sum, pitch, duration, and vowel height appear to be the best correlates of primary stress in Tongan. These measures show that all vowels are higher-pitched, longer, and lower when they have primary stress. The other acoustic cues do not consistently show differences across all vowels, but generally confirm that stressed vowels may be lower, more fronted, louder, and less breathy/noisy than unstressed vowels. The correlates of secondary stress will be assessed in the following section.

3.2 Experiment 1B: Correlates of secondary stress

3.2.1 Introduction

Most acoustic studies of stress have focused on primary stress, but studies of secondary stress have shown that the acoustic correlates may differ from those of primary stress (e.g., Adisasmito-Smith & Cohn 1996). In this experiment, we sought to determine which of these measures are good correlates of secondary stress in Tongan.

3.2.2 Method

The target words were uttered by the same speaker as in Section 3.1. To evaluate secondary stress, ten words were used for each of the five Tongan vowels. The words were of the form ˌCVCVˈCVCV, where the first and second vowels were identical in quality, for example *nenenuni* [ˌne.ne.ˈnu.ni] 'hesitate persistently.' These words are identical to those in the Section 3.1, except for the addition of the demonstrative enclitic *–ni*. The testing procedure and location, labeling and retrieval of the acoustic measures were identical to those in the Section 3.1.

3.2.3 Results and discussion

For each of the measures below, the three tokens for each item were averaged together to get an item mean. Paired samples t-tests were then run comparing the means for stressed and unstressed vowels of each quality.

Figure 7 shows the mean F0 values for each of the Tongan vowels, separated according to stress. For each vowel quality except /a/, secondary-stressed vowels have a higher F0 than unstressed vowels. This F0 difference is usually in the range of about 7 Hz, thus much less than the difference between primary stress and no stress, which ranged from about 45-65 Hz. Paired t-tests reveal that these observed differences are all significant. The relevant statistics are shown in Table 7.

Table 7. Mean and SD of F0 for vowels with secondary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Secondary stress	No stress	Secondary stress	No stress	
/i/	149.82	141.37	5.03	4.42	$p < 0.001$***
/e/	147.11	140.65	2.76	1.99	$p < 0.001$***
/a/	138.29	135.46	11.31	3.19	$p = 0.48$
/o/	148.25	143.67	6.12	4.52	$p < 0.001$***
/u/	149.08	141.38	6.93	5.25	$p < 0.001$***

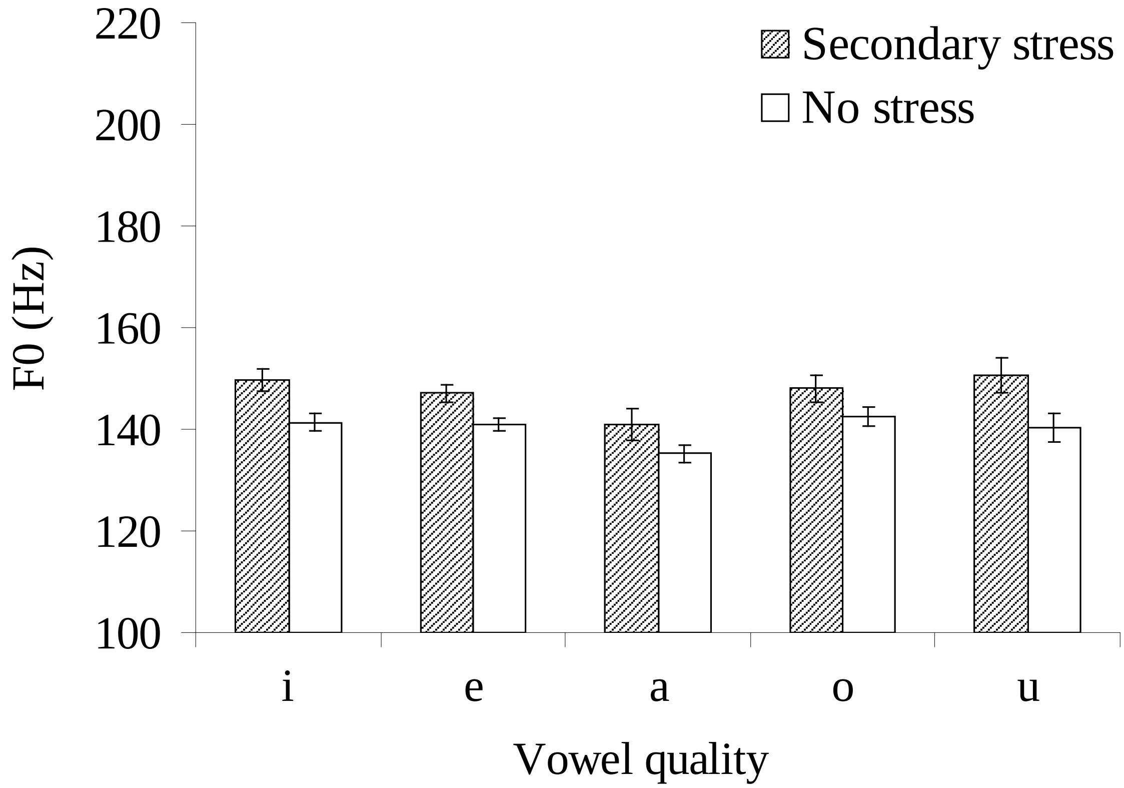

Figure 7. Mean F0 for secondary and unstressed vowels.

Figure 8 shows the mean duration values for each of the Tongan vowels, separated according to stress. For each vowel quality, secondary stressed vowels does not differ in length from unstressed ones. The relevant statistics are shown in Table 2.

Table 8. Mean and SD of duration for vowels with secondary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Secondary stress	No stress	Secondary stress	No stress	Secondary stress
/i/	71.52	71.40	12.23	10.43	0.98
/e/	77.36	75.10	15.78	14.11	0.55
/a/	74.71	73.89	5.65	7.17	0.80
/o/	70.43	72.17	9.59	11.80	0.73
/u/	70.36	70.21	14.24	13.86	0.96

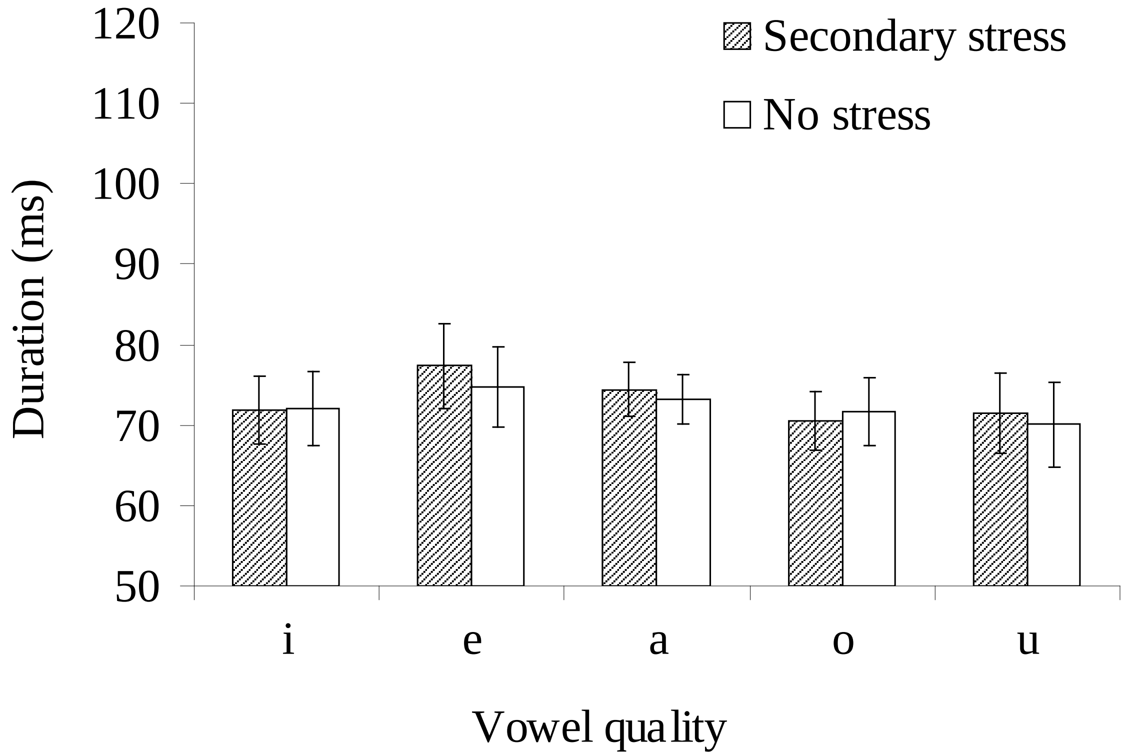

Figure 8. Mean duration for secondary and unstressed vowels.

Figures 9 and 10 show the results for vowel quality, in terms of vowel height (measured by the first formant F1) and vowel frontness (measured by the second formant or F2) for each of the Tongan vowels, separated according to stress. Secondary stressed vowels do not differ in height or retraction from unstressed vowels. The relevant statistics are shown in Tables 9 and 10.

Table 9. Mean and SD of F1 for vowels with secondary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Secondary stress	No stress	Secondary stress	No stress	Secondary stress
/i/	299.63	293.07	13.12	15.53	p = 0.39
/e/	401.55	413.48	34.63	35.92	p = 0.25
/a/	809.55	790.63	65.93	58.60	p = 0.13
/o/	498.95	511.52	45.40	43.92	p = 0.26
/u/	332.10	335.82	15.08	18.17	p = 0.65

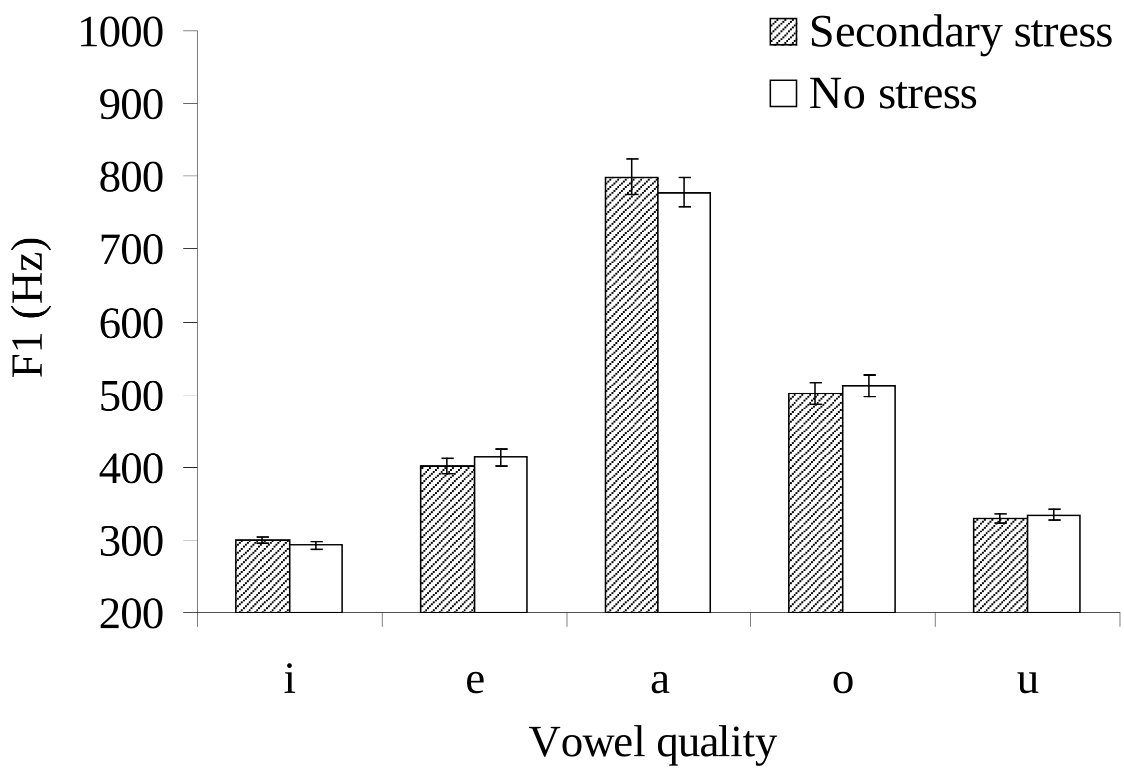

Figure 9. Mean F1 for secondary and unstressed vowels.

Table 10. Mean and SD of F2 for vowels with secondary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Secondary stress	No stress	Secondary stress	No stress	Secondary stress
/i/	2447.81	2491.53	196.34	70.27	p = 0.47
/e/	2258.82	2255.63	132.53	122.41	p = 0.95
/a/	1656.12	1723.90	162.00	98.71	p = 0.33
/o/	1177.09	1163.41	253.14	172.26	p = 0.80
/u/	1135.30	1061.38	413.50	147.68	p = 0.64

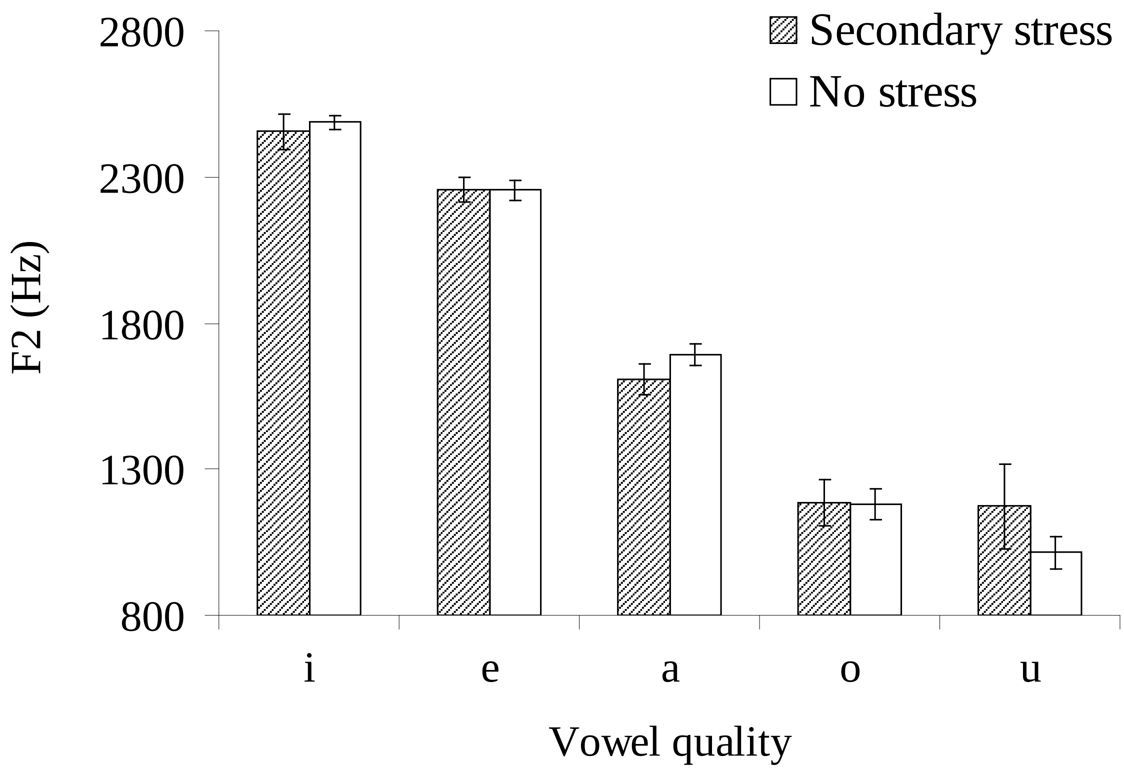

Figure 10. Mean F2 for secondary and unstressed vowels.

Figure 11 shows that secondary stressed /a/ have more energy than its unstressed counterpart. This marginal effect (p = 0.07) is interesting, because primary stressed /a/ does not differ in its intensity from unstressed /a/ and because /a/ is the only vowel not to show a difference in pitch for secondary stress. The vowels that do show a difference in intensity under primary stress, /e, o, u/ here show no differences. The relevant statistics are shown in Table 5.

Table 11. Mean and SD of intensity for vowels with secondary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Secondary stress	No stress	Secondary stress	No stress	Secondary stress
/i/	1.13	1.00	0.69	0.37	p = 0.56
/e/	1.04	1.02	0.37	0.25	p = 0.78
/a/	1.42	0.99	0.92	0.44	p = 0.07
/o/	1.64	1.49	0.65	0.77	p = 0.13
/u/	0.73	0.59	0.42	0.29	p = 0.09

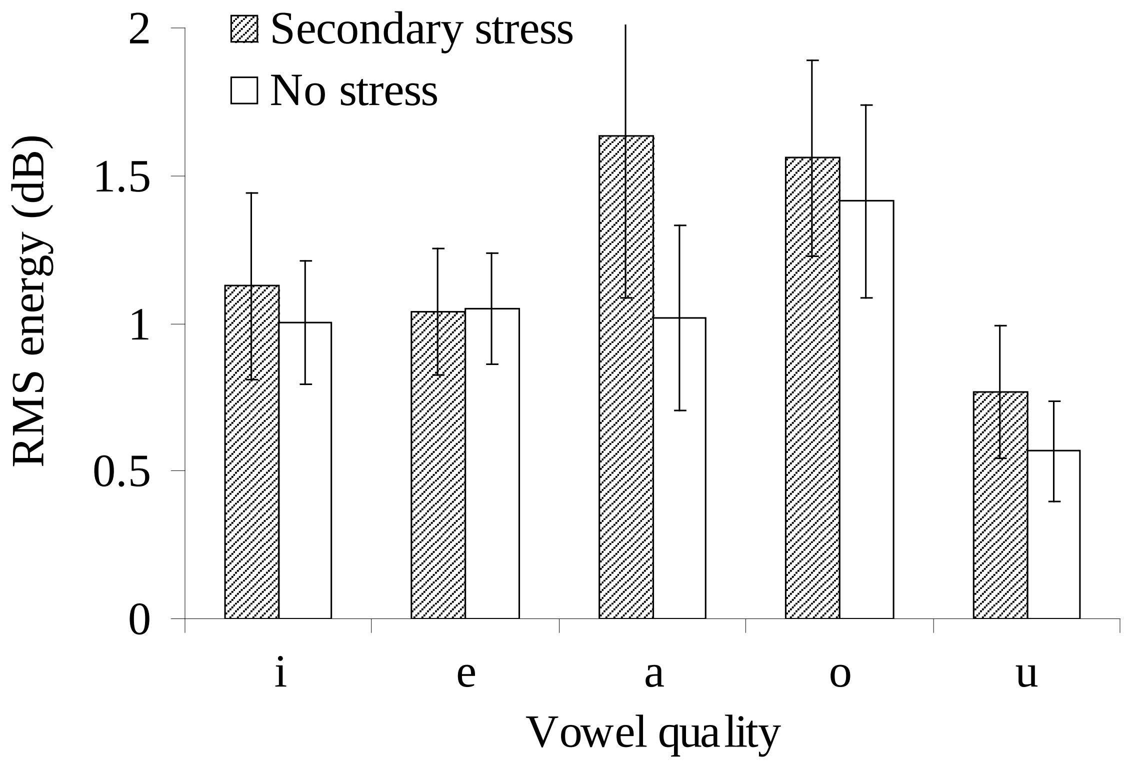

Figure 11. Mean intensity for secondary stress and unstressed vowels.

Figure 12 shows secondary stressed vowels do not differ in voice quality from unstressed vowels, in terms of CPP. The relevant statistics are shown in Table 12.

Table 12. Mean and SD of CPP for vowels with secondary stress and no stress, including p-value of their differences.

	Mean		Standard deviation		P-value
	Secondary stress	No stress	Secondary stress	No stress	Secondary stress
/i/	23.62	25.46	2.39	3.02	p = 0.12
/e/	25.90	26.32	2.18	1.97	p = 0.52
/a/	24.62	24.33	1.26	1.58	p = 0.60
/o/	25.02	24.83	1.80	2.22	p = 0.78
/u/	22.09	23.39	3.27	2.86	p = 0.10

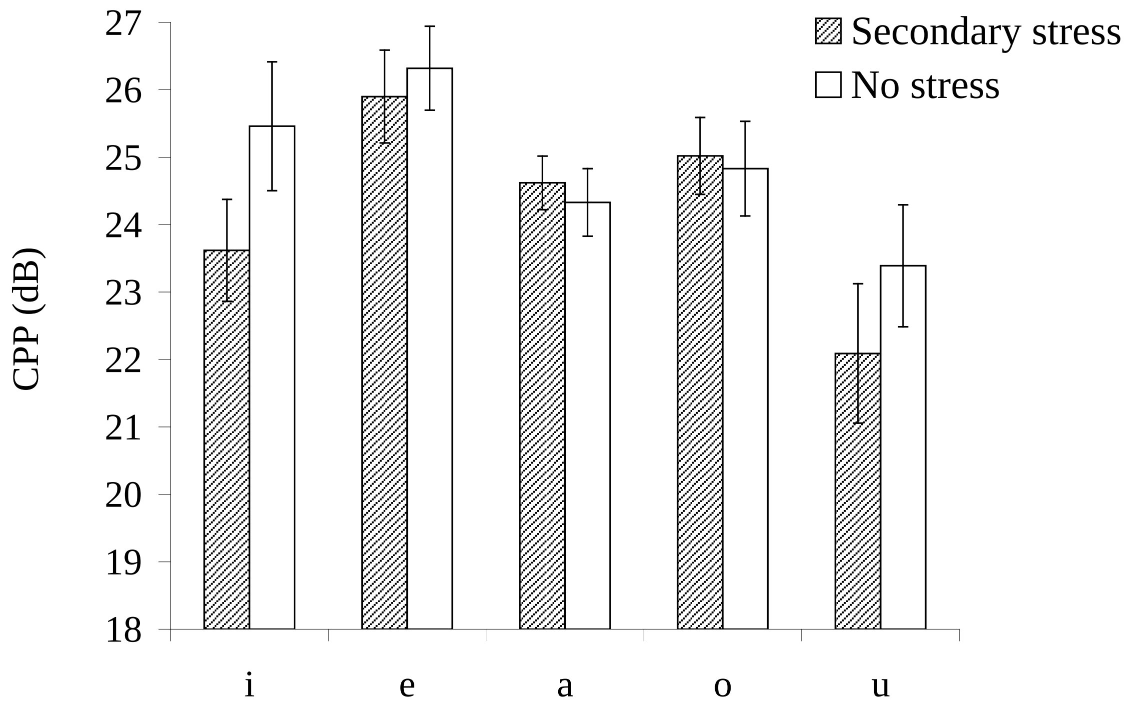

Figure 12. Mean CPP for secondary stress and unstressed vowels.

In sum, pitch appears to be the only consistent correlate of secondary stress in Tongan. However, the mean pitch for secondary stress is higher than that for unstressed vowels by only about 7 Hz, which is substantially less than the difference between primary stressed and unstressed vowels, which ranged from 45-65 Hz.

Table 13 summarizes the significant correlates of pitch found for primary and secondary stress in Experiments 1A and 1B, respectively.

Table 13. Summary of acoustic correlates for found for primary and secondary stress.

	Primary Stress	Secondary Stress
Pitch	Higher	Higher
Duration	Longer	--------
Vowel quality	Lower (higher F1)	--------
Intensity	Louder	--------
Voice quality	More modal (esp. /i, e, a/)	--------

4 Experiment 2: Diagnosing syllable fusion

4.1 Introduction

Syllable fusion is a purported phonological process of Tongan by which certain sequences of vowels, typically analyzed as making up two syllables, may be fused into a single syllable much like a diphthong. Considering the fairly sparse amount of literature dealing with Tongan phonology, the concept of syllable fusion is a topic that arises relatively often. Numerous researchers have commented on the existence of syllable fusion, with disagreement about its exact nature.

In his original grammar of Tongan, Churchward (1953) noted that when the sequences *au, ao, ai, ei, eu, ou* fall in the penultimate position of a word, primary stress may be assigned to the sequence as a whole rather than to only the last vowel; thus, words like *tauhi* 'to keep, look after' may be pronounced as [ˈt͡au͡hi] rather than the expected [taˈuhi].

Feldman (1978) also noticed that when the sequences *ei, ai, ae, ao, au, oi, oe,* or *ou* are adjacent, that they may be pronounced as a single syllable. Poser (1985), who seemingly first used the term "syllable fusion" for this process in Tongan, agrees with Feldman's description as well. However, both Poser and Feldman claim that this process may only occur when stress falls on the first vowel of the sequence and that the sequence is pronounced as two distinct syllables when stress falls on the second vowel. Note that this claim says that the process of syllable fusion is blocked in exactly the case where Churchward claims that one should find it (i.e., words like *tauhi).*

Schütz (2001) generalizes this overall observation by saying that any sequence containing a lower vowel followed by a higher vowel may be a "potential diphthong," meaning that they may form a single syllable.[1] Like Poser and Feldman, however, Schütz notes that this process is blocked in words like *laine* 'line' when primary stress falls on the second vowel of the sequence.

Even though these researchers all seem to agree that syllable fusion exists in Tongan in some form, this fact is not universally accepted. In fact, Taumoefolau (2002), a native speaker of the language, strongly claims that syllable fusion does *not* exist in Tongan, stating "[…] if a diphthong is defined as two unlike vowels within a syllable (instead of just two unlike vowels existing side-by-side and making up two different syllables), then the stress rule of Tongan disallows diphthongs since stress cannot fall on two vowels or two syllables at the same time […] there is no such syllable in Tongan (p. 347)." None of these researchers, however, used phonetic measurements to guide their claims; thus, to our knowledge there have been no acoustic studies conducted to bear on this question.

In Experiment 2, we use acoustic correlates of stress to argue that syllable fusion does happen in Tongan. For those researchers who have claimed that syllable fusion exists, there is a common theme in what types of sequences should be subject to the process. Namely, as Schütz (2001) suggests, sequences of a lower vowel followed by a higher vowel (e.g., *ai*) should be targets for syllable fusion but not sequences of a higher vowel followed by a lower vowel (e.g., *ia*). Based on our personal intuitions working with the language, this generalization seems reasonable. Therefore, for this study we set

[1] Schütz also mentions that *iu* may be eligible to become a diphthong as well, but he was unsure.

up crucial comparisons between lower-to-higher vowel sequences and their higher-to-lower counterparts. In particular, we examined the lower-to-higher sequences *ai, au, ae, ao,* and *ei* and their respective higher-to-lower counterparts *ia, ua, ea, oa,* and, *ie.*[2] For the remainder of this paper, we will refer to the lower-to-higher sequences as *fusing sequences* and the higher-to-lower sequences as *non-fusing sequences* for the sake of clarity.

To demonstrate that syllable fusion has occurred, we must show that fusing sequences behave differently than non-fusing sequences. In the sections to follow, we show this by looking at the stress cues found in Experiments 1A and 1B. We assume that fused sequences (with stress assigned on the first vowel) would have a different syllabic representation than non-fused sequences as follows:

Fused sequences Non-fused sequences

Therefore, fused sequences should show some correlates of stress throughout the entire sequence whereas non-fused sequences should look like sequences of stressed vowels followed by unstressed vowels. That is, the non-fused sequences should have stress cues only on the first part of the sequence. It is important to note that we do not expect fused sequences to act *exactly* like a single stressed vowel (they are different, after all), but they should look markedly different from their corresponding non-fused sequences.

Our precise predictions are as follows. In terms of pitch, we expect that non-fused segments should have a high pitch peak on the first half of the sequence followed by a steep decline while fused sequences should have a flatter contour throughout. Contours for CPP should be the same, such that non-fused sequences are modal followed by a rapid decrease in modal quality while fused sequences maintain a similar voice quality throughout. Similarly, we found that stressed vowels are lower in height than unstressed vowels in Tongan. Therefore, the second part of fused sequences should have a higher F1 like stressed vowels whereas the second part of non-fused sequences should behave like unstressed vowels. Finally, we might expect the overall duration of fused sequences to be shorter than that of non-fused sequences due to the fact that the former is only represented by a single syllable.[3]

[2] We also collected data for *iu* and *ui*, but those data will not be discussed here. Additionally, we did not collected data for lower-to-higher sequences that change backness under the assumption that they might be less likely to fuse. We leave that question for a future project.

[3] We do not make use of intensity in this analysis because of the many confounds related to intensity variations by vowel quality. Moreover, we only expect pitch to differentiate fused and non-fused sequences in secondary stress contexts because only pitch was found to be a significant stress cue for secondary stress in Experiment 1B.

4.2 Method

Target words were uttered by the same speaker as in Experiments 1A and 1B, using the same carrier sentence and recording setup as in those experiments. To look for signs of syllable fusion in sequences where the primary stress falls on the first vowel, we recorded words of the type ˈCVV. Words of the type ˌCVVCVˈVCV, where the first vowel sequence was identical to the second vowel sequence, were also recorded to look at syllable fusion in terms of secondary stress as well as when primary stress falls on the second vowel of a sequence. These longer words were formed by reduplicating the CVV words and adding a *–ni* suffix, which is very productive in Tongan and forms a prosodic word with the preceding word. A full list of the words used is offered in Appendix C.

The words were labeled in three ways using Praat. Each vowel sequence was initially segmented out in its entirety to investigate F0 and CPP contours as well as overall sequence duration. Individual vowels were also labeled so that we could examine individual vowel quality differences. The boundary between the first and second vowel in each sequence was marked at the beginning of formant transitions (the more obvious transition of either F1 or F2 as appropriate for each sequence in question). Finally, the word [foʔiˈlea] was segmented from the carrier phrase; to control for speech rate, durations were calculated as a proportion of this part of the carrier phrase. Labeled sound files were once again run through VoiceSauce to obtain the acoustic measures as in Experiments 1A and 1B.

4.3 Results and discussion

4.3.1 Primary stress

The analysis of primary stress compares fusing sequences (e.g., *ai*) with non-fusing sequences (e.g., *ia*) in ˈCVV words, where the first vowel of the sequence is assigned primary stress. We first examined the pitch contours over the entire length of the vowel sequences. Figure 13 and Figure 14 presents the F0 contours over ten standardized time intervals for the fusing and non-fusing sequences, respectively. Comparing the shapes of the contours, it is clear that the fusing sequences have flatter overall F0 contours than the corresponding non-fusing sequences. This implies that stress is being assigned over the entire sequence rather than only over the first vowel. The non-fusing sequences, on the other hand, show higher pitches over the first part of the sequences with a quick pitch decline, implying that they are behaving more like sequences of stressed vowels followed by unstressed vowels. Moreover, the trendline for the fusing sequences has a lower negative slope than the trendline for the non-fusing ones, further supporting the observation that the fusing sequences have flatter pitch contours.

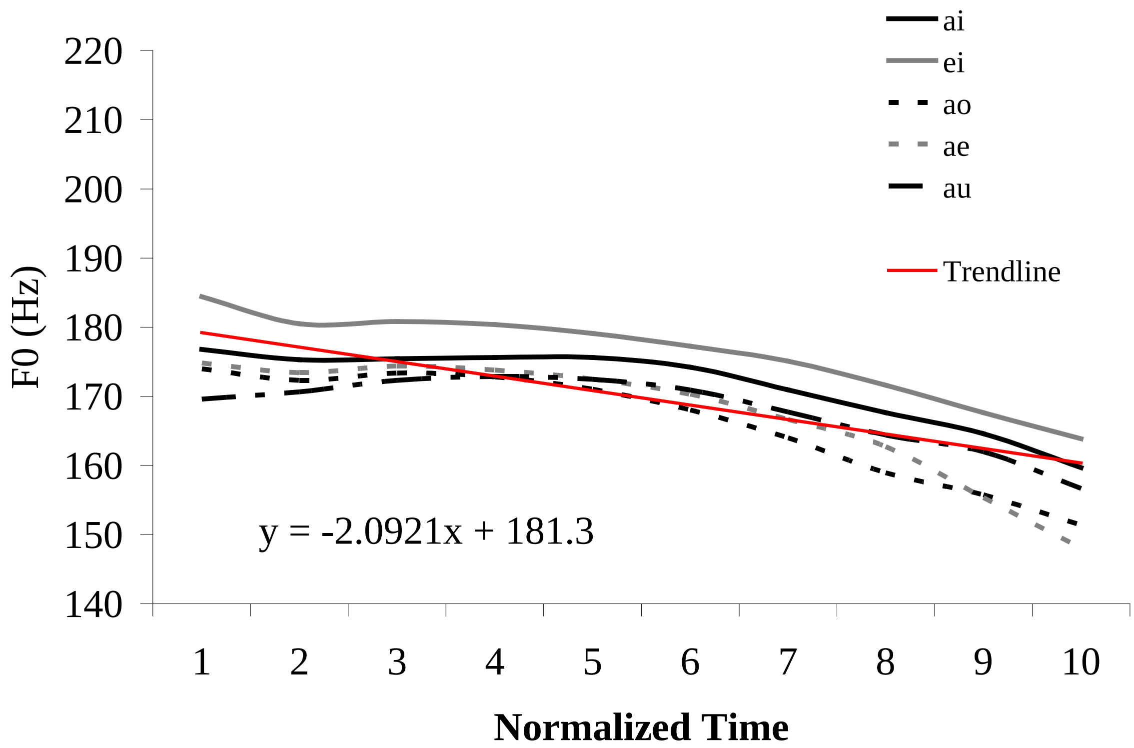

Figure 13. F0 contours (in Hz) for fusing sequences in CVV words.

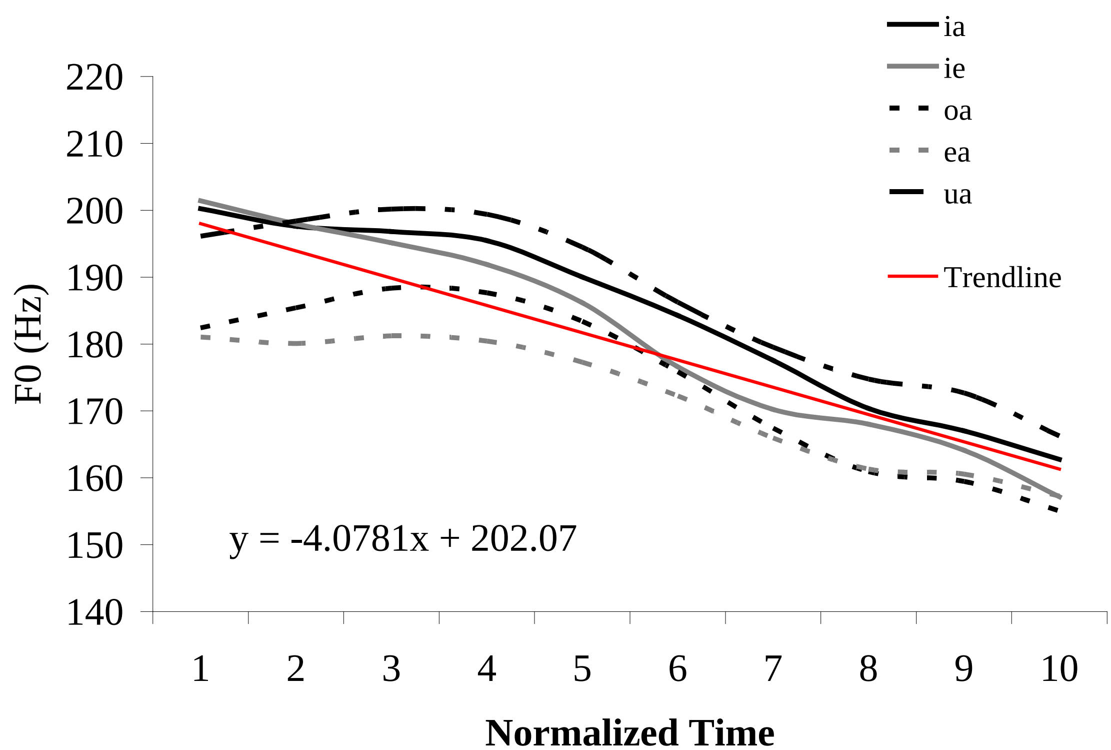

Figure 14. F0 contours (in Hz) for non-fusing sequences in CVV words.

We also looked at CPP contours to determine how voice quality changed over the course of the vowel sequences. Figure 15 and Figure 16 present the CPP contours for fusing sequences and non-fusing sequences, respectively. Looking at the shape of the lines, it is clear that fusing sequences once again have much flatter contours with shallow declines in CPP. This shape is consistent with a fused sequence with stress over the entire sequence because it implies that there is little change in the voice quality throughout the sequence. By comparison, the CPP contours for the non-fusing sequences have large bumps over the first half of the sequence, illustrating a rapid increase in CPP followed quickly by a rapid decrease in CPP before the shallow decline in the second half. This contour indicates that the first half of the non-fusing sequences has more modal voice quality than the second half of the sequences. This result is expected for the non-fusing sequences based on our results from Experiment 1A as well as previous research (Sluijter & van Heuven 1996; Campbell & Beckman 1997) on voice quality in stressed and unstressed syllables. The distinction in CPP contour between fusing and non-fusing sequences may be taken as additional evidence in favor of syllable fusion in Tongan.[4]

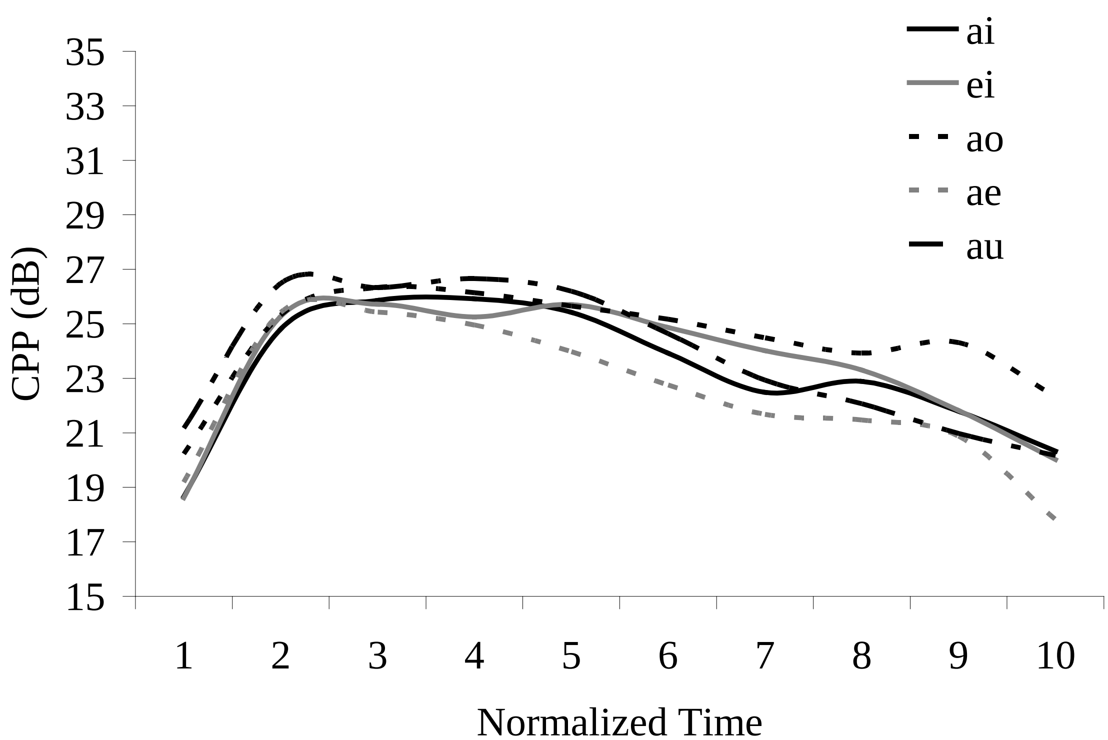

Figure 15. CPP contours for fusing sequences in CVV words

[4] In Figure 16, *ua* does not look like the other non-fusing sequences in that it has a flatter contour. This is likely due to the fact that primary stressed /u/ was not found to be distinguished by CPP.

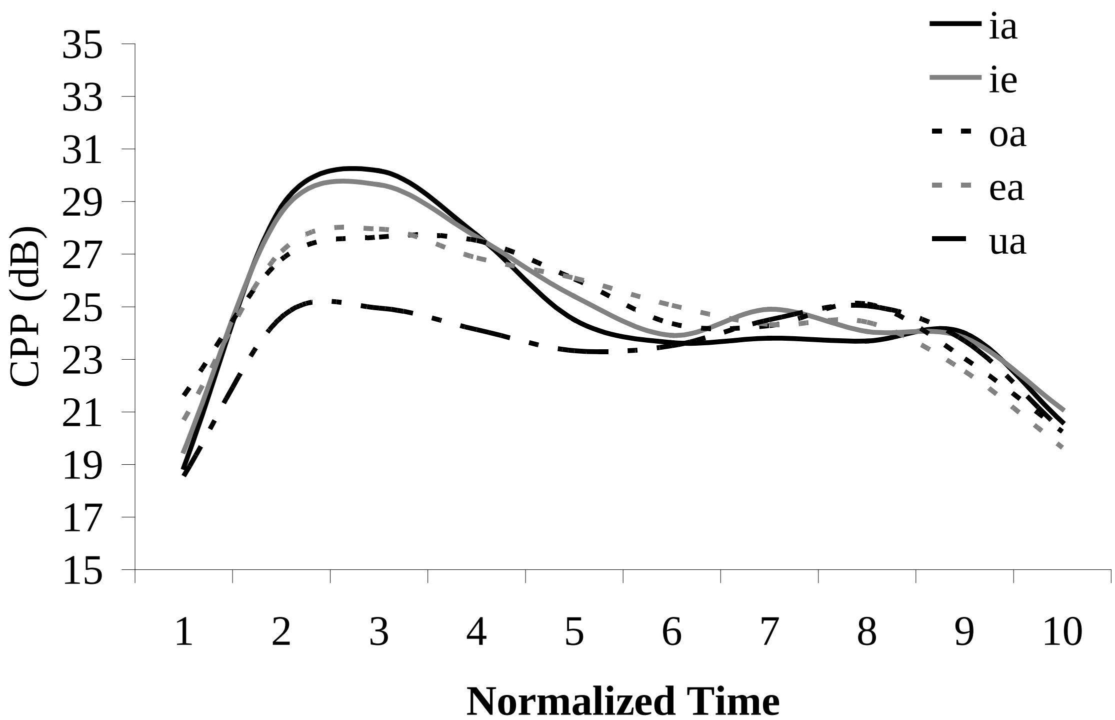

Figure 16. CPP contours for non-fusing sequences in CVV words.

The next place that we looked for differences in fusing and non-fusing sequences was in vowel height. In Experiment 1A, we found that stressed vowels are lower in vowel height (i.e., have a higher F1) than unstressed vowels. To examine how this difference may bear on the question of syllable fusion, we measured the F1 of the second vowels in VV sequences, comparing V2s in stressed position (i.e., V4 of ˌCVVCVˈVCV) with V2s in unstressed position following primary stress (i.e., V2 of ˈCVV). Figure 17 illustrates the exact comparisons that were made using the sequences *ai* and *ia* as an example:

	Fusing Sequence	Non-fusing Sequence
Stressed	[ˌkaikaˈini]	[ˌkiakiˈani]
Unstressed	[ˈkai]	[ˈkia]

Figure 17. V2 comparisons made depending on position for fusing and non-fusing sequences (measured vowel underlined).

For non-fusing sequences, we expect that the V2s in stressed position should have higher F1 values than the V2s in unstressed position based on our results from Experiment 1A. If the fusing syllables are actually undergoing syllable fusion, resulting in stress throughout the sequence, then the V2s for fusing sequences should have very little difference in F1 in these two positions.

Figure 18 shows the difference in F1 between the V2 in stressed position and the V2 in unstressed position for each of the fusing sequences and their non-fusing counterparts. As the figure illustrates, each of the non-fusing sequences has a large positive difference in F1, showing that the V2s are indeed lower in quality in stressed position than in unstressed position. On the other hand, the V2s of the fusing syllables have small (and sometimes negative) differences in F1 depending on position, indicating that the V2s in ˈCVV words are acting like stressed vowels. These results provide evidence for the syllable fusion process because they imply that fusing syllables have stress throughout the sequence rather than only on the first vowel of the sequence.

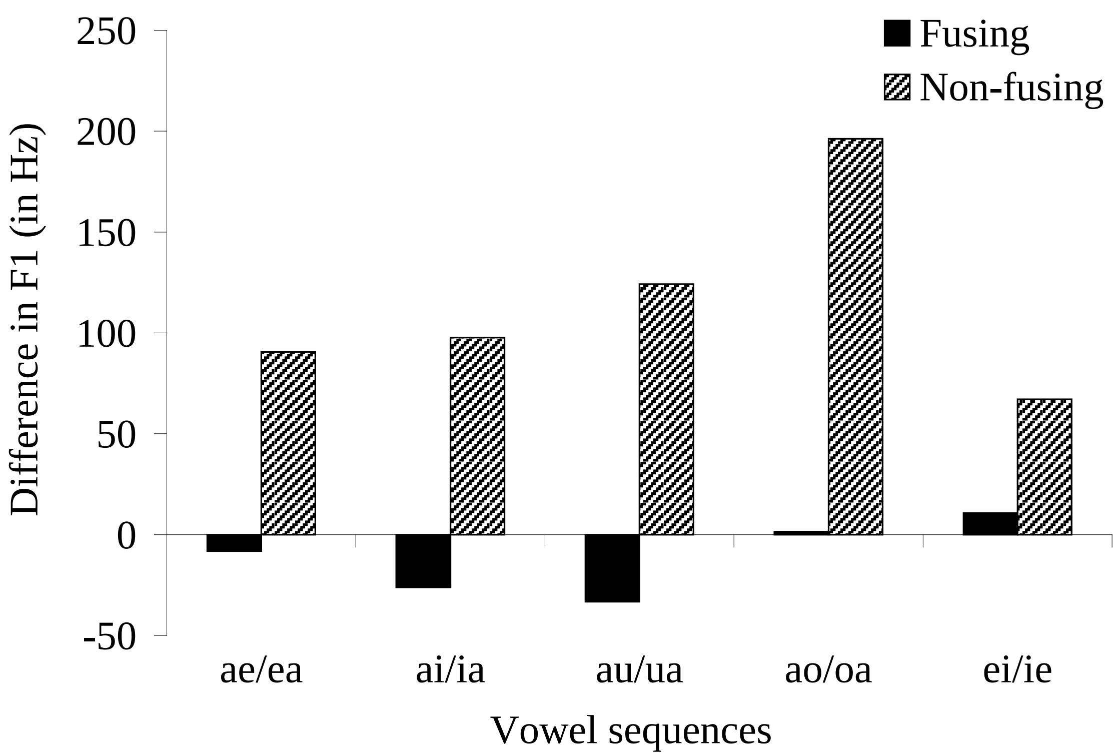

Figure 18. Difference in F1 between V2s in stressed and unstressed positions in VV sequences.

4.3.2 Secondary stress

For cues to fusion in sequences with secondary stress, we looked only at two measures: pitch and duration. Pitch was used because it was the only consistent cue to secondary stress (except for /a/). Duration was used because the question of fusion is not solely a matter of stress, given that a fused sequence should, in theory, be shorter than a non-fused sequence. This is due to the fact that a fused sequence consists of one syllable, whereas the non-fused one consists of two.[5]

We plotted the time courses of F0 for fusing sequences, shown in Figure 19. The figure shows that these sequences, whose stress in theory falls on the first vowel, have a rather flat pitch contour, with a slope of -0.84, similar to the fusing sequences with primary stress shown in Figure 13.

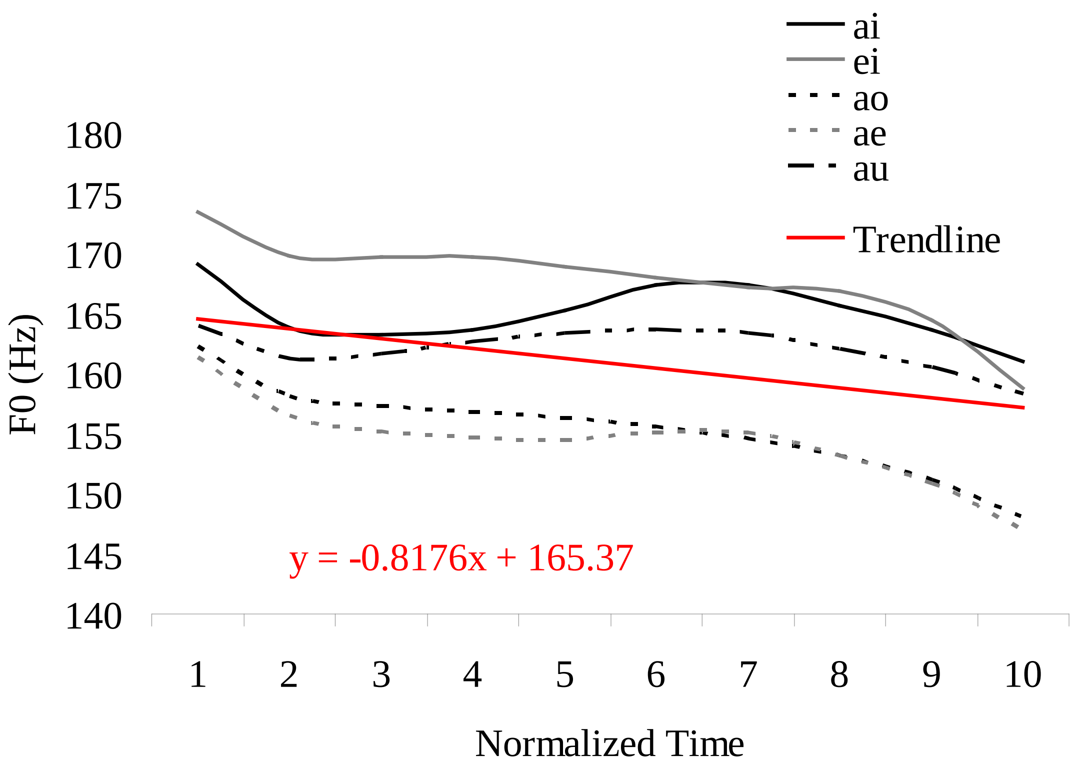

Figure 19. F0 contours for fusing sequences with secondary stress.

The time courses for non-fusing sequences are shown in Figure 20. The steeper slope of these contours (-2.51) indicates that the pitch peak is localized on the first vowel of the sequence, and then drops for the second vowel. This suggests that these sequences have two distinct pitch targets, in contrast to the fusing sequences shown above. This pitch contour difference between fusing and non-fusing sequences mirrors that found for sequences with primary stress above.

[5] Duration was not used for sequences with primary stress, because these have the confound of being in word-final position, which would lengthen the duration. It is not possible to have a primary stressed sequence that is not word-final in Tongan.

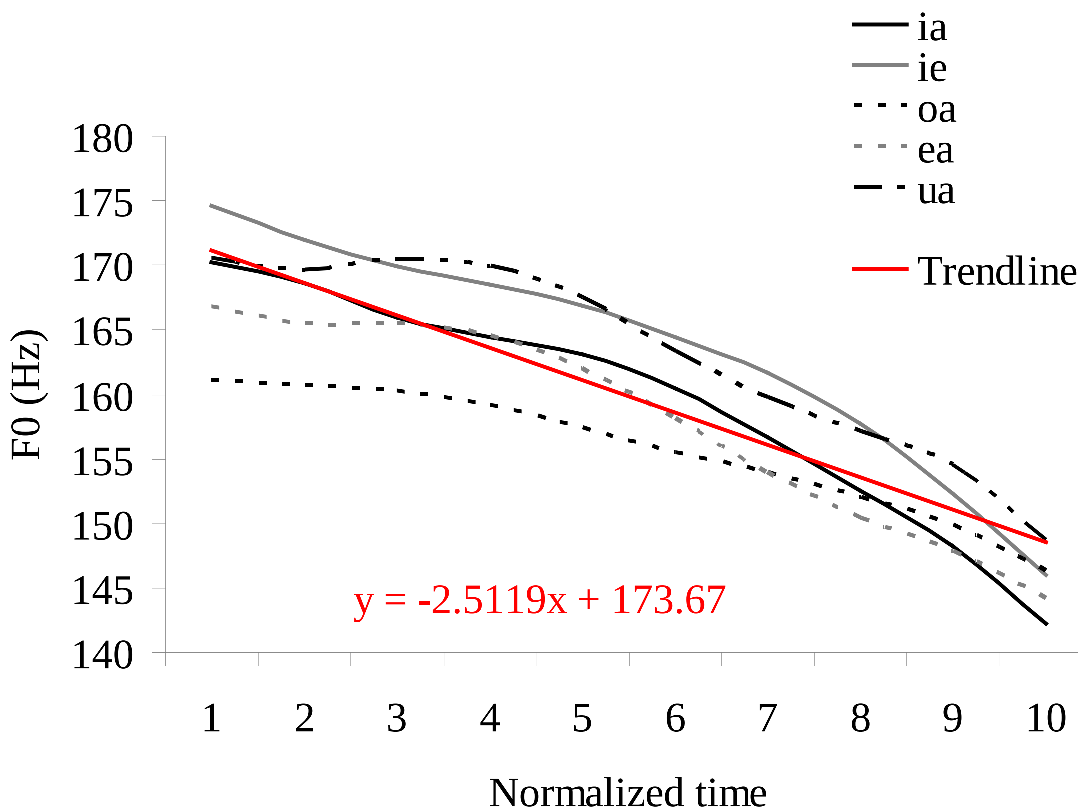

Figure 20. F0 contours for non-fusing sequences with secondary stress.

The difference in duration between fusing and non-fusing sequences with secondary stress is shown in Figure 21. The results show that /ae, ao, au/ are shorter than /ea, oa, ua/, as expected if fusion occurred. This pattern was not found for /ai, ei/, however. We suspect this could be due to shorter duration of /i/ in the non-fusing cases.

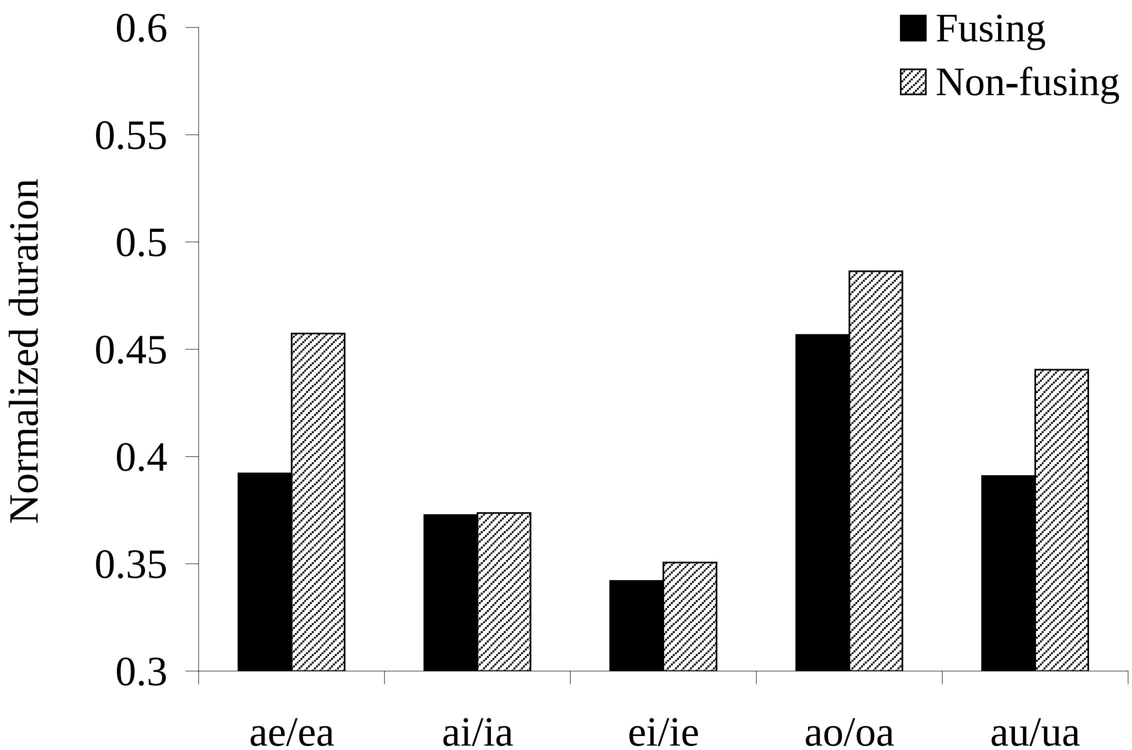

Figure 21. Difference in duration for fusing and non-fusing sequences.

The differences in pitch contours and duration between fusing and non-fusing sequences suggest that, even under secondary stress, fusion can occur.

5 Conclusions

To conclude, we have shown that there exist several correlates of primary stress in Tongan. Pitch, duration, and vowel quality appear to be strong cues, but intensity and voice quality can differentiate stressed from unstressed tokens for certain vowels. The differences were in the expected direction, with stressed vowels being higher pitched, longer, lower, more intense and less breathy than their unstressed counterparts. For secondary stress, only F0 was found to differentiate stressed from unstressed vowels, but the effect was much less than for primary stress.

We then used these stress cues to argue that syllable fusion does occur in Tongan. In fused sequences, both vowels show signs of stress. For primary stress, fused sequences show a flatter pitch and CPP contour, whereas non-fused sequences have a sharp decline in pitch and CPP in the first half. In addition, the second half of fused sequences has an F1 that more comparable to that of stressed vowels. On the other hand, the second half of non-fused sequences has a lower F1 than stressed vowels, implying that it is unstressed. All this strongly suggests that stress is only on the first vowel of a non-fused sequence, but on both vowels of a fused sequence.

For secondary stress, the same pattern holds for the F0 contour. Pitch was the only cue found to distinguish vowel with secondary stress from unstressed ones. Other cues were less informative, which is not surprising given their inability to distinguish secondary stress in Experiment 1B. The duration results also provide evidence for fusion for sequences with secondary stress, in that fused sequences are generally shorter than non-fused ones. We take these tests as converging acoustic evidence that syllable fusion is an active phonological property affecting lower-to-higher vowel sequences in Tongan.

References

Adisasmito-Smith, N. & A. C. Cohn. 1996. Phonetic correlates of primary and secondary stress in Indonesian: A preliminary study. *Working papers of the Cornell phonetics laboratory* 11, Cornell.

Anderson, V. and Y. Otsuka (2003). Phonetic Correlates of Length, Stress and Definitive Accent in Tongan. *Proceedings of the Fifteenth International Congress of Phonetic Sciences, Universitat Autònoma de Barcelona*, 2047-2050.

Anderson, V. & Y.Otsuka. 2006. The phonetics and phonology of “Definitive Accent” in Tongan. *Oceanic Linguistics* 45(1), 25-42.

Boersma, P. & D. Weenink. 2009. Praat: doing phonetics by computer (version 5.1.14) [Computer program]. Retrieved August 30, 2009, from http://www.praat.org/.

Campbell, N. & M. Beckman. 1997. Stress, prominence, and spectral tilt. In Botinis, A., G. Kouroupetroglou, & G. Carayiannis (Eds.) *Intonation: Theory, Models and Applications (Proceedings of an ESCA Workshop, September 18-20, 1997, Athens, Greece*). ESCA and University of Athens Department of Informatics, 67-70.

Cho, T. & P. A. Keating. 2010. Effects of initial position versus prominence in English. *Journal of Phonetic* 37(4), 466-485.

Churchward, C. M. 1953. *Tongan grammar*. Oxford: Oxford University Press.

Everett, K. M. 1998. The acoustic correlates of stress in Pirahã. *The Journal of Amazonian Languages* 1, 104-162.

Feldman, H. 1978. Some notes on Tongan phonology. *Oceanic Linguistics* 17(2), 133-139.

Gordon, M. & A. Applebaum. 2010. Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association* 40, 35-58.

Hillenbrand, J., R. A. Cleveland & R. L. Erickson. 1994. Acoustic correlates of breathy voice quality. *Journal of Speech and Hearing Research* 37, 769-778.

Kahn, D. 1976.Syllable-based generalizations in English phonology. Ph.D. Dissertation, MIT. [Distributed by the Indiana University Linguistics Club, Bloomington].

Kawahara, H., L. Masuda-Katsuse & A. de Cheveigné. 1999. Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication* 27, 187–207.

Kochanski, G., E., Grabe, J. Coleman, & B. Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America* 118, 1038-1054.

Lewis, M. Paul (ed.), 2009. Ethnologue: Languages of the World, Sixteenth edition. Dallas, Tex.: SIL International. Online version**:** http://www.ethnologue.com/**.**

Lieberman, P. 1960. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America* 32, 451–454.

Poser, W. J. 1985. Cliticization to NP and Lexical Phonology. In J. Goldberg et al. (eds.), *Proceedings of West Coast Conference in Formal Linguistics* 4, Stanford Linguistics Association, CSLI, 262–272.
Schütz, A. J. 2001. Tongan accent. *Oceanic Linguistics* 40, 307-323.
Shue, Y.-L., P. A. Keating & C. Vicenik. 2009. VoiceSauce: A program for voice analysis. Poster presented at the 158th meeting of the Acoustical Society of America, San Antonio, TX.
Sjölander, K. 2004. The Snack Sound Toolkit [Computer program]. Retrieved May 25, 2010, from http://www.speech.kth.se/snack/.
Sluijter, A. M. C. & V. J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4), 2471-2485.
Taumoefolau, M. 2002. Stress in Tongan. *MIT Working Papers in Linguistics, 44*, MIT.
Zuraw, K., O'Flynn, K., & Ward, K. 2010. Marginal prosodic contrasts in Tongan loans. UCLA Phonology seminar presentation, June 2, 2010.

Appendices

Appendix A

Wordlist used for Experiment 1A

/a/		**/e/**		**/i/**	
Word	**Meaning**	**Word**	**Meaning**	**Word**	**Meaning**
mafana	warm (of food, water)	melemo	to be drowned	kikila	to look with widely open eyes, stare
talamu	chew	tekena	to be pushed up or out	kikilo	to roll the eyes
panaki	to be near, close	kekena	going yellow (of leaves)	kikite	to have the power of seeing into the future
papaka	to be nervous	nenefu	blurred, indistinct	kilisi	to saw, mince meat
manafa	piece of open ground	pepenu	flexible but difficult to break	mimili	to rub roughly
manatu	to think of	nenenu	keep hesitating	mimisi	to suck up
papani	to besmear	tetepa	to look cross-eyes, to squint	ninimo	to suffer from vertigo
pakaka	dry and rough or stiff	tetenga	painful because of a squeeze	pipiki	to hold on or adhere
takafi	outer cover for something	kekete	really full (from food)	pipine	(of hair) clogged with dirt
makaka	rough (speech,behavior)	teteka	(of the eyes) continually rolling about	sinifu	unmarried wife

/o/		/u/	
Word	**Meaning**	**Word**	**Meaning**
kolonga	camping place of pigeon-catching	putuki	to plant close together
kotofa	to appoint a time	kukuta	to keep a firm grip on oneself
momoko	cold	mumuni	to shade (the eyes) with the hand
popongo	to tend or look after	tufunga	skilled workman
popoto	(of two or more) to get along alright together	nunumi	gather together fast (not folding it up)
tokoni	to help	tukuku	kind of bird
tokoto	to lie (down)	punusi	patching the tapa
totofa	to strike out a new path for oneself	mumutu	to cut off roughly
nonofo	to live together in one house	pununga	nest
tokosi	few	tutuku	to disperse

Appendix B

Wordlist used in Experiment 1B

/a/	/e/	/i/	/o/	/u/
mafanani	melemoni	kikilani	kolongani	putukini
talamuni	tekenani	kikiloni	kotofani	kukutani
panakini	kekenani	kikiteni	momokoni	mumuni'i
papakani	nenefuni	kilisini	popongoni	tufungani
manafani	pepenuni	mimilini	popotoni	nunumini
manatuni	nenenuni	mimisini	tokoni'i	tukukuni
papani'i	tetepani	ninimoni	tokotoni	punusini
pakakani	tetengani	pipikini	totofani	mumutuni
takafini	keketeni	pipineni	nonofoni	punungani
makakani	tetekani	sinifuni	tokosi'i	tutukuni

Appendix C

Wordlist used in Experiment 2

ˈCVV words		ˌCVVCVˈVCV words	
mae	fao	faofaoni	maemaeni
tae	kao	kaokaoni	taetaeni
pae	tea	teateani	paepaeni
tei	sea	seaseani	teiteini
sei	lea	lealeani	seiseini
kei	kie	kiekieni	keikeina
tui	fiu	fiufiuni	tuituini
kui	niu	niuniuni	kuikuini
mui	kiu	kiukiuni	muimuini
mai	sia	siasiani	maimaini
kai	kia	kiakiani	kaikaini
fai	lua	lualuani	faifaini
lau	tua	tuatuani	laulauni
tau	fua	fuafuani	tautauni
vau	loa	loaloani	vauvauni
lao	poa	poapoane	laolaoni