Stress correlates and vowel targets in Tongan

Marc Garellek and James White[1]
marcgarellek@ucla.edu; jameswhite@ucla.edu

Abstract

In this study, we determine the acoustic correlates of primary and secondary stress in Tongan. Vowels with primary stress show differences in F0, intensity, duration, F1, and voice quality, but F0 is the best predictor of primary stress. Vowels with secondary stress are mainly cued by a difference in F0. With regards to the effects of stress on the vowel space, we find that all five Tongan vowels are higher in the vowel space (have lower F1) when unstressed, with no differences in F2. Moreover, there is no reduction in the overall size of the vowel space. We interpret this pattern as evidence that unstressed vowels in Tongan are not undergoing centralization, nor are they otherwise reduced. Rather, Tongan speakers have separate targets for stressed and unstressed vowels.

1 Introduction

The goal of this paper is to determine which acoustic measures correlate with both primary and secondary stress in Tongan (Malayo-Polynesian, Austronesian; Blust 2009), and to determine which of the measures best predict stress in the language. Studies on the acoustic correlates of stress date from pioneering work on English (Fry 1955). However, the work in this area has focused on only a few languages, coming from a limited set of language families. There have been few acoustic studies of stress in Polynesian languages, and even fewer of Tongan in particular (but see Anderson and Otsuka 2003, 2006). Analyzing data from a wide array of cross-linguistic studies is vital for understanding how stress is realized in language in general, such as providing insight on which aspects of stress are universal and which are language-specific. For example, Gordon and Applebaum (2010:35-36) cite cases of languages in which typologically-common acoustic measures of stress do not distinguish stressed from unstressed vowels, because such measures serve as the primary cue to other contrasts in the language, e.g. F0 in languages with lexical tone.

Cross-linguistic studies have shown that multiple acoustic measures may correlate with stress in vowels. Typically, stressed vowels may 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, Gordon & Nafi 2012), and longer duration (Lieberman 1960; Everett 1998; Gordon & Applebaum 2010). Differences in F1 and F2, associated with differences in vowel quality, have also been found, including higher F1

[1] Author names in alphabetical order.

(Cho & Keating 2009; Gordon & Applebaum 2010). Researchers have also found differences in measures associated with voice quality or phonation (Sluijter & van Heuven 1996). Moreover, not all of these acoustic measures necessarily correlate with stress for any given language, and secondary stress may be cued differently than primary stress (Adisasmito-Smith & Cohn 1996, Gordon & Applebaum 2010, Plag, Kunter, & Schramm 2011).

Changes in vowel quality as a function of stress are often discussed from the perspective of synchronic precursors to phonological vowel reduction. For instance, Crosswhite (2001) reviewed common patterns in phonological, stress-based vowel reduction, and identified two main patterns: centralization vs. merging of contrasts. Indeed, stressed vowels are often more peripheral: high vowels are higher when stressed, and low vowels lower. This has been found in languages with five-vowel systems similar to the one found in Tongan, e.g. Castilian Spanish (Ortega-Llebaria & Prieto 2011). Thus, we might expect unstressed vowels in Tongan to undergo centralization, yielding lower high vowels and higher low vowels.

The paper is organized in the following manner: we first give a brief overview of the phoneme and stress systems of Tongan. We then discuss our methodology followed by the results of our study for each measure. Finally, we end with a discussion of our findings, their implications, and the conclusions of the paper.

2 Tongan Background

Tongan has twelve consonant phonemes and five vowel phonemes, presented in Figure 1. In the consonants, Tongan has a voicing distinction only between the labiodental fricatives /f/ and /v/. For vowels, there is arguably a distinction between long and short vowels, as can be seen from pairs such as *pepe* [ˈpepe] 'butterfly' and *pēpē* [ˌpeːˈpeː] 'baby'. Long vowels have alternatively been analyzed as sequences of two short vowels in separate syllables (Taumoefolau 2002, Anderson & Otsuka 2006), but the exact nature of these "long" vowels will not be important in this paper.

Primary stress in Tongan always falls on the penultimate mora of a phonological word. Secondary stress assignment depends on morphology (Feldman 1978) and can be variable for loanwords (Zuraw, O'Flynn, and Ward 2010), but always falls on the leftmost mora in the examples used in this study.

Figure 1. Tongan consonant inventory (top) and vowel inventory (bottom).

	Bilabial	Labiodental	Dental	Velar	Glottal
Plosive	p		t	k	ʔ
Fricative		f v	s		h
Nasal	m		n	ŋ	
Lateral Approximant			l		

	Front	Central	Back
Close	i		u
Mid	e		o
Open		a	

3 Method

3.1 Participants

Four female native Tongan speakers living in the Los Angeles area participated in the study (approximate ages between 45 and 60 years old). All had moved to the United States from Tonga as adults and reported still communicating in Tongan on a daily basis. The participants received a small monetary compensation for their participation.

3.2 Materials

For primary stress, we used CVˈCVCV words in which the first vowel was the same as the second vowel, for example *nenenu* [neˈnenu] 'to hesitate persistently.' This allowed us to compare stressed and unstressed (short) vowels of the same type within the same word so that extraneous factors that may vary across repetitions (e.g., speaking rate) could be well controlled. Ten words of this type were created for each of the five Tongan vowels. For secondary stress, the same set of words was used, but a CV suffix, usually the highly-productive demonstrative suffix *–ni*, was added to each word. The resulting ˌCVCVˈCV-CV words allowed us to once again compare the first and second vowels, which now had secondary stress and no stress, respectively. A wide variety of consonants was used in each consonant position in order to control for consonantal effects on the target stressed/unstressed vowels.

3.3 Procedure

The words were collected into a wordlist written in Tongan orthography, which was read by the speakers. Each word was repeated in the carrier phrase *Angimui 'a e fo'ilea ko e ______ kiateau* ([ˌaŋiˈmui ˈʔae ˌfoʔiˈlea ˈkoe ______ ˌkiateˈau]) 'Repeat the word ______ for me.' Three repetitions were collected for each word, yielding a total of 30 tokens per speaker for each vowel for both primary and secondary stress. The recordings were made in a UCLA Phonetics Lab 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 22,050 Hz.

The first and second vowels of each word were labeled in Praat textgrids (Boersma and Weenink 2009). The boundaries for vowels were segmented according to the beginning and end of a clear second formant. The labeled sound files were then run through VoiceSauce (Shue et al. 2011) to obtain the acoustic measures, which were calculated for every millisecond. VoiceSauce calculates F0 using the STRAIGHT algorithm (Kawahara, Masuda-Katsuse, & de Cheveigné 1999). VoiceSauce also outputs the duration of the labeled segment as well as values for F1, F2, and Root Mean Square (RMS) energy. The formants were measured using the Snack SoundToolkit (Sjölander 2004). We also include here two acoustic correlates of voice quality: H1*-H2* and cepstral peak prominence (CPP). H1*-H2* is a measure of the difference in amplitude between the first and second harmonics. Its values have been corrected for formants (hence the use of asterisks) following the correction by Hanson (1997) and Iseli, Shue, & Alwan (2007), in order to enable cross-vowel comparison. H1*-H2* is perhaps the most commonly used harmonic measure of voice quality. Values of H1*-H2* are typically higher for breathy voice when compared to modal voice, and lower for creaky or laryngealized voice when compared to modal voice (Klatt & Klatt 1990, Gordon & Ladefoged 2001). CPP, calculated using the algorithm from Hillenbrand, Cleveland, and Erickson (1994), is a measure of noise and aperiodicity. Both aspiration noise during breathy voice or aperiodic voicing during creaky voice may result in lower values of CPP (Garellek & Keating 2011, Garellek 2012). Table 1 provides a summary of the acoustic measures recorded with a brief description of each measure.

The average values for each measure were calculated automatically by VoiceSauce, and the results were then saved to a text file for subsequent analysis.

Table 1. Summary of acoustic measures.

Measure	Description
Fundamental frequency (F0)	Frequency of lowest harmonic, correlated with perceived pitch. Measured in Hertz (Hz).
Duration	Duration of the vowel, measured in milliseconds.
RMS energy	Root mean squared energy, corresponding to intensity/loudness.
First formant (F1)	First formant, measured in Hz. Correlates with vowel height.
Second formant (F2)	Second formant, in Hz. Correlates with vowel frontness.
H1*-H2*	Corrected difference in amplitude between the first and second harmonics, in decibels (dB). Correlates with voice quality (higher = breathier).
Cepstral peak prominence (CPP)	Measure of regularity and magnitude of harmonics above the noise floor (lower CPP = noisier signal, e.g. due to aspiration or irregularity).

4 Results

4.1 Method of Statistical Analysis

In the following sections, we determine whether each of the acoustic measures was a statistically significant correlate of primary and secondary stress in Tongan. The values of each measure were analyzed using linear mixed effects models. These were implemented in R (R Development Core Team 2008) using the *lmer()* function of the *lme4* package (Bates, Maechler, & Dai 2008), following Baayen (2008a:ch. 7). Separate models were fitted for primary and secondary stress. The models contained a fixed effect for presence of stress (stress or no stress), and three random intercepts: speaker, word, and repetition. By including these variables as random effects, we effectively controlled for any effect that they may have had on the results, so that we can determine what effect stress alone had on a particular measure. Overall, the random effects listed here were included in the model because they significantly improved model fit, which was assessed using an ANOVA test comparing likelihood between models (see Baayen 2008a). Inclusion of random slopes did not improve model fit, so none were included. Including Repetition as a random effect was found to significantly improve model fit, probably because some speakers had a tendency to speak more quickly in later repetitions. For each model, we report t-values provided in the model output, as well as p-values obtained using the *pvals.fnc()* function of the *languageR* package (Baayen 2008b), which estimates *p*-values by conducting Markov Chain Monte Carlo (MCMC) sampling with 10,000 simulations.

In the cases where the overall model revealed a significant effect, additional linear mixed-effects models were run on subsets of the data corresponding to each of the five Tongan vowels to determine if the overall pattern was also found for individual vowels. These within-vowel models included the same fixed and random structures as the overall models.

4.2 Fundamental Frequency (F0)

Figure 2 and Table 2 show the mean values of F0 for primary and secondary stressed vowels and their unstressed counterparts. Vowels with primary stress have significantly higher F0 values overall (by about 50 Hz) than those without stress. Post-hoc within-vowel comparisons show that, for each of the five vowels, primary stress results in significantly higher F0 values relative to unstressed vowels. Vowels with secondary stress also have significantly higher F0 values than those without stress overall, but this difference of about 9 Hz is much smaller than the difference found for primary stress. Post-hoc within-vowel comparisons show that the difference in F0, though small, is significant for each of the five vowels individually.

Figure 2. Mean F0 (in Hz) by vowel, for primary stress (left panel) and secondary stress (right panel). Errors bars represent standard error of the mean.

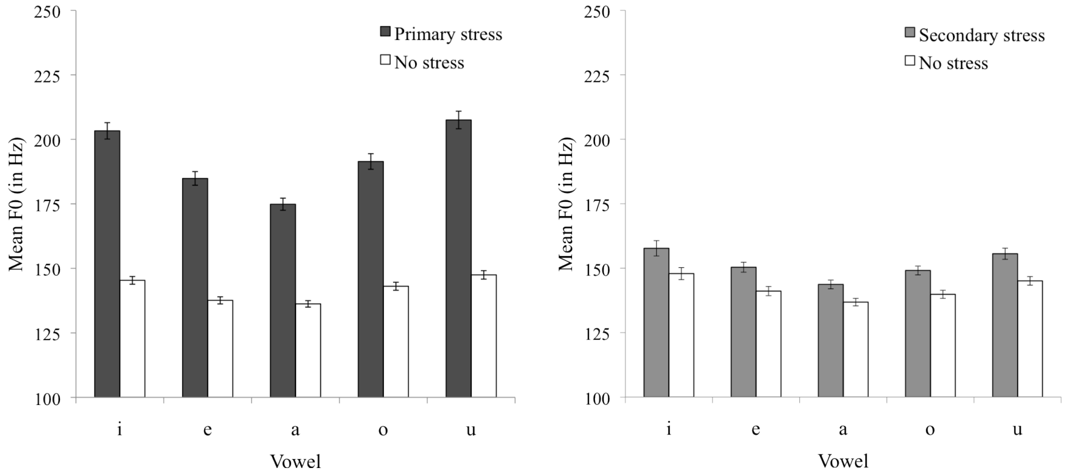


Table 2. Mean F0 (in Hz) overall and by vowel, for primary stress (left panel) and secondary stress (right panel). P-values represent differences between stress and no stress.

	Mean primary	Mean no stress	T-value	*P*-value
Overall	192.31	141.93	53.24	.0001
/i/	203.30	145.36	29.59	.0001
/e/	184.86	137.60	28.73	.0001
/a/	174.86	136.24	22.38	.0001
/o/	191.43	143.07	21.51	.0001
/u/	207.51	147.48	25.15	.0001

	Mean secondary	Mean no stress	T-value	*P*-value
Overall	151.37	142.18	10.49	.0001
/i/	157.71	147.89	4.38	.0001
/e/	150.36	141.11	4.73	.0001
/a/	143.70	136.83	4.10	.0002
/o/	149.11	139.86	6.03	.0001
/u/	155.60	145.08	5.50	.0001

4.3 Duration

Figure 3 and Table 3 present the mean durations for vowels with primary and secondary stress as well as their unstressed counterparts. We find that vowels with primary stress are significantly longer in duration (by about 30 ms) than unstressed vowels. Within-vowel comparisons show that this difference in duration holds for all five Tongan vowels individually. For secondary stress, we find a surprising effect: vowels with secondary stress are slightly *shorter* than unstressed vowels. This difference is significant overall, but by-vowel comparisons indicate that the difference is only significant for /a/ and marginally for /o/. The difference for the other three vowels, though in the same direction, does not reach significance.

It is possible that the shortened duration found for vowels with secondary stress was due to initial word position rather than to stress itself. This strikes us as unlikely, because word initial positions are associated with increased duration in other languages (Turk & Shattuck-Hufnagel 2000). Nevertheless, due to this possible confound, the fact that some of the within-vowel comparisons did not reach significance, and the overall small magnitude of the difference, we conclude that the effect of secondary stress on duration is weak at best.

Figure 3. Mean duration (in ms) by vowel, for primary stress (left panel) and secondary stress (right panel). Errors bars represent standard error of the mean.

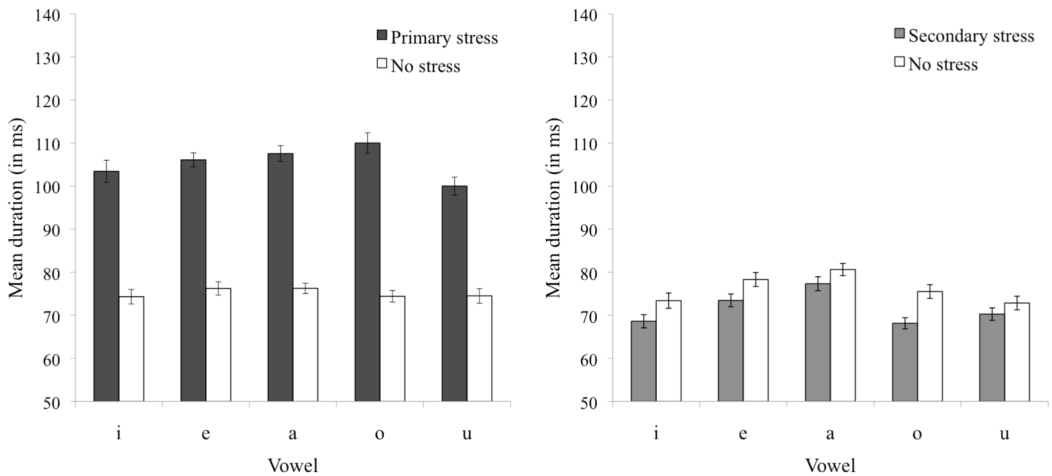


Table 3. Mean duration (in ms.) overall and by vowel, for primary stress (left panel) and secondary stress (right panel). P-values represent differences between stress and no stress.

	Mean primary	Mean no stress	T-value	*P*-value
Overall	105.36	75.16	29.08	.0001
/i/	103.44	74.32	10.00	.0001
/e/	106.10	76.23	17.08	.0001
/a/	107.56	76.26	15.96	.0001
/o/	110.01	74.41	14.55	.0001
/u/	100.02	74.51	12.26	.0001

	Mean secondary	Mean no stress	T-value	*P*-value
Overall	71.50	76.10	3.34	.0008
/i/	68.62	73.40	.93	.354
/e/	73.46	78.31	.97	.334
/a/	77.34	80.63	2.94	.005
/o/	68.16	75.52	2.05	.042
/u/	70.28	72.85	.89	.384

4.4 RMS Energy

Figure 4 and Table 4 show the mean values of RMS energy for vowels with primary and secondary stress as well as their unstressed counterparts. We find significantly greater energy in vowels with primary stress than in unstressed vowels, and within-vowel comparisons indicate that this difference is significant for each of the five Tongan vowels individually. We also find that vowels with secondary stress have significantly higher energy than unstressed vowels overall, but within-vowel comparisons show that this difference is only significant for central and back vowels: /a/, /o/ and (marginally) /u/.

Figure 4. Mean RMS energy by vowel, for primary stress (left panel) and secondary stress (right panel). Errors bars represent standard error of the mean.

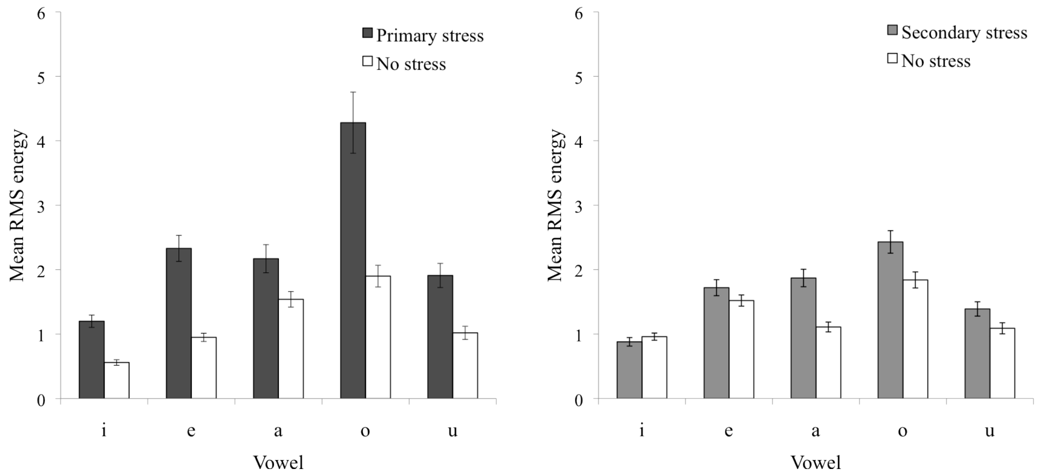


Table 4. Mean RMS energy overall and by vowel, for primary stress (left panel) and secondary stress (right panel). P-values represent differences between stress and no stress.

	Mean primary	Mean no stress	T-value	*P*-value
Overall	2.35	1.19	9.84	.0001
/i/	1.20	.56	6.80	.0001
/e/	2.33	.95	7.77	.0001
/a/	2.17	1.54	3.76	.0004
/o/	4.28	1.90	4.79	.0001
/u/	1.91	1.02	4.87	.0001

	Mean secondary	Mean no stress	T-value	*P*-value
Overall	1.66	1.30	5.08	.0001
/i/	.88	.96	-1.11	.261
/e/	1.72	1.52	1.14	.266
/a/	1.87	1.11	5.02	.0001
/o/	2.43	1.84	3.44	.001
/u/	1.39	1.09	2.07	.042

4.5 F1 and F2

Table 5 shows the mean F1 values for vowels with primary and secondary stress and their unstressed counterparts. Overall, the F1 for vowels with primary stress is significantly higher than the F1 for unstressed vowels by about 57 Hz. Within-vowel comparisons indicate that the difference in F1 is significant for all five of the Tongan vowels in the same direction: higher F1 for vowels with primary stress. For secondary stress, we find that a small difference in the same direction (i.e., higher F1 for secondary stress) just reaches significance overall, but within-vowel comparisons reveal that this overall difference is only driven by a large difference for the vowel /a/. The other four vowels do not have significantly different F1 values for vowels with secondary stress and unstressed vowels.

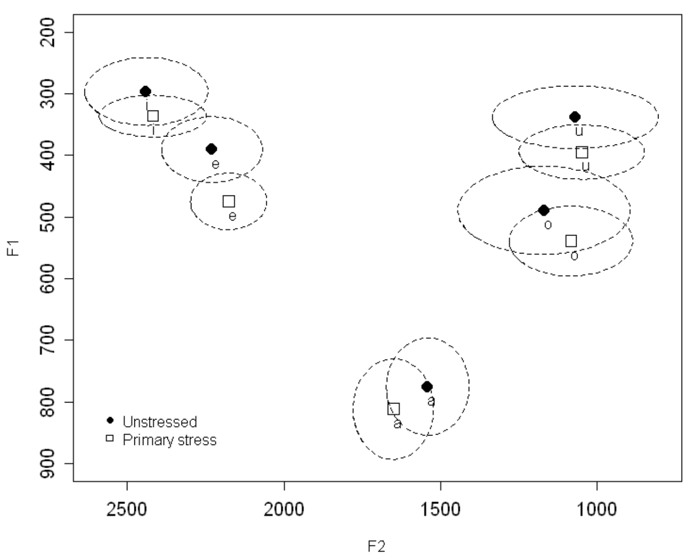
Table 5. Mean F1 overall and by vowel for primary stress (left panel) and secondary stress (right panel). P-values represent differences between stress and no stress.

	Mean primary	Mean no stress	T-value	P-value
Overall	515.75	458.98	14.24	.0001
/i/	336.22	296.39	6.23	.0001
/e/	474.66	390.54	19.61	.0001
/a/	811.20	774.97	4.11	.0001
/o/	541.78	501.38	3.37	.001
/u/	400.47	340.94	8.92	.0001

	Mean secondary	Mean no stress	T-value	P-value
Overall	460.62	456.50	2.13	.034
/i/	304.62	306.65	-.10	.921
/e/	403.92	411.83	-1.42	.155
/a/	750.06	717.01	3.52	.0002
/o/	498.74	503.43	-.85	.393
/u/	354.99	351.08	.44	.661

Looking at the F2 values, we find no significant overall difference in F2 for vowels with primary stress (mean = 1738 Hz) and their unstressed counterparts (mean = 1746 Hz), $t = -.16$, $p = .87$. We also find no significant overall different in F2 for vowels with secondary stress (mean = 1743 Hz) and their unstressed counterparts (mean = 1736 Hz), $t = .52$, $p = .60$.

Figure 5. Vowel plot for primary stress vs. unstressed vowels. F1xF2 clouds show one standard deviation from mean value.



As seen in the vowel plot in Figure 5, vowels with primary stress in Tongan are generally lower in the vowel space (i.e., have a higher F1) than their unstressed counterparts. Thus, there is a general shifting-up of all vowels in the vowel space when unstressed. This shift in F1 is not accompanied by a significant change in F2.

It should also be noted that the vowel space in Figure 5 shows little to no overlap between the five vowels even when they are unstressed. For each vowel, there is some overlap between the two stress conditions (e.g., between stressed /u/ and unstressed /u/), but the data show that in Tongan, both stressed and unstressed vowels are well dispersed. Implications of this pattern will be discussed in section 4.3.

4.6 Voice quality measures (H1-H2 and CPP)

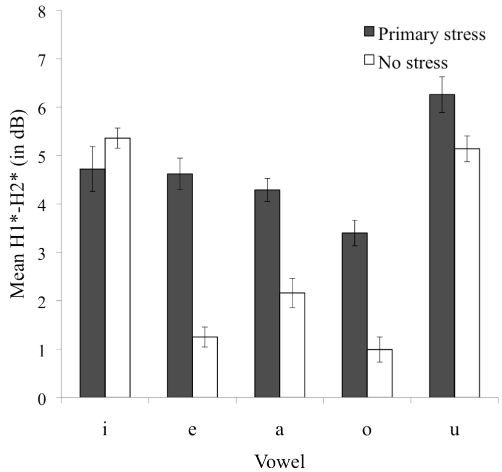
The results for H1*-H2* show a significant main effect for primary stress (see Table 6). Within-vowel comparisons show that all vowels but /i/ have significantly higher H1*-H2* values when they bear primary stress than when they are unstressed. For /i/, the unstressed vowel has a higher value of H1*-H2*, for reasons that remain unclear. The overall average of 4.68 dB for vowels with primary stress is high, suggesting that vowels with primary stress are breathy. To confirm this, we report below the results for CPP. If vowels are breathier when they bear primary stress, it is expected that they should also be noisier (i.e., they should have a lower CPP value: Hillenbrand, Cleveland & Erickson 1994).

There was no significant difference in H1*-H2* between vowels with secondary stress (mean = 2.56 dB) and their unstressed counterparts (mean = 2.38 dB), $t = 1.11$, $p = .27$.

Table 6. Mean H1*-H2* (in dB), overall and by vowel, for primary stress. P-values represent differences between stress and no stress.

	Mean primary	Mean no stress	T-value	P-value
Overall	4.68	3.02	8.48	.0001
/i/	4.72	5.36	-2.19	.0334
/e/	4.62	1.25	10.99	.0001
/a/	4.29	2.16	7.21	.0001
/o/	3.40	.99	7.58	.0001
/u/	6.26	5.14	2.13	.0338

Figure 6. Mean H1*-H2* by vowel, for primary stress (left panel) and secondary stress (right panel). Errors bars represent standard error of the mean.



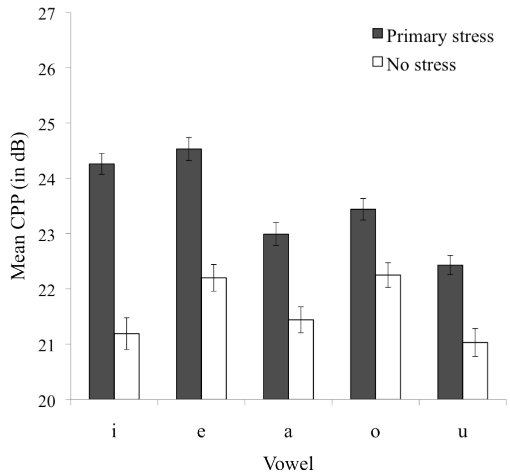
The results for CPP also show a significant main effect for primary stress. Within-vowel comparisons show that all vowels have significantly higher values when they bear primary stress than when they are unstressed. A higher CPP value is an indication of a clearer (i.e. less noisy and more periodic) vowel. At first glance, the CPP values would appear to contradict the H1*-H2* results reported above, which seem to indicate that primary stressed vowels are breathier. These voice quality results will be discussed in more detain in Section 4.1.

There was no significant difference in CPP between vowels with secondary stress (mean = 22.45 dB) and their unstressed counterparts (mean = 22.49 dB), $t = .13$, $p = .89$.

Table 7. Mean CPP (in dB), overall and by vowel, for primary stress. P-values represent differences between stress and no stress.

	Mean primary	Mean no stress	T-value	P-value
Overall	23.52	21.60	14.09	.0001
/i/	24.26	21.19	9.99	.0001
/e/	24.53	22.20	8.23	.0001
/a/	22.99	21.44	6.28	.0001
/o/	23.44	22.25	2.72	.0076
/u/	22.43	21.03	5.09	.0001

Figure 7. Mean CPP by vowel, for primary stress (left panel) and secondary stress (right panel). Errors bars represent standard error of the mean.



Given that H1*-H2* and F0 tend to be positively correlated (Garellek & Keating, 2011; Esposito 2010), it is possible that the higher values of H1*-H2* under primary stress are a result of higher F0 rather than an independent effect. To test this, we re-ran a linear regression predicting H1*-H2* as a function of stress, but this time included F0 as a random effect. The results showed that H1*-H2* was still significantly higher for vowels with primary stress than for unstressed vowels. Thus, even when F0 is taken into account, stressed vowels in Tongan have a higher H1*-H2* than unstressed ones.

4.7 Logistic regression analysis

In the previous sections, we found that several acoustic measures can distinguish stressed from unstressed vowels in Tongan, and this is true for both primary and (to a lesser extent) secondary stress. However, it is still unclear which of these measures best predict primary vs. no stress, and secondary stress vs. no stress. In this section, we answer this question by means of logistic regression.

We first ran a model predicting the presence of primary stress vs. no stress. The logistic regression model had as fixed effects the seven acoustic measures analyzed in the previous section: F0, F1, F2, CPP, H1*-H2*, RMS energy, and duration. The acoustic measures were centered to reduce collinearity among them. The model also included speaker, word, and repetition as random effects. The output of the model indicates which measures are the best predictors of primary stress regardless of speaker or vowel quality. The models were generalized linear mixed-effects logit models, which provide p-values in the summary (Baayen 2008a).

The results of the regression model for primary stress are shown in Table 8. Of the seven measures, only F0 and duration were found to be significant. Both measures have positive estimates, indicating that an increase in F0 or duration is associated with an increase in the likelihood of a stressed vowel, consistent with the results of the previous section. The greater *z*-score of F0 compared to duration indicates that the former is the more important predictor of primary stress. Even though the other measures (except F2) were found to be significantly different for primary stress (relative to no stress) when considered on their own, this logistic regression model indicates that they do not significantly improve performance of the model over and above the effects of F0 and duration.

Table 8. Results of logistic regression model predicting primary stress vs. no stress

	Estimate	Standard Error	*z*-score	*p*-value
Intercept	2.49	2.30	1.08	0.28
F0	0.23	0.03	7.90	<0.0001***
Energy	-0.21	0.17	-1.24	0.22
H1*-H2*	-0.06	0.13	-0.46	0.65
CPP	0.02	0.14	0.16	0.87
F1	<0.01	0.00	1.27	0.21
F2	< -0.01	0.00	-0.15	0.88
Duration	0.09	0.02	3.96	<0.0001***

A similar logistic regression model was fitted to determine the best predictor of secondary stress (vs. no stress). We included in this model the same fixed and random effect structure as for the primary stress model. The results, shown in Table 9, indicate that F0, duration, and energy were found to be significant predictors of the presence of secondary stress. By comparing the *z*-scores, the relative importance of these significant predictors is found to be F0, followed by duration, then by energy. However, recall that for vowel duration there is the potential confound with word position.

Table 9. Results of logistic regression model predicting secondary stress vs. no stress

	Estimate	Standard Error	*z*-score	*p*-value
Intercept	0.27	0.49	0.55	0.58
F0	0.05	0.01	7.45	<0.0001***
Energy	0.15	0.07	1.99	0.05*
H1*-H2*	-0.01	0.03	-0.15	0.88
CPP	0.01	0.03	0.41	0.68
F1	< 0.01	0.00	0.95	0.34
F2	< -0.01	0.00	-0.62	0.53
Duration	-0.02	0.01	-4.48	<0.0001***

Thus, even though other measures beyond F0 can differentiate stressed from unstressed vowels when considered individually, F0 is found to be the most significant predictor of stress, both primary and secondary.

5 Discussion

In this study we find that multiple measures distinguish between stressed and unstressed vowels. In particular, vowels with primary stress are marked by higher F0, lower F1, longer duration, higher energy, higher H1*-H2*, and higher CPP relative to vowels without stress. But even though multiple measures correlate with stress in Tongan, logistic regression finds that F0 was the main predictor of stress. Because F0 is such a robust indicator of stress, the other measures, except for duration, fail to add any significant improvement to the logistic model over the baseline provided by F0. These results are robust across our four speakers.

We also find that a different set of measures correlates with secondary stress than with primary stress in Tongan, similar to Adisasmito-Smith and Cohn (1996)'s findings for Indonesian. In particular, only F0 and duration are found to be consistently different in vowels with secondary stress and unstressed vowels, a subset of those found to be significant cues for primary stress.

4.1 Voice quality and stress

The voice quality results show that H1*-H2* and CPP are generally higher for vowels with primary stress than for unstressed vowels. Higher H1*-H2* values are associated with less laryngealization and/or breathier voice quality, whereas higher CPP values indicate a more modal, more periodic vowel. Thus, when taken together, the two measures indicate that primary stressed vowels in Tongan are more modal than unstressed vowels. Because unstressed vowels have lower H1*-H2* and CPP values, they are assumed to be creakier or less periodic than vowels with primary stress. The aperiodicity in creaky phonation is likely to lower both H1*-H2* and CPP, as seen in other languages like Mazatec, Zapotec, and Yi (Garellek & Keating 2011; Keating et al. 2011). Thus, the inclusion of CPP in this study is important, in that it helps clarify the H1*-H2* results. Normally higher values of H1*-H2* (or its uncorrected counterpart) are associated with increased breathiness (Bickley, 1982, Klatt & Klatt, 1990). H1*-H2* varies along a continuum not clearly delimited in terms of creaky-modal-breathy qualities, so CPP allows for a clearer interpretation of H1*-H2*. As such, the combination of both spectral balance and periodicity measures provides a more accurate picture of voice quality than either of the measures alone.

Our results also indicate that voice quality and stress interact similarly in Tongan to what has been shown in other languages. Campbell & Beckman (1997) also found similar spectral changes in English vowels. They calculated H1-H2 (in their study, H2-H1, uncorrected for vowel formants), and three of the four speakers showed lower values of H2-H1 (thus, higher H1-H2) for accented vowels compared to stressed or unstressed ones. The results in Campbell & Beckman (1997) are consistent with ours, because vowels with primary stress also bore accent, and these showed higher values of H1*-H2*.

Other work on the voice quality of stressed vowels has also shown an increase in higher-frequency energy for stressed vowels in Dutch (Sluijter & van Heuven 1996). Although we did not calculate higher-frequency spectral energy, our finding that stressed vowels have higher values of H1*-H2* does not imply that higher-frequency energy is weaker. H1*-H2* has been found to be correlated with open quotient (OQ), i.e. the portion of the glottal cycle during which the vocal folds are abducted (Holmberg et al., 1995). On the other hand, higher-frequency energy is associated with more abrupt vocal fold closure, which is typical of creaky voice (Hanson et al., 2001). Thus, although during typical breathy voice both H1*-H2* and spectral tilt measures tend to be high (perhaps due to higher OQ and less abrupt closure, respectively), a higher value of H1*-H2* during stress does not necessarily mean a decrease in high-frequency energy.

We hypothesize that the higher values of H1*-H2* under stress could serve as a cue to listeners: A prominent H1 and lower aperiodicity would be perceptually useful for retrieving pitch information, thereby enhancing the higher F0 during stress.

Therefore, this study provides further evidence that stress affects voice quality. The results also show that both harmonic and inharmonic (noise) measures should be used to analyze voice quality (Simpson, 2012), and that closer examination of these effects on various components of the spectrum is warranted.

4.2 Pitch accent vs. word stress

It is possible that the findings of this study are associated with pitch-accented vowels (i.e., phrasal prominence), rather than stress (lexical prominence). We believe that for primary stress, it is effectively impossible to disambiguate between these two levels of prominence in our case. In Tongan, each content word typically forms its own accentual phrase, thereby necessarily bearing a pitch accent, in addition to edge tones (Vicenik & Kuo 2010). Moreover, since focus in Tongan is expressed primarily through syntactic means with no overt prosodic changes (Vicenik & Kuo 2010), we were unable to elicit unaccented content words through post-focus de-accenting. Although this is a limitation in the current study, we expect that the same problem would arise in phonetic studies of stress in other languages that have both lexical stress and obligatory accentual phrase pitch accents, e.g. Farsi (Jun, 2005). Therefore, more work is needed to determine how one could disambiguate stress from accent in languages with prosodic systems of the type found in Tongan.

Although in this study vowels with primary stress always bore a pitch accent, vowels with secondary stress did not. Thus, we can conclude that the higher F0 that is characteristic of vowels with secondary stress is due to lexical stress rather than the presence of a pitch accent. Because secondary stress clearly has an effect on F0, it is likely that part of the difference in F0 seen for primary stress is also due to lexical stress itself and not only to phrasal accent.

4.3 Phonetic targets and the effects of stress on the vowel space

Recall that the Tongan vowel space was neither expanded nor reduced in unstressed vowels relative to vowels with primary stress (Fig. 5). Instead, all five vowels were shifted up in the vowel space (i.e., had lower F1) when unstressed with no change in the overall size of the vowel space. This pattern of results is informative for our understanding of phonetic targets and how they are realized within the context of a stress system. As we describe below, the Tongan results do not fit in with commonly held assumptions about how vowel quality is affected by stress (or lack of stress).

Typically, changes in the vowel space for unstressed vowels are discussed in terms of "undershoot" or vowel reduction. In Lindblom's (1990) "Hyper- and Hypoarticulation" theory, the input to the speech system at the time of production represents an ideal goal that the speaker intends to produce. In certain speech conditions in which the duration of speech sounds is reduced (e.g., casual speech or unstressed vowels), articulatory targets may not be fully reached, resulting in what is commonly called undershoot. Similarly, in the articulatory phonology framework (Browman & Goldstein 1986, 1990), each speech sound is associated with a set of articulatory gestures. In running speech, gestures in close proximity overlap. Under conditions where speech sounds have shorter durations, these gestures have greater overlap due to the temporal compression. As a result, the gestural targets may not be fully realized.

Additional evidence for the principle of an ideal articulatory target, as well as for undershoot in contexts where speech sounds have shorter durations, comes from perception and production experiments. Johnson, Flemming, and Wright (1993) discuss what they call the "hyperspace" effect. Speakers were recorded pronouncing words containing various vowels in casual speech. In a separate task, the same speakers were able to adjust the F1 x F2 dimensions of synthesized vowels by clicking different positions on a screen to hear the resulting vowel. For the same set of vowels that they had previously produced, they were asked to choose the F1 x F2 combination that best matched their own production of each vowel (without getting to hear the recordings of their own productions). The vowels chosen by the speakers in the perception task were more extreme than the vowels actually produced by those speakers in casual speech. In fact, the vowels chosen in the perception task were closer to hyperarticulated speech whereas the vowels produced in causal speech resulted in a reduced vowel space. These results indicate that speakers' targets (and what they believe they are saying) are actually from an expanded vowel space—they represent extreme, hyperarticulated vowels. In casual speech, the vowels are subject to undershoot because the speakers do not reach those hyperarticulated targets, resulting in a smaller vowel space.

Over time, the tendency to reduce the vowel space for unstressed vowels due to articulatory undershoot (a phonetic effect) may lead to phonological patterns of vowel reduction that we see in many of the world's languages (e.g., see Flemming (2005), Barnes (2012)). Crosswhite (2001) discusses two common phonological vowel reductions systems: centralization of the unstressed vowels (e.g., as in English) and merging vowel contrasts (e.g., as in Catalan and Italian). Yet even in languages without phonological vowel reduction, we expect a tendency for a phonetically reduced acoustic vowel space in unstressed vowels (Flemming 2005).

In Tongan, we found that unstressed vowels indeed had shorter durations than vowels with primary stress (see Table 3). Therefore, it is reasonable to expect that unstressed vowel in Tongan should be subject to undershoot and a reduced vowel space (Flemming 2005). However, the particular pattern found in Tongan does not follow from any of the theories discussed above. As the vowel space plot in Fig. 5 shows, the overall size of the Tongan vowel space is not reduced for unstressed vowels as predicted by the hyperspace effect; rather, the size of the vowel space is comparable for stressed and unstressed vowels. Moreover, the vowels are not headed towards either of the two phonological systems of vowel reduction discussed by Crosswhite (2001): the unstressed vowels are not moving towards the center of the vowel space (as in centralization) and they are not moving closer to each other (as in contrast merging). Rather, all of the vowels, including the high vowels, are higher in the vowel space when unstressed. The fact that even high vowels in Tongan have lower F1 when unstressed is important: high vowels are predicted to have roughly the same F1 (or slightly *higher* F1) when unstressed if the effects on the vowel space are solely due to undershoot (see Flemming 2005). Based on these observations, we conclude that the differences in F1 for unstressed vowels are not merely due to Tongan speakers falling short of their targets (undershoot).

However, it is clear that Tongan speakers are not reaching the same targets that they reach for stressed vowels. If they were, there would be no difference in F1 between stressed and unstressed vowels. Thus it appears that Tongan speakers do not have a single target for each vowel, falling short of that target when the vowel is not stressed. Instead, Tongan speakers appear to have a separate target for unstressed vowels involving a somewhat lowered F1.

We propose that the relationship between Tongan stressed and unstressed vowels is not one characterized by undershoot or the hyperspace effect, but rather by a shifted vowel space that retains both its overall size and the relative distance between the vowels within that space.

4.3.1 Considering an alternate explanation: Effects of consonant closure on F1

We briefly consider one alternate explanation for the lower F1 found for unstressed vowels—namely that consonants closures surrounding a vowel can lower the vowel's F1 going into and coming out of the closure (see Johnson 2002). Even though our stressed and unstressed vowels have the same set of consonants surrounding them, the unstressed vowels have shorter durations meaning that the surrounding consonants could affect a greater proportion of the unstressed vowels, resulting in lower mean F1 for all the unstressed vowels. To evaluate this possibility, we took a subset for each of the five vowels in which vowels with primary stress and unstressed vowels had an equal mean and standard deviation for duration (on average 54 tokens, half stressed and half unstressed, were included for each vowel).

If the lower F1 for unstressed vowels were due only to the surrounding consonant closures, the difference should be erased when comparing stressed and unstressed tokens with equal duration. But if the stress itself has some effect on F1, then the difference should remain when duration is controlled. Rerunning the models using just the subsets with an equal mean duration, the difference in F1 remains significant for all vowels except /a/. Thus we conclude that for all vowels except /a/, primary stress results in a higher F1 independent of any effect of duration or the surrounding consonants. These results also provide further support for the conclusion that phonetic undershoot, which depends on the difference in duration between stressed and unstressed vowels, cannot solely explain the lowered F1 for unstressed vowels in Tongan.[2]

4.3.2 Motivation for the shifted vowel space in Tongan

---

[2] Note that phonetic undershoot may be responsible for the F1 differences for the low vowel /a/ because the F1 difference for /a/ did not remain significant when duration was controlled.

One possible motivation for the shifted vowel space is perceptual clarity: enhancing the contrast between stressed and unstressed vowels without sacrificing the vowel quality contrast. Vowel reduction of the type found in English is effective at making stressed vowels very distinct from unstressed vowels, but the distinction between many vowel qualities is lost in unstressed vowels. This reduction strategy, however, would be counterproductive in a language with few distinctive phonemes and relatively simple syllable structure such as Tongan. In such languages, distinctions between different vowel qualities are highly informative even for unstressed vowels because losing those contrasts would result in many merged words. At the same time, stress plays an important role in morphological processes in Tongan, such as the definitive accent (see Anderson & Otsuka 2006). Thus, enhancing the contrast between stressed and unstressed vowels via slightly modified F1 values, without threatening the contrast between vowel qualities, would by hypothesis be perceptually beneficial.

This interpretation also implies that the shift has been phonologized and does not merely result from articulatory factors alone. As a result, we predict that Tongan speakers should be able to use F1 as a cue to stress in a perceptual task with other measures held constant. We leave this prediction for future work. We conclude by noting that Tongan is unlikely to be unique in exhibiting the shifted system of unstressed vowels. As such, these findings underscore the need to examine a wider selection of languages to increase our understanding of how stress may affect the vowel space.

6 Conclusions

The goal of this paper was to determine which acoustic measures correlate with both primary and secondary stress in Tongan. The results from four female speakers indicate that vowels with primary stress are marked by higher F0, lower F1, longer duration, higher energy, and more regular voice quality relative to vowels without stress. Vowels with secondary stress are marked by higher F0 and shorter duration. The logistic regression analysis found that F0 was the main predictor of primary and secondary stress. We found a lowering of F1 for all unstressed vowels, including high vowels, together with no change in F2. This shift in the vowel space with no corresponding change in overall size is inconsistent with an explanation based on phonetic undershoot alone, indicating the Tongan speakers have separate targets for stressed and unstressed vowels.

7 Acknowledgments

We would like to thank Pat Keating, Kie Zuraw, Hilda Koopman, and the members of the UCLA Phonetics Lab for helpful discussion. An earlier version of this work was presented at the 161st meeting of the Acoustical Society of America and the 18th meeting of the Austronesian Formal Linguistics Association.

8 References

Adisasmito-Smith, Niken & Abigail C. Cohn. 1996. Phonetic correlates of primary and secondary stress in Indonesian: A preliminary study. *Working Papers of the Cornell Phonetic Laboratory* 11, Cornell University.

Anderson, Victoria & Yuko 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, Victoria & Yuko Otsuka. 2006. The phonetics and phonology of "Definitive Accent" in Tongan. *Oceanic Linguistics* 45, 25–42.

Baayen, R. Harald. 2008a. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. Harald. 2008b. languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics". R package version 0.953.

Barnes, J. 2012. Phonetics and phonology in Russian unstressed vowel reduction: A study in hyperarticulation. Ms, Boston University.

Bates, Douglas, Martin Maechler & Bin Dai. 2008. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-28. http://lme4.r-forge.r-project.org/.

Bickley, Corine. 1982. Acoustic analysis and perception of breathy vowels. *MIT Speech Communication Working Papers* 1, 73–83.

Blust, Robert. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.

Boersma, Paul & David Weenink. 2009. Praat: Doing phonetics by computer (version 5.1.14). http://www.praat.org/ (30 August 2009).

Browman, Cathe P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology* 3, 219–252.

Browman, Cathe P. & Louis Goldstein. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18, 299–320.

Campbell, Nick & Beckman, Mary E. 1997. Stress, prominence, and spectral tilt. In Antonis Botnis, Georgios Kouroupetroglou & George Carayannis (eds.), *Intonation: Theory, Models and Applications (Proceedings of the ESCA Workshop on Intonation),* 67–70.

Cho, Taehong & Patricia A. Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37, 466–485.

Crosswhite, Katherine. 2001. *Vowel reduction in Optimality Theory*. New York: Routledge.

Esposito, Christina M. 2010. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38, 303–316.

Everett, Keren L. 1998. The acoustic correlates of stress in Pirahã. *The Journal of Amazonian Languages* 1, 104–162.

Feldman, Harry. 1978. Some notes on Tongan phonology. *Oceanic Linguistics* 17, 133–139.

Flemming, Edward. 2005. A phonetically-based model of vowel reduction. Manuscrip, MIT.

Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27, 765–768.

Garellek, Marc. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics* 40, 152–161.
Garellek, Marc & Patricia Keating. 2011. The acoustic consequences of phonation and tone interactions in Mazatec. *Journal of the International Phonetic Association* 41, 185–205.
Gordon, Matthew & Ayla Applebaum. 2010. Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association* 40, 35–58.
Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29, 383–406.
Gordon, Matthew & Latifa Nafi. 2012. Acoustic correlates of stress and pitch accent in Tashlhiyt Berber. *Journal of Phonetics* 40, 706–724.
Hanson, Helen M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101, 466–481.
Hanson, Helen M., Kenneth N. Stevens, Hong-Kwang Jeff Kuo, Marilyn Y. Chen & Janet Slifka. 2001. Towards models of phonation. *Journal of Phonetics* 29, 451–480.
Hillenbrand, James, Ronald A. Cleveland & Robert L. Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37, 769–778.
Holmberg, Eva B., Roger E. Hillman, Joseph S. Perkell, Peter Guiod & Susan L. Goldman. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech, Language, and Hearing Research* 38, 1212–1223.
Iseli, Markus, Yen-Liang Shue & Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustical measures related to the voice source. *Journal of the Acoustical Society of America* 121, 2283–2295.
Johnson, Keith. 2002. *Acoustic and auditory phonetics*. 2nd edition. Oxford: Blackwell.
Johnson, Keith, Edward Flemming & Richard Wright. 1993. The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 69, 505–528.
Jun, Sun-Ah. 2005. Prosodic typology. In S.-A. Jun (ed), *Prosodic Typology: The Phonology of Intonation and* Phrasing, Oxford: Oxford University Press, pp. 430–458.
Kawahara, Hideki, Ikuyo Masuda-Katsuse & Alain 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.
Keating, Patricia, Christina Esposito, Marc Garellek, Sameer ud Dowla Khan & Jianjing Kuang. 2011. Phonation contrasts across languages. *Proceedings of the Seventeenth International Congress of Phonetic Sciences,* Hong Kong, 1046–1049.
Klatt, Dennis & Laura Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87, 820–857.
Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America* 32, 451–454.
Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modeling*, 403–439. Dordrecht: Kluwer.

Ortega-Llebaria, Marta & Pilar Prieto. 2011. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech* 54, 73–97.

Plag, Ingo, Gero Kunter & Mareile Schramm. 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39, 362–374.

R Development Core Team. 2008. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org (14 January 2009).

Shue, Yen-Liang, Patricia Keating, Chad Vicenik & Kristine Yu. Voicesauce: A program for voice analysis. *Proceedings of the Seventeenth International Congress of Phonetic Sciences,* Hong Kong, 1846–1849.

Simpson, Adrian P. 2012. The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics* 40, 477–490.

Sjölander, Kåre. 2004. The Snack Sound Toolkit [Computer program], http://www.speech.kth.se/snack/. Retrieved May 25, 2010.

Sluijter, Agaath M. C. & Vincent J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100, 2471–2485.

Taumoefolau, Melenaite. 2002. Stress in Tongan. *MIT Working Papers in Linguistics* 44, MIT.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28, 397–440.

Vicenik, Chad & Grace Kuo. 2010. Tongan Intonation. Poster presented at the 160th Meeting of the Acoustical Society of America. Cancun, Mexico.

Zuraw, Kie, Kathleen O'Flynn & Kaeli Ward. 2010. Marginal prosodic contrasts in Tongan loans. UCLA Phonology Seminar presentation, June 2, 2010.