This is a pre-publication version of: Evans, J. P., Chu M., Aston J. A. D., & Su C. (2010). Linguistic and human effects on F0 in a tonal dialect of Qiang. *Phonetica*, 67, 1-2.
The final version of this paper is available at http://content.karger.com/produktedb/produkte.asp?doi=319380.

Linguistic and human effects on F0 in a tonal dialect of Qiang

Jonathan P. Evans[a], Man-ni Chu[a,b], John A. D. Aston[c,d], Chao-yu Su[a]

[a]Institute of Linguistics, Academia Sinica
[b]Institute of Linguistics, National Tsinghua University
[c]Institute of Statistical Science, Academia Sinica
[d]CRiSM, Department of Statistics, University of Warwick

While both human and linguistic factors affect F0 in spoken language, capturing the influence of multiple effects and their interactions presents special challenges, especially when there are strict time constraints on the data-gathering process. A lack of speaker literacy can further impede the collection of identical utterances across multiple speakers. This study employs linear mixed effects analysis to elucidate how various effects and their interactions contribute to the production of F0 in Luobuzhai, a tonal dialect of the Qiang language. In addition to the effects of speaker sex and tone, F0 in this language is affected by previous and following tones, sentence type, vowel, position in the phrase, and by numerous combinations of these effects. Under less-than-ideal data collecting conditions, a single experiment was able to yield an extensive model of F0 output in an endangered language of the Himalayas.

1. Introduction.

Qiang is a Tibeto-Burman language spoken by about 110,000 people in northern Sichuan province, China. It is comprised of dialects with lexical tone as well as those that lack lexical pitch distinctions. Tonal and toneless varieties are not mutually intelligible, and it is not clear how many distinct tonal dialects of Qiang exist.[1] In 2006, when the data for this study were gathered, the village of Luobuzhai was reported to be the largest Qiang-speaking community, with about one thousand residents. Eyewitnesses report that the Wenchuan earthquake of May 12, 2008 caused the collapse of nearly all buildings in Luobuzhai, and the deaths of about two hundred villagers.

In Luobuzhai Qiang, High (á) and Low (à) tones can distinguish lexical items. Minimal pairs include /ʁú/ 'back of body', /ʁù/ 'goose', and /tɕhé/ 'narrow', /tɕhè/ 'goat'. A rising tone was reported by Wen & Fu (1943) and Wen (1951), but this appears to serve mainly a morphological function at the end of the verb phrase (as in Mianchi, another tonal dialect of Qiang; cf. Evans 2008). Neither source mentions lexical or phrasal stress, although in Northern Qiang dialects, lexical stress affects vowel quality and devoicing (LaPolla & Huang 2003).

The purpose of this study was to discover as many linguistic and human contributors to F0 production in Luobuzhai as possible, given the limitation that only one day of recording was available. Because the location was not easy to reach, subsequent data collection has not been feasible. Other obstacles added to the challenge – Luobuzhai is not a written language, so all recorded utterances had to be easy for subjects to remember and to repeat. Previous

[1] Liu (1998a) lists seven distinct Southern Qiang varieties, and Sun (1981) lists five. Native speakers claim that some of these varieties are mutually intelligible, contrary to the claims in Liu (1998a) and Sun (1981).

fieldwork had shown that Qiang speakers were not likely to repeat anomalous morpheme combinations. This conservatism ruled out the use of nonsense words or phrases, which are often used in phonetic research to fill out a matrix of minimally differentiated utterances, or to isolate a certain phenomenon. There is no dictionary for Luobuzhai; there is a short wordlist in Wen (1951); however, the words in this source did not correspond well to the words we encountered in situ. Thus, in the time available for this project, it was not possible to obtain a large list of minimally differentiated words, as has been advocated for phonetic studies:

[T]he most effective observation may be made in minimal contrast comparisons. For example, when all the other factors are kept constant while the surrounding tonal contexts are systematically varied. (Xu, 2006)

While the minimal contrast method can yield great descriptive precision, there is a potential weakness to it. Because tones are linguistic entities, they are never uttered unaccompanied, but always with some vowel, initial consonant (or zero), etc. Thus, if a factor is held constant throughout the entire experiment, such as initial /m/ in Xu (1999), the resulting model neither reveals how much of the response is contributed by that factor, nor what the F0 pattern would look like if a different initial consonant were used. Presumably, in that study, the initial consonant was held constant to avoid speaker exhaustion; adding consonantal voicing variation within Xu's experimental design would have doubled each speaker's utterances to 960 tokens.

Numerous linguistic and non-linguistic factors have been demonstrated to affect fundamental frequency (F0) of the vowel portion of a syllable. Linguistic effects may be divided into properties of the syllable and its structure, and properties of larger prosodic units. Syllable-level properties include tone (with its obvious reflection in pitch), stress (Fry, 1958; Gussenhoven, 2004; Kohler, 2008) vowel height (higher vowels have higher intrinsic F0: Esposito, 2002; Ohala & Eukel, 1987; Whalen & Levitt, 1995) and the laryngeal properties of initial and following consonants (voiceless ones raise F0 of an adjacent vowel (Hombert, Ohala, & Ewan, 1979; Mazaudon & Michaud, 2008; Ohde, 1984)).

Larger prosodic units also affect F0. In tonal languages, tones of previous and following syllables can exert assimilatory force, pulling the target syllable's frequency closer to that of a neighbor. In certain circumstances, speakers dissimilate tones, such as the raising of H(igh) before L(ow) in Mandarin (Xu 1999), or the deletion of a H tone that is adjacent to another H tone in Bantu and other languages (Meussen's rule).

Depending on the type of sentence, F0 may be shifted up or down over the whole sentence, or over part of it (Liu & Xu, 2005; Yuan, Shih, & Kochanski, 2002; Zeng, Martin & Boulakia, 2004). In some cases, changing the sentence type or the frame sentence changes phrasal stress properties, further altering F0 (Huss, 1978; Werner & Keller, 1994). Syllable location within a sentence's intonation curve plays a role, as do larger intonation units, so that a paragraph-final syllable may be significantly lower in F0 than a paragraph-internal but sentence-final syllable (Ladd & Silverman, 1984; Sluijter & Terken, 1993; Tseng, 2006a). In this study, we have also hypothesized a word effect, which encodes the notion that there may be differences between the F0 patterns of different lexical items that are not captured by the above-mentioned set of effects. For example, corpus frequency, a factor external to phonology, has been shown to affect F0 in Cantonese (Zhao & Jurafsky, 2009). In addition, certain effects may be difficult to model because they occur too infrequently to have been noticed by researchers; this "sparsity problem" has been described for timing models (van Santen 1994, 1997; van Santen & Shih, 2000). In our case, a lack of extensive prior familiarity with Luobuzhai meant that before the experiment we were not aware of particular language-specific F0 effects, other than tone.

In addition to linguistic properties, there are properties of the speakers themselves that affect F0 production. Age, sex, and health condition all contribute to F0; for example, females have been found to demonstrate a significant decrease in F0 with increasing age (Sulter, Schutte, & Miller, 1996), while for males, the F0 fall that begins in puberty lasts until about age 35, and then begins rising again at about 55 years (Traunmüller & Eriksson, 1995). Speakers also have their own idiosyncratic pitch means and ranges. Table 1 lists the factors that were examined in this experiment, along with their expected effects on F0:

Table 1. Factors potentially affecting F0.

Factors	**Expected effects**
Tone.	Direct F0 reflection of tone category.
Lexical stress.	Broader F0 range in some tone languages.
Vowel height.	Higher vowels have higher intrinsic F0.
Consonant voicing	Voiceless consonants raise F0 on the adjacent edge of the neighboring vowel.
Previous/following tone.	Assimilatory or dissimilatory shifting is possible.
Location of syllable within word.	Downtrend is expected.
Sex of speaker.	Men typically have lower F0 than women.
Age of speaker.	Decrease in F0 with age, at least for females.
Sentence condition – statement, question, etc.	F0 may be shifted up/down over the whole sentence, or over part of it.
Word.	Allows F0 to vary in ways not yet understood.
Speaker idiosyncrasies.	F0 affected by customary pitch range, weight of vocal cords, health condition,…

The result of the experiment was to elucidate the effects on vowel F0 of tone, speaker sex, neighboring syllables, sentence type, phrasal declination, and vowel, as well as various interactions among these effects. Consonant voicing, speaker age, and relative intensity (taken as a surrogate for stress) were not found to be significant. The status of lexical stress in this language is not yet fully understood, and would require studies that lie beyond the scope of this experiment.

2. Material and methods.

2.1. Material.

Data for this study were recorded on location in August, 2006. A list of nineteen nouns exemplifying the range of tonal and segmental variation was selected with the help of a primary consultant; the list is given in Appendix A. An attempt was made to find nouns whose F0 properties covered the widest possible range, and could fit within the frame sentence, "I'm thinking of ___." Southern Qiang is not a written language; all example words were discussed in Chinese and in Qiang before being recorded. Because of an oral, rather than literate, culture, speakers had to find forms acceptable before they would say them; semantic anomalies which fit the tonal patterns being sought were rejected by the consultants and were not recorded.

We used the following exchange between speakers to elicit unmarked statements, yes/no questions, and contrastive statements:

(1) Frame sentences.

a: qà _ bè.lù. nə̀ _ bè.lù-nə́-má?
1sg _ think 2sg _ think-2sg-Q
"I want a(n) __. Do you want a(n) __?" or
"I am thinking about (a) __. Are you thinking about (a) __?"

b: mì-ʁú, qà __ bè-lù.
NEG-COP 1sg __ think
"No, I want/am thinking about (a) __."

A recording of this series of exchanges yielded the clause, "want/think about (a) __," within an unmarked declarative (1.a), a question (1.a'), and under contrastive focus (1.b). In all of the sentences, a L toned syllable precedes and follows the word in the frame. Due to the fact that the language consultants were aware of the word that was of interest in each case, it appeared to receive slight phrasal prominence; the degree of prominence was much smaller than is heard in English emphatic contrastive utterances when near homophones are being distinguished, as in, "I said CHICKen, not KITCHen!".

2.2 Speakers and recordings.

Four male native speakers (ranging from 34 to 65 years old) and four female speakers (31 to 62 y.o.) gathered for an elicitation session in the home of one of the speakers. All of the speakers lived in Luobuzhai village at the time of recording and used Luobuzhai Qiang as their most frequent mode of communication.

Speakers were trained to hold a microphone (Sony ECM-77B) in such a way as to yield optimal recording quality. By placing the back of the microphone-holding thumb against the point of the chin, very clear recordings with minimal aspiration noise were made, in spite of recording in a social environment. Data were recorded directly onto a Macintosh iBook computer, via USBPre pre-amplifier (M-Audio corporation); adjustable pre-amplification was required, due to differences in speaker intensities. Speech was recorded via Praat software (praat.org) at 22 kHz sampling rate, the rate recommended by Ladefoged (2003, p.95) for identifying characteristics of both vowels and consonants. Speakers were asked to use their own normal comfortable speaking styles; they appeared to maintain relatively constant and natural speaking rates and pitch ranges. Sound files were checked during the session, and unclear utterances were re-recorded.

2.3 Acoustic measures.

Pitch contours on vowels were identified via Praat's autocorrelation algorithm (Boersma, 1993). Syllable nuclei were sampled at 11 equidistant points, using a Praat script based on one written by Douglas Honorof. The script sampled F0 values at the beginning of the vowel, at intervals of 10% duration, and at the end of the vowel. For each syllable, the median F0 value of these 11 points was selected for use in the analysis. The resulting databases were reorganized, in order to prepare them for statistical analysis by the R software package (R Development Core Team (Designers), 2009). The eight speakers' production of F0 over the quadrisyllable /'tshà tʂ̞ú qù qù/ 'storage room door' can be visually represented as follows (syllables have been time-normalized to aid presentation):

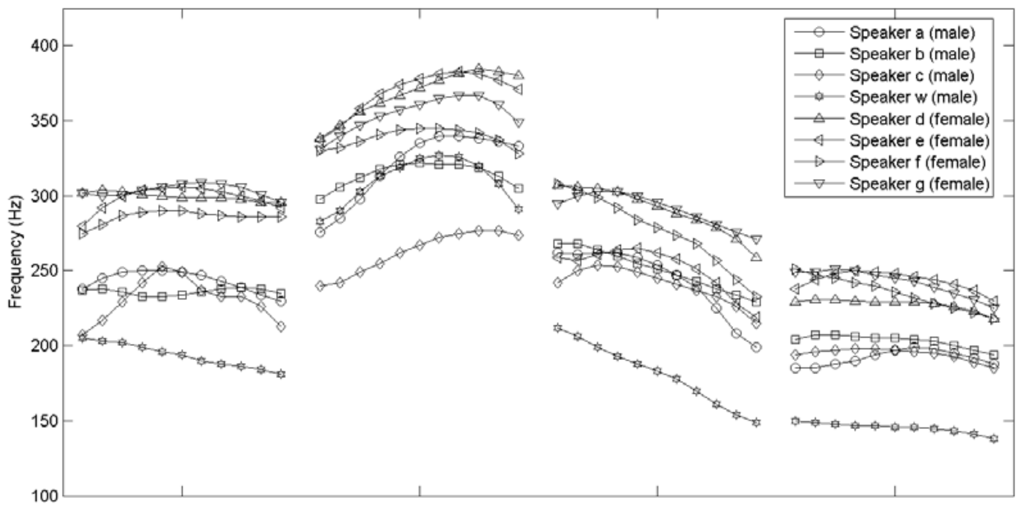

Fig. 1. Time-normalized vowel F0 plot for each speaker's pronunciation of /tshà tʂ̞ú qù qù/, 'storage room door'.

The analysis used Hertz, rather than a transform of frequency (Bark scale, semi-tones, etc.), because male and female speakers' standard deviations of frequency were not statistically distinguishable, and thus variance transformation was unnecessary. The data also appeared to follow a normal distribution of error (having checked statistical diagnostics on the final model – data not shown), one of the assumptions needed for the following statistical analysis.

2.4 Analytical methods.

On the analytical side, it was decided to use linear mixed effects analysis (LME), because of the ability of this method to extract the contribution of a large number of effects and their interactions, and also its ability to model the inherent random aspects of the data, along with the usual fixed effects in multiple linear regression; cf. Henderson (1953), Cunningham & Henderson (1968), Laird & Ware (1982). There are many studies in the literature that combine multiple coefficients to estimate acoustic effects, including multiplicative models (e.g., of timing; Shih & Ao, 1997).

Previous linguistic research has employed LME; cf. Baayen (2008) and Baayen, Davidson, & Bates (2008) on lexical reaction times and Johnson (2008) on sentence acceptability judgments. In phonetic research, Nguyen & Fagyal (2008) use linear mixed effects analysis to study speaker variation in French vowel harmony, Myers & Tsay (2008) use LME to verify the categoricality of tone changes in Taiwanese Southern Min, and Plag (2006) employs LME to elucidate English stress pattern choices. Myers & Tsay (2008) discuss those features of LME that set it apart from other mixed effects analytical tools. While in the aforementioned studies, LME is used primarily as a test procedure for a particular effect of interest, here we were also interested in constructing a more comprehensive model from hypothesized effects in the literature.

After considering the various effects on linguistic F0 that are found in the literature, the factors in Table 1 were selected for investigation, with intensity peak location serving as a surrogate for stress. In addition to the effects mentioned above, the following interactions between effects were also investigated (a*b indicates the interaction between effects a and b):

*Tone*word_position*. The tone category (H, L) may interact with its position in the word (syllable number); integer values range from 1 to 4. This interaction is suspected, due to the observation that High syllables typically decline more rapidly over a prosodic domain than do Low syllables; i.e., pitch range compression, cf. Myers (1999), Shih (2000), Yip (2002).

*Previous*initial_voice*. This interaction codifies a cross-linguistic trend that the influence of the previous tone category can be transmitted though a voiced syllable-initial consonant more effectively than through a voiceless consonant.

*Previous_tone*target_tone*next_tone*. Assuming that these three effects may interact allows for the most assimilatory or dissimilatory effects between the target syllable and the preceding and following tones or word boundaries. Here and elsewhere, the algorithm also examines subsets of interactions, such as *previous_tone*next_tone*.

*Gender*target_tone*condition*. This interaction allows for females and males to differ in the way that sentence conditions (statement, question, contrast) interact with tone in the production of F0. For example, contrastive emphasis could lead to a greater divergence between H and L for one sex but not the other. It also allows for males and females to differ in their F0 spread between tones. For the Northern Qiang dialect of Ronghong, impressionistic studies indicated that speakers do not employ contrastive emphasis, a claim that this study was designed to investigate (cf. LaPolla & Huang, 2003).

*Target_tone*intensity_peak*condition.* Sentence condition has been shown to shift H and L tones differently (Xu, 1999). This combination tests whether this interaction is present, and whether *intensity_peak* interacts with tone and condition in a complex way.

*Age*speaker_sex*. This interaction allows for male and female voices to respond differently to the aging process, as mentioned in section 1.

*Gender*word_position*. Since female voices are typically higher than males, they could tolerate a more rapid slope of declination.
The abovementioned effects are *controlled* – that is, the researcher stipulates which word is to be said, how many male/female speakers are recorded, etc. However, other possible effects are *measured,* such as formants. Because we are estimating a measurable value, F0, it

is easier to base estimations on controlled factors rather than measured values, as in the latter case, the effect of measurement (and errors introduced thereby) must be accounted for. For this reason, instead of using a numerical value for F1 to mark vowel height, which has been shown to affect F0, we use vowel phoneme labels. Intensity peak assignment was based on visual observation of the Praat intensity curve; where speakers differed, 75% agreement (six of eight speakers) was considered adequate to assign this peak. Three polysyllabic words did not have clear intensity patterns; no syllables in these words were marked for intensity peak. In order to more closely correspond to a lexical stress analysis, monosyllables were treated as containing an intensity peak; cf. Appendix A. There were no clear secondary intensity peaks, so only primary peaks are marked. This study does not claim that the syllables with the greatest intensity carry lexical stress; rather the experiment investigates whether there is any correlation between intensity and F0 in these data.

In addition to these fixed effects and their combinations, *speaker* and *word* were examined as random effects. Random effects are drawn from a larger population, and cannot be exhaustively sampled. They are presumed to come from a distribution throughout the population and from which the sample in the experiment is taken. In this study, speaker and word are both representative samples of large populations. In studies that do not acknowledge the random nature of the speaker, various techniques have been employed to reduce this effect. Researchers may attempt to reduce the random speaker effect by using a large number of subjects, a design which may then be taken to represent the whole population. Studies may also restrict the study to a selective sub-population, such as only male speakers, or those within a certain age range. Neither of these solutions truly accounts for the random effect of the speaker; both methods are unable to explicitly state how random effects in the sample can be generalized to the population. While a large number of subjects will always be better for estimation, it is also preferable for the random nature of the sample to be accounted for. The question may be raised as to the validity of extending the results of this model to a new group of Luobuzhai speakers. This study does not claim that adding more speakers would leave the model unchanged. Nevertheless, if it is assumed that the speakers have a normal distribution around a fixed point, and if the speakers have been chosen randomly, then it may be expected that another group of speakers would cluster around the same value. This extension to new speakers is accounted for by taking speaker as a random effect.

The question of how the random word effect would be generalized to new words follows the same reasoning as given for speakers. Because speaker and word are random effects, if they are omitted from the model, then the estimation of the remaining fixed effect coefficients would be less accurate.

One advantage of the LME approach is that it does not require a completely balanced design, where all combinations of effects have an equal number of observations (although this is preferable in terms of estimating parameters if possible). A balanced design can prove difficult to achieve when there are higher order interactions present that need to be estimated using actual (not fabricated) words with set tonal patterns. While in this study most main effects had approximately the same number of observations (except for the following tone category as the word boundary was designated a low tone as mentioned above, yielding approximately twice the number of following L as following H), the range for interactions terms was greater. Taking the most extreme example, HLH and HHH had 24 observations to estimate their effects, while LLL had 236 observations. The HPD confidence intervals in Table 2 allow one to gauge the uncertainty in the parameter estimates, which is affected in part by the number of relevant samples.

It could be argued that the use of the linearity assumption across all variables is not justifiable over a long time range. For example, if syllables decline at a regular rate, then speaker pitch is projected to go below 0 Hz after ten to twenty syllables. However, over the short range used in this study (up to four syllables), there was no obvious benefit to using log(F0) or other transformation. For longer utterances, Shih (2000) has shown that asymptotic declination gives a more accurate estimation of Mandarin Chinese F0. However, over the short ranges used in this study, the deviation from linearity of an asymptotic declination would likely not be noticeable.

It is also to be expected that at the end of the intonation group, F0 is reset to a higher, starting level (Ladd, 1988; Tseng, 2006b). However, the assumption of linearity appears reasonable over the range of data examined here (up to four phrase-internal syllables). For these two reasons (asymptoticity and F0 reset), it should be noted that the declination rate estimated here should only be considered applicable to the portion of the frame sentences that was sampled.

The study selected median F0 as the measurement upon which the model is constructed, because of the resistance of the median to extreme values. Coincidentally, most of the median values were at or near the vowel midpoint, which has been selected in numerous other studies as a summary statistic.

The model produced by this study builds its estimation on just one feature of the data for each syllable, namely median F0. However, speakers produce and listeners perceive information along the entire duration of the syllable. Thus, an LME analysis employs a static, point-like model to describe continuous data. This concern is especially important in describing languages with phonemic contour tones, in which there is an intentional and meaningful F0 trajectory over the course of the tone-bearing unit. If one employed a *time series LME* model (which is substantially more complicated due to the correlation in the measurements across time), akin to those developed for econometrics, then the resulting model would be able to show changing effects, such as the degree to which a previous tone has a greater effect on the left edge of a syllable than on the right edge, etc. This particular method is explored in Aston, Chiou, & Evans (2010); implementation of a continuous model is much more statistically and computationally complex than the LME model. The components responsible for 97% of the variance in the Aston, et al. (2010) model are essentially the same as those found significant in the single-value LME model presented here. Thus, the model that is simpler to implement yields much the same information as the more complicated model, and is easier to interpret.

In spite of the limitations of basing a model on individual values, it should be noted that perception studies have shown that a single value can be a strong indicator of tone category in Cantonese, a language with three level and three contour tones on open syllables (Gandour, 1981; Khouw & Ciocca, 2007). Khouw & Ciocca found that mean F0 yielded a

strong correspondence between the tones that speakers pronounced and the tones that listeners perceived, even when phonemic contour tones were presented.
The statistical approach of top-down modeling (or backwards elimination) was used. This method of arriving at a final model starts with the largest, most complex reasonable model for which the amount of data is sufficient to yield an analysis. For a single effect or interaction, a small t value ($-2 \leq t \leq 2$) indicates that the effect/interaction contributes to the model in an insignificant way, and is a candidate for removal from the model; the smallest such effect or interaction is removed, and the resulting smaller model is run, followed by a parametric bootstrap test (Faraway, 2006) undertaken to verify the insignificance of the removed effect or interaction. The process is repeated until the effect with the smallest contribution to the model exceeds the t value limits, or the bootstrap fails to verify insignificance. We note that other methods of arriving at a final model, such as bottom-up modeling (starting with a single variable and adding complexity) could have been used, and would likely have resulted in a different final model. However, bottom-up modeling would still require the definition of the most complex model deemed reasonable by the investigators (as in our case) although it would never be examined. While we have employed a top-down approach due to the most complex model, representing the most complete set of effects and interactions that could affect the data, being formulated from looking at prior literature and adding possible effects, other approaches could well have been used. For a more complete description of variable selection methods, including backwards elimination, cf. Draper & Smith (1998).

Our goal was not necessarily the sparsest model, but rather the most descriptive one, subject to not having inconsequential effects. Hence, we did not use a penalized model selection technique, where a cost is associated with each term in the model (e.g., AIC, BIC, etc.). Similarly, overfitting is always a concern; nevertheless, because all of the included effects have elsewhere been demonstrated to influence speaker F0, they have been kept in the final model. Cross-validation of the final model was desired, but with only eight speakers sampled, it would not have yielded a reliable analysis. The analysis was performed using the R software package and the *lme4* (Bates & Maechler, 2009), *MASS* (Venables & Ripley, 2002) and *LanguageR* (Baayen, 2007) libraries of functions. For more on the implementation of LME in R, cf. Baayen (2008); Baayen, Davidson, & Bates (2008); Bates & Pinheiro (1998); Faraway (2006); Pinheiro & Bates (2000).

This study shares some similarities with the general superpositional model of van Santen (Mishra, van Santen, & Klabbers, 2006; van Santen, Mishra, & Klabbers, 2003), in that the magnitude of effects is not predetermined by the investigator, but is rather determined by an unbiased analysis of the data.

3. Results.

3.1 Overview of the final model.

Table 2 shows the contributions of the fixed effects and effect interactions in the final model; effects are presented in the order of the following discussion.

Table 2. Fixed effects that alter F0 in Luobuzhai.

	Estimate (Hz)	S.E.	t value	HPD 95% Lower	HPD 95% Upper
Baseline	382	9	41	368	397
Tone = L	-69	4	-18	-77	-62
Sex = Male	-88	12	-7	-106	-69
Syllable number	-24	2	-15	-27	-21
Tone = L, Syllable num.	6	3	2	0	11
Male, Syllable num.	4	1	5	3	6
Prev. T = H	14	5	3	5	22
Prev. T = L	4	4	1	-4	11
Next T = L	28	5	6	18	36
Tone = L, Next T = L	-47	6	-9	-57	-36
Prev. T = H, Tone = L	-14	8	-2	-29	1
Prev. T = H, Next T = L	-15	6	-3	-27	-3
Prev. T = L, Tone = L	-5	7	-1	-17	8
Prev. T = L, Next T = L	-7	6	-1	-18	5
Prev. T = H, Tone = L, Next T = L	45	9	5	26	61
Prev. T = L, Tone = L, Next T = L	25	7	3	10	38
Question	-44	2	-23	-48	-41
Contrastive sentence	-26	2	-13	-30	-22
Male, Question	23	2	11	19	28
Male, Contrastive	5	2	2	1	9
Tone = L, Contrastive	3	2	1	-1	7
Tone = L, Question	6	2	3	2	10
Vowel = a	3	2	1	-3	7
Vowel = e	12	3	4	6	17
Vowel = i	18	3	6	12	23
Vowel = u	15	3	5	9	21

By starting with the baseline (i.e., intercept) F0 and adding the relevant effect values, the response F0 is generated. The following figure shows measured and estimated (male and female) F0 for /dʐə́.ɕí/ 'day before yesterday', uttered in a declarative statement. The initial syllable of this word represents the baseline syllable chosen by R for this dataset: a high-toned word-initial syllable, followed by another high syllable, uttered in a declarative statement by a female speaker. To get the actual estimated median F0 of the baseline syllable, the syllable effect of 24 Hz must be subtracted.

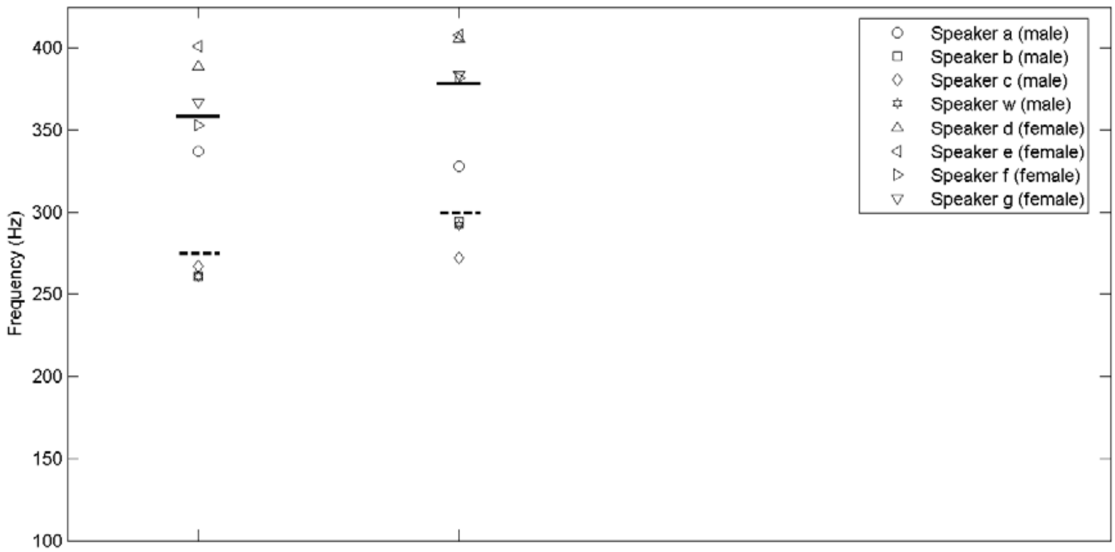

Fig. 2. Median F0 values of the vowel portions of /dʐ̩ə́.ɕí/ 'day before yesterday', with estimated male (dashed line) and female (solid line) F0 medians.

The scaling of this figure has been kept the same as that of Figure 1 in order to highlight the fact that the baseline F0 values are at the top of speakers’ spoken ranges.

Random effects add variance to a model, rather than moving the response variable up or down in the way that fixed effects do. The random effect *speaker* adds a standard deviation of 17 Hz to the response, while *word* adds 9 Hz. As the random effects are assumed to be normally distributed, this means that approximately 95% of the population should be within ±34 Hz (2 standard deviations) of the fixed effects component of the model, whereas 95% of randomly chosen words will be within ±18 Hz of the fixed effects component of the model.

The following sections present the findings relevant to tones, declination, and gender (3.2); tone-tone interactions (3.3); sentence type and gender (3.4); vowels (3.5); the random effects speaker and word (3.6); and effects missing from the final model (3.7).

3.2 Tones, speaker sex, and declination.

The final model allows separation of the influences of tone, speaker sex, and declination. No linguistic effect can be uttered in isolation, as other values must also be specified; e.g., tones are always uttered on a particular vowel, with preceding and/or following syllables, by a male or female speaker, etc. Nevertheless, the model does give values for individual effects.

Figure 3 shows the estimated values for /tshà tʂ̩ú qù qù/, 'storage room door', the F0 traces of which appear in Figure 1. This figure shows the way that the model combines tones, speaker sex, and declination.

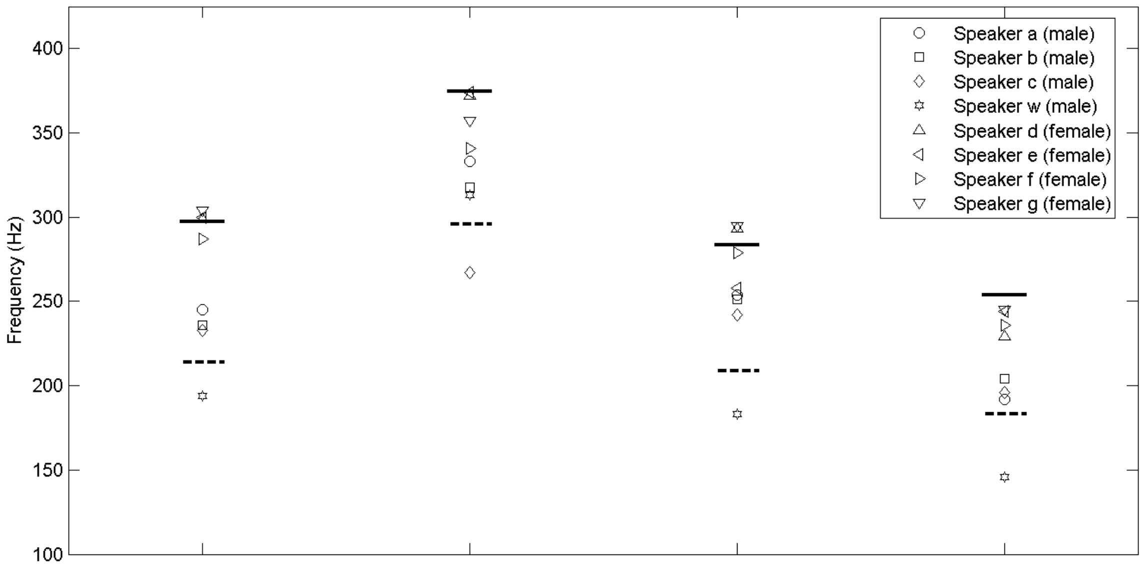

Fig. 3. Female (solid) and male (dashed) estimation of median F0 for /tshà tʂ̩ú qù qù/, 'storage room door'.

The model shows a difference of 69 Hz between H and L tones (shown in Table 2 as Tone = L). A distance of 69 Hz is similar to that found at the midpoint of Cantonese H (55) and L (21) tones (Khouw & Ciocca, 2007).

Males and females differ by 88 Hz in their H tones, which can be seen in the second syllable of Figure 3. Thus, for the baseline utterance of H toned schwa followed by H, the model shows women producing an F0 of 382 – 24 = 358 Hz, and men 294 – 24 = 270 Hz. For females this is at the higher end of the range of 140 to 400 Hz found by Boothroyd (1986); for males, it is substantially higher than Boothroyd's finding of 70 to 200 Hz. For L tones, the baseline values for females and males are 289 Hz and 201 Hz, respectively.

High tones decline by 24 Hz per syllable, starting with the first syllable in the word (in Table 2, this is the syllable number). For declination slope, H must be specified, because of the interaction between L and syllable number, which adds back 6 Hz for L tones, yielding a L tone declination rate of 18 Hz per syllable, which can be seen in the third and fourth syllables of Figure 3. That is, High tones decline more rapidly than Low tones (pitch range compression), which is not surprising, given that Low tones start out 69 Hz lower than High, and are thus closer to a speaker's minimum F0. We specify this as the short phrase declination rate, because our data do not permit extending the generalization to sequences longer than four syllables. There is a very small gender effect on declination rate: maleness reduces declination by 4 Hz per syllable to 20 Hz for H and 14 Hz for L. While this is a small effect on each syllable, by the end of a quadrisyllabic word, it has added up to 16 Hz.

Without taking into account other effects, by the fourth syllable of a word, females have changed H tone F0 by (4 x -24 = -96 Hz), which puts them into the male range for initial syllables (88 Hz lower than female). The model quantifies downtrend in Luobuzhai, as well as showing how it interacts with both gender and tone.

3.3 Tone-tone interactions.

Adjacent tones interact in ways that are not purely assimilatory. As an example, the effects of adjacent tones on F0 output of a L-toned second syllable may be seen in Table 3, where the syllable being modeled is underlined. *Previous tone* and *following tone* have the levels H, L. Note that *previous tone* and *following tone* have different sets of values; the original values for both were H, L, # (word boundary). However, there were not enough data to probe # for the *following tone* condition, so it was replaced by L, the tone of the subsequent syllable in the frame sentences. Preceding word boundary is specified in the baseline values, so it is not explicitly referenced in Table 2. Values in the table are departures from L, which is already displaced -69 Hz from the baseline; syllable number is not given, because all examples are the second syllable of the word in the frame.

Table 3. Effects of adjacent tones on a L-toned second syllable.

Effect	HLL	LLL	HLH	LLH
Prev. T = H	14		14	
Prev. T = L		4		4
Next T = L	28	28		
Tone = L, Next T = L	-47	-47		
Prev. T = H, Tone = L	-14		-14	
Prev. T = H, Next T = L	-15			
Prev. T = L, Tone = L		-5		-5
Prev. T = L, Next T = L		-7		
Prev. T = H, Tone = L, Next T =L	45			
Prev. T = L, Tone = L, Next T = L		25		
total displacement (Hz)	11	-2	0	-1

The values in the last row of Table 3 demonstrate that the fundamental frequency of L tone is raised 11 Hz by preceding H when L follows. However, the other three environments are not distinguished from one another. Because following L and word boundary are both indicated by L, the model estimates the same values for HLL as for HL. In Figure 3 the third syllable L is rather close in F0 to the first syllable L; the declination effect has been somewhat offset by raising in the HLL environment.

Interactions involving H tones show similar, although not identical behavior. Table 4 indicates that H is raised most by preceding H (14 Hz) and by following L (28 Hz); note that Table 4 is shorter than Table 3 because of the structure of the baseline.

Table 4. Effects of adjacent tones on a H-toned second syllable.

Effect	HHL	LHL	HHH	LHH
Prev. T = H	14 Hz		14	
Prev. T = L		4		4
Next T = L	28	28		
total displacement (Hz)	42	32	14	4

Tables 3 and 4 suggest that H tone may be more phonologically active than L. This is evidenced by the negligible impact of L on L (Table 3: LLL, LLH columns) versus the perseveratory influence of H on both H and L tones (Table 3: HLL; Table IV). In addition, second syllable H ranges over 38 Hz, which is a more expressive range than the 13 Hz range for second syllable L. Taken together, these tonal behaviors suggest that the contrast between H and L may be a contrast between H and zero at the phonological level. Tonal underspecification, or privativity, is common among Bantu languages, and has been observed in the Mianchi dialect of Southern Qiang (Hyman, 2001; Evans, 2008). Further phonological research into the tonal processes of this language would be needed to verify privativity of tone in Luobuzhai.

It has been argued that tones in Southern Qiang are a development subsequent to the split between Northern and Southern Qiang dialect groups (Evans, 2001; Liu, 1998b). Across Southern Qiang varieties, it is common for L and H to be the phonologically basic tones, with other tones occurring in highly restricted contexts, such as in borrowings from local varieties of Sichuanese Mandarin. In addition, it is common to find so-called accentual behavior of tone, such as culminativity of H (Mianchi dialect) and obligatory H (Jiuzi/Muka dialect) (Evans, 2001, 2006, 2008). Neither restriction is observed in Luobuzhai. In addition, for Qiang dialects that allow more than one H syllable per word (e.g., Jiuzi/Muka, Longxi), Evans (2001) observed that there was never more than one peak (i.e., *HLH). However, this restriction is also not found in Luobuzhai.

3.4 Sentence type and speaker sex.

The effects of sentence condition (statement vs. contrast vs. yes/no question) are also significant, along with interactions between condition and speaker gender and between condition and tone. The following table shows the displacements from statement F0 patterns for men and women:

Table 5. Interactions between gender and sentence type.

	Question	Contrast
Female	-44 Hz	-26
Male	-21	-19

Contrasts lower F0 in female speech in this sentence position by 26 Hz, relative to unmarked statements (23 Hz for L tones); this could be an indication of the difference made by contrastive phrasal stress, which is an object for further study. Questions lower female F0 in this position by 44 Hz (38 Hz for L tones); Thus, for females, object NP’s are separated by about 15 to 25 Hz into categories of statement, question, and contrastive statement. It is not yet known to what extent these shifts are due to contrastive phrasal stress, or to boundary tones at the end of yes/no questions.

In this sentence position in contrasts and questions, males add back 5 Hz and 23 Hz, respectively. That is, relative to unmarked statements, males lower H tone F0 in contrasts by (-26 + 5 = -19 Hz) and in questions by (-44 + 23 = -21 Hz), a difference of just 2 Hz and not distinguishable; F0 of male L tone syllables in these contexts is similarly not distinguishable. The conclusion may be drawn that females mark object position with different F0 values in unmarked statements, contrastive statements, and questions, while male speakers do not distinguish contrasts and questions in this position; see Figure 4. The direction of change (lower F0 for object position in questions and contrastive statements) did not follow our expectations, and is not to be generalized to the sentence level, without further research into male and female intonation patterns.

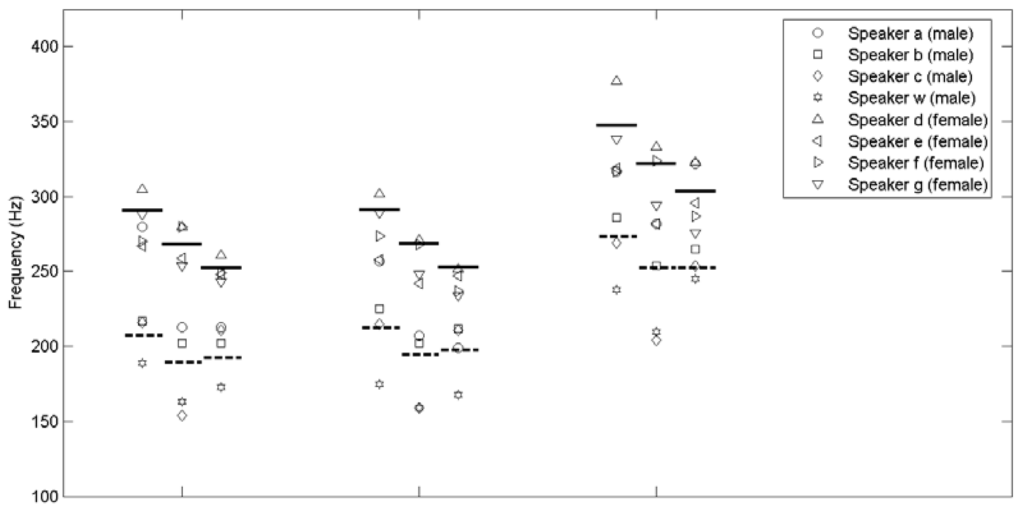

Fig. 4. Male (dashed) and female (solid) sentence condition effects on /dʐ̨ù dʐ̨ù gé/ 'ruler'. For each syllable position, the forms are given in the sequence statement, contrast, question.

In Figure 4, the fact that men and women separate statement, contrast, and question into different numbers of categories can be seen from the estimated values (dashed and solid lines), and also from the measurements themselves.

3.5 Vowels.

The vowel effects /a/ (3 Hz), /e/ (12 Hz), /i/ (18 Hz), /u/ (15 Hz) reflect divergence from /ə/, the unmarked vowel. These values are in accord with the 4 to 25 Hz effect reported in Ohala & Eukel's (1987) survey of about ten languages, and with the 12 to 15 Hz effect found in Whalen & Levitt’s (1995) survey of 31 languages.

3.6 Speaker and word.

In the final Luobuzhai model, speakers alter F0 by a standard deviation of 17 Hz. For the eight speakers who participated in this study, the estimated speaker effects in Hz are: W (m.) -15, A (m.) 24, B (m.) 12, C (m.) -20, D (f.) 12, E (f.) 5, F (f.) -9, and G (f.) -8 Hz. It is not known why the male speaker effects are greater than the female speaker effects. However, a random (instead of fixed) component for gender, when also included in the model, was not found to be significant. This suggests that the range of variation in Hz does not vary between the men and women, as was also found in the preliminary analysis. If the individual speaker effect for a given speaker is added to the model, estimation for that speaker is improved.

The model also shows a word effect with s.d. of 9 Hz; that is, there are some prosodic properties of lexical items that are not accounted for by the fixed effects in the model. These could reflect effects that were not anticipated when the experiment was designed. One such effect would be lexical frequency, how often a given word is used in discourse. This variable has been shown to affect F0 in Cantonese (Zhao & Jurafsky, 2009) and Mandarin (cf. discussion of *jīntiān* 'today' in Shih, 2000). For poorly documented languages, data on how often a particular word occurs is not available, hence cannot be directly modeled. Estimates of the word effects of the lexical items elicited in this study are included in Appendix A. The word effect does not codify the same thing as any of the fixed effects, because effects are specified at the syllable level, while all the syllables in a word share the same word effect.

There is also a residual effect with s.d. of 16 Hz: this is the amount of variation that is not explained by the model, and reflects the inability to perfectly describe speaker behavior. This range of error is comparable to the 13 Hz RMS error found in a model of a single speaker's read Mandarin speech (Kochanski, et al., 2003). Other random effects (including a random effect for speaker gender) were considered, but these were found to be insignificant in the model. In addition, diagnostics were performed on the residuals and no evidence for deviation from the linear mixed effects model was found.

3.7 Effects missing from the final model.

One omission from the final model is that the voicing status of the initial consonant is not a significant effect on median F0. It is probable that the voicing of the initial alters only a short portion of the vowel, such that the median F0 value is not affected. Hombert (1975) & Hombert, Ohala & Ewen (1979) show that in English (which lacks lexical tones), F0 perturbations caused by qualities of the initial consonant last more than 100 ms into the vowel, while in the tonal languages Yoruba and Thai (Gandour, 1974), this effect lasts only 40 to 60 ms and 30 to 50 ms respectively. The explanation given by Hombert et al. is that, "There may be a tendency in tone languages (which does not exist in non-tonal languages) actively to minimize the intrinsic F0-perturbing effect of prevocalic consonants – probably so that the different tones will be maximally distinct perceptually." Our model shows that any perturbation of F0 due to voicing in Luobuzhai was not statistically detectable in median F0.

Whether or not a syllable contained the word's peak of intensity also did not significantly move F0. If this peak does indeed correspond to the location of lexical stress, then for this language, F0 is not a reliable indicator of stress. It should be noted that intensity peaks were not obvious at the time of utterance, although they were clear in the Praat tracks, and that no other prototypical reflections of stress were noticeably present (duration, intensity, vowel quality changes, etc.). In addition to consonantal voicing and intensity, age was also not found to affect F0 production, either for men or women. Given that change in adult F0 with age has been observed in other populations, and that such a change is probably due to physical changes in the vocal apparatus, it may be that a larger sample with a more carefully designed age distribution would quantify an age effect in Luobuzhai.

4. Discussion.

Luobuzhai Qiang vowel F0 is affected by tone, speaker sex, neighboring tones, sentence type, phrasal declination, and by the vowel itself. In addition to these main fixed effects, certain interactions among them also affect F0: tone and declination (L syllables decline more slowly than H), speaker sex and declination (females decline more rapidly than males), tonal combinations (preceding and following H and L interact in complex ways with the target syllable), and speaker sex and type of sentence (females make three distinctions where males make two). Consonant voicing, speaker age, and relative intensity (taken as a possible surrogate for stress) were not found to be significant.

In the estimation of F0 of a population of speakers, it is necessary to consider the speaker and the word being spoken as random effects; otherwise, there is no statistically principled way to draw conclusions from the model about the population of speakers or words in general. Linear Mixed Effects analysis calculates the variance components that show how these variables affect F0 estimation. The model quantifies how much variance is contributed by speakers and by lexical items. Once known, the F0 contribution of a particular speaker or word can be treated as a fixed effect, leading to improved estimation.

Figures 2 through 4 show that estimation does not have to lie within the data, it can be above or below it. This deviation indicates that the values in the figures are not just summaries of the relevant measured values, but are estimated from the model.

While not a phonological study, the model suggests that H has more of an influence on neighboring, especially following, syllables. Further phonological analysis of Luobuzhai should investigate whether there are two lexical tones or whether the categorical opposition is between toned (H) and toneless (zero, L).

The analytical tool of Linear Mixed Effects Analysis has produced an extensive quantification and description of Luobuzhai F0 from one dataset, without having to fill out a matrix with nonsense words, or anomalous sentences. Given the social setting in which this language is spoken, and the non-literate backgrounds of the subjects, it does not appear possible to record a large, one-effect-at-a-time dataset, of the type often used with literate subjects. Nevertheless, the end result of the present model is an extensive, if not exhaustive, listing of both human and linguistic effects and effect interactions that combine in the production of F0 in an endangered language of Southwest China. LME has also been a tool to elucidate suggested areas for further exploration, such as the extent of male-female differences in contrasts and questions, and whether the H/L tonal opposition can be reduced to merely the presence or absence of tone.

References:

Aston, J. A. D.; Chiou, J. M.; Evans, J. P.: Linguistic Pitch Analysis using Functional Principal Component Mixed Effect Models. Journal of the Royal Statistical Society. *59.2*: 297–317 (2010).

Baayen, R. H.: LanguageR: Data sets and functions with Analyzing Linguistic Data: A practical introduction to statistics R package (version 0.4) [Computer software] (2007).

Baayen, R. H.: Analyzing linguistic data: A practical introduction to statistics (Cambridge University Press, Cambridge 2008).

Baayen, R. H.; Davidson, D. J.; Bates, D. M.: Mixed-effects modeling with crossed random effects for subjects and items. Retrieved from: http://www.mpi.nl/world/persons/private/baayen/publications/baayenDavidsonBates.pdf /(2008).

Bates, D.; Maechler, M.: lme4: Linear mixed-effects models using S4 classes. R package (version 0.999375-31) [Computer software] (2009).

Bates, D. M.; Pinheiro, J. C.: Computational methods for multi-level models (Tech. Memo No. BL0112140-980226-01TM). Murray Hill, NJ: Lucent Technologies, Bell Labs. (1998).

Boersma, P.: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the Institute of Phonetic Sciences. University of Amsterdam *17*: 97–110 (1993).

Boothroyd, A.: Speech acoustics and speech perception. Texas: Pro-Ed series in Communication Disorders. (1986).

Draper, N. R.; Smith, H.: Applied Regression Analysis, 3rd Edition. (Wiley 1998)

Cunningham, E. P.; Henderson, C. R.: An iterative procedure for estimating fixed effects and variance components in mixed model situations. Biometrics *24*(1): 13-25 (1968).

Esposito, A.: On vowel height and consonantal voicing effects: Data from Italian. Phonetica *59*: 197-231 (2002).

Evans, J. P.: Contact-Induced tonogenesis in Southern Qiang. Language and Linguistics *2*(2): 63–110 (2001).

Evans, J. P.: How do you analyze a three-tone system, when all three tones are high? Presented at 39th ICSTLL, Sept. 14–17, 2006, University of Washington, Seattle, 21 pp. (2006).

Evans, J. P.: 'African' tone in the Sinosphere. Language and Linguistics *9*(3): 463–490 (2008).

Faraway, J.: Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models (Chapman & Hall/CRC, Boca Raton 2006).

Fry, D.: Experiments in the perception of stress. Language and Speech *1*: 126-152 (1958).

Gandour, J.: Consonant types and tone in Siamese. Journal of Phonetics *2*: 237–250 (1974).

Gandour, J.: Perceptual dimensions of tone: Evidence from Cantonese. Journal of Chinese Linguistics *9*: 20–36 (1981).

Gussenhoven, C.: The phonology of tone and intonation (Cambridge, Cambridge University Press 2004).

Henderson, C. R.: Estimation of variance and covariance components. Biometrics 9(2): 226-252 (1953).

Hombert, J. M.: Towards a theory of tonogenesis: An empirical, physiologically and perceptually-based account of the development of tonal contrasts in language; Doctoral diss. University of California, Berkeley. (unpublished 1975).

Hombert, J. M.; Ohala, J. J.; Ewan, W. G.: Phonetic Explanations for the Development of Tones. Language *55*(1): 37–58 (1979).

Huss, V.: English word stress in the post-nuclear position. Phonetica *35*: 85–105 (1978).

Hyman, L. M.: Privative tone in Bantu. In S. Kaji (Ed.), Cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese accentology, and other topics, pp. 237-257. (Institute for the Study of Languages and Cultures, Tokyo 2001).

Johnson, K.: Quantitative Methods in Linguistics (Blackwell, Malden, MA 2008).

Khouw, E.; Ciocca, V.: Perceptual correlates of Cantonese tones. Journal of Phonetics *35*:104–117 (2007).

Kochanski, G.; Shih, C.; Jing, H.: Quantitative Measurement of Prosodic Strength in Mandarin. Speech Communication *41*(4): 625–645(2003).

Kohler, K.: The perception of prominence patterns. Phonetica *65*: 257-269 (2008).

Ladd, R.: Declination 'reset' and the hierarchical organization of utterances. J. Acoust. Soc. Am 84(2): 530-544 (1988).

Ladd, D.; Silverman, K.: Vowel intrinsic pitch in connected speech. Phonetica *41*: 31-40 (1984).

Laird, N. M.; Ware, J. H.: Random-Effects Models for Longitudinal Data. Biometrics *38*, 963-974 (1982).

Ladefoged, P.: Phonetic data analysis: an introduction to fieldwork and instrumental techniques (Blackwell, Malden, MA 2003).

LaPolla, R J.; Huang, C.: A Grammar of Qiang, with Annotated Texts and Glossary (Mouton de Gruyter, Berlin 2003).

Liu, F.; Xu, Y.: Parallel encoding of focus and interrogative meaning in Mandarin intonation. Phonetica *62*: 70-87 (2005).

Liu, G.: Mawo Qiangyu Yanjiu (Sichuan Nationalities Press, Chengdu 1998a).

Liu, G.: Lun Qiangyu shengdiao de chansheng he fazhan. Minzu Yuwen *2*:1–8 (1998b).

Mazaudon, M.; Michaud, A.: Tonal contrasts and initial consonants: A case study of Tamang, a 'missing link' in tonogenesis. Phonetica *65*: 231-256 (2008).

Mishra, T.; van Santen, J.; Klabbers, E.: Decomposition of pitch curves in the general superpositional model, Proceedings of Prosody 2006, Germany. (2006).

Myers, S.: Downdrift and pitch range in Chichewa intonation. In J. Ohala; Y. Hasegawa; M. Ohala; D. Granville; A. Bailey (Eds.). Proceedings of the XIVth International Congress of Phonetic Sciences (pp. 1981–1984). Berkeley: Linguistics Department, University of California. (1999).

Myers, J.; Tsay, J.: Neutralization in Taiwan Southern Min Tone sandhi. In Y. E. Hsiao, H.-C. Hsu, L.-H. Wee, and D.-A. Ho (Eds.). Interfaces in Chinese phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday, pp. 47-78. Language and Linguistics Monograph Series Number W-8. (Academia Sinica, Taipei 2008).

Nguyen, N.; Fagyal, Z.: Acoustic aspects of vowel harmony in French. Journal of Phonetics *36*(1): 1–27 (2008).

Ohala, J. J.; Eukel, B. W.: Explaining the intrinsic pitch of vowels. In R. Channon & L. Shockey (Eds.). In Honor of Ilse Lehiste pp. 207–215 (Foris, Dordrecht 1987).

Ohde, R. N.: Fundamental frequency as an acoustic correlate of stop consonant voicing. The Journal of the Acoustical Society of America *75*(1), 224–230 (1984).

Pinheiro, J. C.; Bates, D. M.: Mixed-Effects Models in S and S-Plus: Statistics and Computing Series. (Springer-Verlag, New York 2000).

Plag, I.: The variability of compound stress in English: structural, semantic and analogical factors. English Language and Linguistics *10*(1): 143–172 (2006).

R Development Core Team (Designers): The R Project for Statistical Computing [Computer software and manual]. Retrieved from: http://www.R-project.org/ (2009).

Van Santen, J.; Assignment of segmental duration in text-to-speech synthesis. Computer Speech and Language *8*: 95–128(1994).

Van Santen, J.: Prosodic modeling in text-to-speech synthesis. Proceedings of Eurospeech-97, Rhodes, Greece. (1997).

Van Santen, J.; Mishra, T.; Klabbers, E.: Estimating phrase curves in the general superpositional intonation model. ISCA Speech Synthesis Workshop 5, Pittsburgh, PA. (2003).

Van Santen, J.; Shih, C. :Suprasegmental and segmental timing models in Mandarin Chinese and American English. Journal of the Acoustical Society of America *107*(2): 1012–1026 (2000).

Shih, C.: A Declination Model of Mandarin Chinese. In A. Botinis (Ed.). Intonation: Analysis, Modelling and Technology pp. 243–268 (Kluwer Academic Publishers, Dordrecht 2000).

Shih, C.; Ao, B.: Duration study for the Bell Laboratories Mandarin Text-to-Speech system. In J. van Santen, R. Sproat, J. Olive, J. Hirschberg (Eds.). Progress in Speech Synthesis pp. 383-399 (Springer-Verlag, New York, 1997).

Sluijter, A.; Terken, J.: Beyond sentence prosody: Paragraph intonation in Dutch. Phonetica *50*: 180-188 (1993).

Sulter, A. M.; Schutte, H. K.; Miller, D. G.: Standardized Laryngeal Videostroboscopic Rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency and age. Journal of Voice *10*(2): 175-189 (1996).

Sun, H.: Qiangyu Jianzhi. (Nationalities Press, Beijing 1981).

Tseng, C. Y.: Higher Level Organization and Discourse Prosody. The Second International Symposium on Tonal Aspects of Languages: 23–34 (2006a).

Tseng, C.Y.: Prosody Analysis. Advances in Chinese Spoken Language Processing, pp. 57-76. (World Scientific Publishing, Singapore 2006b).

Traunmüller, H.; Eriksson, A.: The frequency range of the voice fundamental in the speech of male and female adults. (manuscript 1995).

Venables, W. N.; Ripley, B. D.: Modern Applied Statistics with S (4th ed.) (Springer, NY 2002).

Wen, Yu.: Wenzhou Qiangyu Cihui Jianbian (Luopuzhai Fangyan). Zhongguo Wenhua Yanjiu Huikan *10* (1951).

Wen, Y.; Fu, M.: Wenchuan Lobuzhai Qiangyu yinxi. Studia Serica *3*(2):14–25 (1943).

Werner, S.; Keller, E.: Prosodic aspects of speech. In E. Keller (Ed.), Fundamentals of speech synthesis and speech recognition: Basic concepts, state of the art, and future challenges pp. 23–40 (John Wiley, Chichester 1994).

Whalen, D. H.; Levitt, A. G.: The universality of intrinsic F0 of vowels. Journal of Phonetics *23*: 349–366 (1995).

Xu, Y.: Effects of tone and focus on the formation and alignment of F0 contours. Journal of Phonetics *27*: 55–105 (1999).

Xu, Y.: Principles of tone research. Proceedings of International Symposium on Tonal Aspects of Languages pp. 3–13 France: La Rochelle. (2006).

Yip, M.: Tone. (Cambridge University Press, Cambridge 2002).

Yuan, J.; Shih, C.; Kochanski, G.: Comparison of declarative and interrogative intonation in Chinese. In B. Bel & I. Marlien (Eds.), Proceedings of the First International Conference on Speech Prosody pp. 711–714 France: Aix-en-Provence. (2002).

Zeng, X.; Martin, P.; Boulakia, G.: Tones and Intonation in Declarative and Interrogative Sentences in Mandarin. Proceedings of the Tonal Aspects of Languages, March 28–31, pp. 235–238 Chian: Beijing. (2004).

Zhao, Y.; Jurafsky, D.: The effect of lexical frequency and Lombard reflex on tone hyperarticulation. Journal of Phonetics *37*(2): 231-247 (2009).