The experimental state of mind in elicitation: illustrations from tonal fieldwork *

Kristine Yu

Abstract

Inspired by Hyman (2007, 2001), we cast fieldwork elicitation with an experimental state of mind. We begin by illustrating how principles of experimental design underlie Pike (1948)'s classic toneme discovery procedure. These principles include the consideration of confounding variables that can obscure the relation between the manipulated independent variable and the affected dependent variable, as well as the explicit statement of linking hypotheses about the map between unobservables, like tonal concepts, and quantities we can use to indirectly observe them, such as pitch patterns. We show that recognizing the role of these principles in toneme discovery allows us to use the same principles to generalize to elicitation methodology for fieldwork ranging from acoustic studies of tonal realization to explorations of tonal allophony and alternation and the syntax-prosody interface. We close with an in-depth look at considerations for linking hypotheses between tonemes and observable phonetic parameters of the voice source.

1 Introduction

This paper revisits a very old method for studying tone languages: "how do you say X?" Summed up like this, elicitation from linguistic consultants might not seem to have the capacity to be particularly well-defined or rigorous as a method. But that would be a mischaracterization, as is evident from Pike (1948)'s brief summary of elicitation methodology for discovering tonemes in his classic tome on how to study a tone language:

> The procedure indicated is basically a method of controlling free, conditioned, key, mechanical, morphological, and sandhi tonal changes by inserting lists of words into selected contexts so as to reduce the number of variables at any one time and give the investigator the opportunity of observing the significant linguistic pitch in its simplest contrastive forms Pike (1948, p. vii)'s.

*We thank our wonderful Kirikiri consultant Alfius Polita and the other members of the Kirikiri team from the 2011 ANU tone workshop: Anthony Woodbury, Rebecca Hetherington, Peter Appleby, Rosey Billington, Lea Brown, Isolde Kappus, and Sutriani Narfafan, and we gratefully thank Steven Bird, Mark Donohue, Larry Hyman, and Mark Liberman for organizing the inspiring tone workshops. We also thank Brian Dillon and John Kingston for invaluable discussions. This paper is dedicated to Will Leben, my first mentor in linguistics.

Pike's description evokes an *experimental state of mind* in elicitation. By referring to the experimental state of mind, we take inspiration from Hyman (2001)'s conception of fieldwork as a state of mind: just as Hyman remarks that "it is possible to be a fieldworker without constantly going to the field," so is it possible to be an experimentalist in pursuing fieldwork without ever stepping foot in a lab. Bringing the lab to the field, e.g. conducting psycholinguistic experiments in the field, is a burgeoning line of research—see for instance, the forthcoming special issue of *Language and Cognitive Processes*, Laboratory in the Field Jaeger et al. (2014). However, the focus of this paper is how an experimental state of mind can inform and enrich traditional elicitation methodology in fieldwork.

In this paper, we expound upon Hyman (2007)'s remark that "elicitation is experimental phonology" and illustrate how to guide elicitation in terms of experimental design and analysis. The principles illustrated are applicable regardless of the linguistic phenomenon and data being elicited, but for this paper, we focus on: (1) the problem of discovering the tonemes of a language and (2) how to elicit and analyze phonetic data to tackle this problem. We present illustrative examples from our experiences in discovering the tones of Kirikiri (Papua New Guinea, Lakes Plain) in fieldwork conducted with team members at the Prosodic Systems in New Guinea Workshop in December 2011, supplemented by experiences in fieldwork on Samoan (Samoa, Polynesian), Mandarin (China, Sino-Tibetan), and Cantonese (China, Sino-Tibetan).

The structure of the rest of the paper is as follows: we begin by using Pike's toneme discovery procedure as a starting point for illustrating the role of experimental design in elicitation in §2. Then, we show how we can use principles of experimental design to generalize beyond Pike's elicitation methods in tonal fieldwork in §3. Since assumptions about how we can observe reflexes of directly unobservable tonemes underpin any conclusions we can draw in elicitations, we close in §4 with an extended discussion of a key component of the experimental design—the map between the unobservable cognitive concept of a toneme and observable acoustic and perceptual properties of speech, including f0 and pitch and beyond.

In addition to the body of the paper presented here, a major component of the paper is a set of tutorials on collecting and analyzing phonetic data from elicitations. These tutorials and supporting files are referred to in the body of the paper and presented as supplementary material online at `www.krisyu.org` and `www.github.com/krismyu/ldc-kiy`. We've chosen to prepare on-line tutorials to facilitate interactivity with code and supporting media.

2 Components of an experimental state of mind

In this section, we walk through Pike (1948)'s toneme discovery algorithm (§2.1) and then show how it can be conceived as an application of principles of experimental design in two stages (§2.2, §2.3).

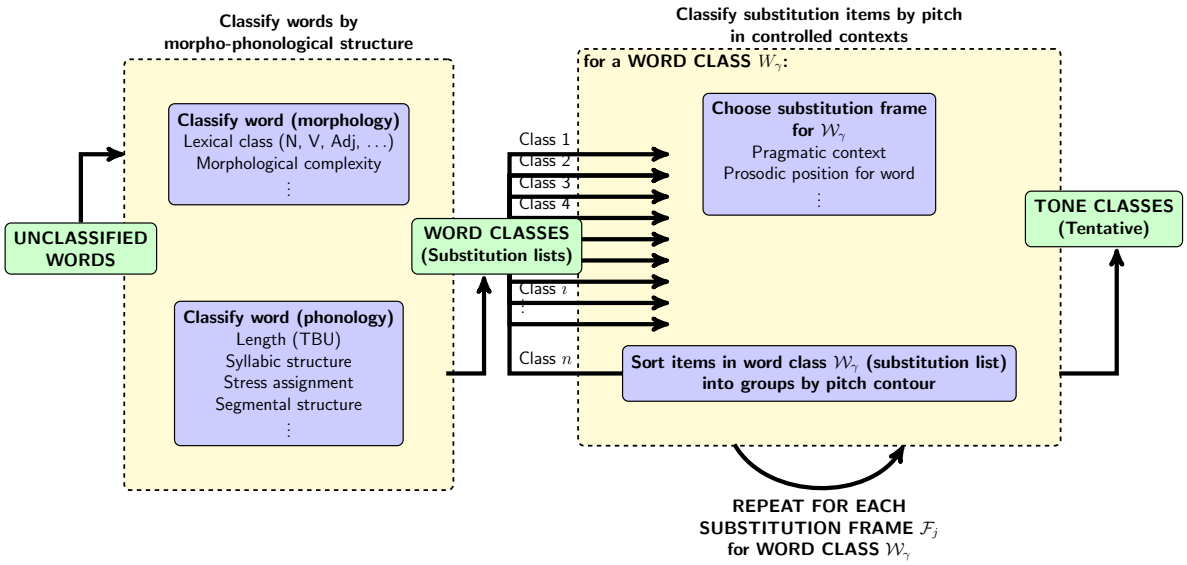


Figure 1: A schematic summarizing the discovery procedure for tonemes in Pike (1948, Ch. 4). There are two main steps: (1) classification of words by morpho-phonological structure into word classes (*substitution lists*), and for each substitution list, (2) classification of the items in the list into tentative tonal classes by their pitch patterns in a phrasal (possibly sentential) context, a *substitution frame*. The second step is repeated for all applicable substitution frames for the substitution list, for all substitution lists, and proposed tonal classes may be adjusted after each iteration. Abbreviations: TBU = tone bearing unit, N = noun, V = verb, Adj = adjective.

2.1 Pike (1948)'s toneme discovery procedure: a walk-through

Pike's elicitation methodology for discovering tonemes in Pike (1948, Ch. IV, p. 48-54) consists of two main steps:[1]

1. Classification of words into word classes (*substitution lists*) of uniform morpho-phonological structure
2. Classification of words into groups that are tonally uniform in controlled contexts

Figure 1 schematizes the entire two-step procedure; Figure 2 illustrates the first step of classification of words into word classes in detail, and Figure 3 exemplifies the second step of examining the tonal properties of words in controlled contexts.

As Pike states, the purpose of the first classification of words by morpho-phonological properties is to control properties of the words such that any variation in pitch patterns of words is highly likely to be due *only* to tonemic contrast:

> The first classification brings together words which are somewhat alike in phonetic and grammatical structure. Such a grouping tends to reduce the hazards

[1] Pike further describes methods for determining the number of tonemes and their phonological description in Ch. V, but we focus on his first two steps for the purposes of illustrating the experimental state of mind in elicitation in this paper.

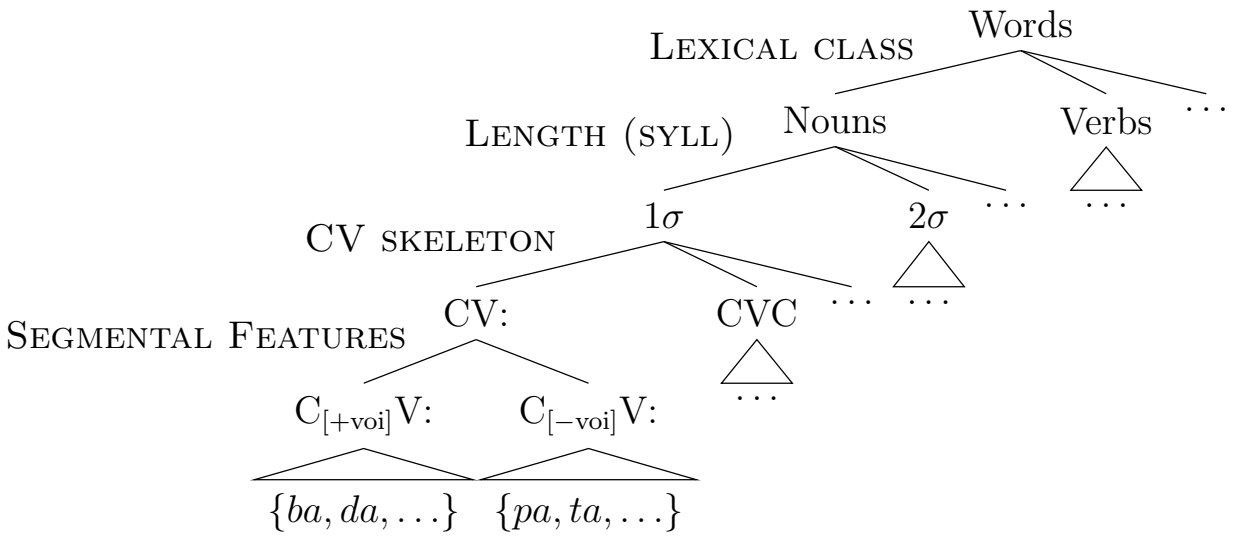


Figure 2: A classification tree depicting the initial step of partitioning words by morpho-phonological structure in the discovery procedure for tonemes in Pike (1948, Ch. 4). Words are partitioned by lexical class, length (in syllables), CV skeleton, and segmental features (here, onset voicing). The sets of terminal elements $\{ba, da, \ldots\}$, $\{pa, ta, \ldots\}$ are word classes Pike calls *substitution lists*. These word classes are uniform in morpho-phonological structure and are to be elicited in controlled contexts in the second step of the discovery procedure.

> introduced in the analysis of these words by segments which cause nonphonemic modification of tonemes. The ear is distracted in its listening for pitch when the forms of the items under attention are not comparable. (Pike, 1948, p. 48)

The morpho-phonological properties listed for consideration in this first classification step in Figure 2 (lexical class, length, CV-skeleton, etc.) should not be taken to be a rigid prescription for exactly how to partition words into classes for pitch pattern comparison. The set of properties listed may be neither necessary nor sufficient for toneme discovery in a particular language. Moreover, the output word classes, the *substitution lists* from the classification, may well be adjusted in the course of fieldwork to yield a coarser or finer-grained partition of the words. For instance, the classification tree in Figure 2 currently does not distinguish vowel quality, but in the course of tonal fieldwork, if one suspected an interaction between vowel height and pitch patterns, a natural step would be to include vowel height in word classification. This heightened attention to vowel quality in classification would yield a finer-grained partition of the words and introduce an additional level of depth to the classification tree.

The output of the first classification step, the word classes Pike calls *substitution lists*, is the input to the second classification step schematized in Figure 3. For each substitution list, a set of *substitution frames* is generated. As shown in the middle box labeled "substitution frames" in Figure 3, the specification of these frames includes phrasal-level prosodic structure as well as pragmatic context.[2] For each iteration of

[2]Pike (1948, p. 51) actually discusses controlling "emotional context... to prevent intonational changes". We take "emotional context" to have a meaning roughly equivalent to that of that of *pragmatic context* as a cover term for situational context that may interact with prosody.

eliciting a substitution list in the context of a substitution frame, the items in the substitution list are sorted into tentative tonal groups by their particular pitch patterns in the substitution frame. Comparison of hypotheses for tonal groupings from different substitution frames for a substitution list provides converging or diverging evidence for particular proposed tonemes. These hypotheses may eventually also be compared to generalize not only over substitution frames within a substitution list, but also across substitution lists.

A final important point about the tonal classification step is the emphasis on the assignment of group labels to words rather than the nature of the labels themselves. Pike writes that: "up to this point there has been no essential need for tonal transcription. It is the grouping as such which has been important" (Pike, 1948, p. 55): in Figure 3, what's critical in the rightmost "pitch patterns" box is not the IPA tonal transcription, but the colors of the boxed tones, which indicate group membership.[3]

The output of Pike's second toneme discovery step can be thought of as a list of mappings from words to glass jars in a pantry storing jars of words. We might have faded written labels on the jar lids, but what really serves as the jar labels in the pantry is the set of identifying and distinguishing properties of the contents inside a jar. Indulging a bit with this pantry metaphor: take a shelf storing various kinds of pasta—we might not know whether to label a jar as *gemelli* or *bucatini*, but it's easy enough to keep the different pasta jars straight and to glance at the jars and get down "the jar with the little spirals" as opposed to "the jar with those long tubes." We discuss considerations for the nature of tonal class labels further in §4 when we examine the map between tonal classes and acoustic and perceptual dimensions.

Now that we've laid out Pike's toneme discovery procedure, we unpack it into two sequential stages: (1) treating tonal class as a *latent variable*, as something to infer from the unexplained variability in pitch contours over words (§2.2), and (2) explicitly manipulating (putative) tonal classes as *independent variables* in testing hypotheses about the partition of words into tonal classes (§2.3). We provide running illustrations of each stage from fieldwork on Kirikiri.

2.2 Experimental design in early toneme discovery: tonal class as a latent variable

At the earliest stages of studying a language, we may not be even sure if it has lexical tonal classes, and even if we are confident that there does exist tonal classes in the language, we don't know enough about them to manipulate TONAL CLASS as an *independent variable* and explicitly group words into different tonal classes.[4] Our research

[3]This observation about the importance of sorting and clustering pitch patterns in contrast to the negligible role of the choice of representation of pitch patterns in the early stages of toneme discovery, also discussed at the Berkeley Tone Workshop in 2011 (NSF Project Prosodic Systems in New Guinea: Integrating computational and typological approaches to linguistic analysis), was the key motivation for the development of computational software to aid this clustering (see this volume).

[4]Note that the independent variables are manipulations of each elicitation item constructed and elicited by the fieldworker (the *experimental unit*) and uttered by the consultant, not manipulations of words in the language of study. When we discuss manipulating syllable structure or tonal class as independent variables, we are not claiming to manipulate the syllable structure or tonal class of a word (that's left to nature! Such a study with the word rather than the elicitation item as the experimental unit would not be an experiment, but an *observational study* (Rosenbaum, 1999).) Instead, we manipulate the properties of an elicitation

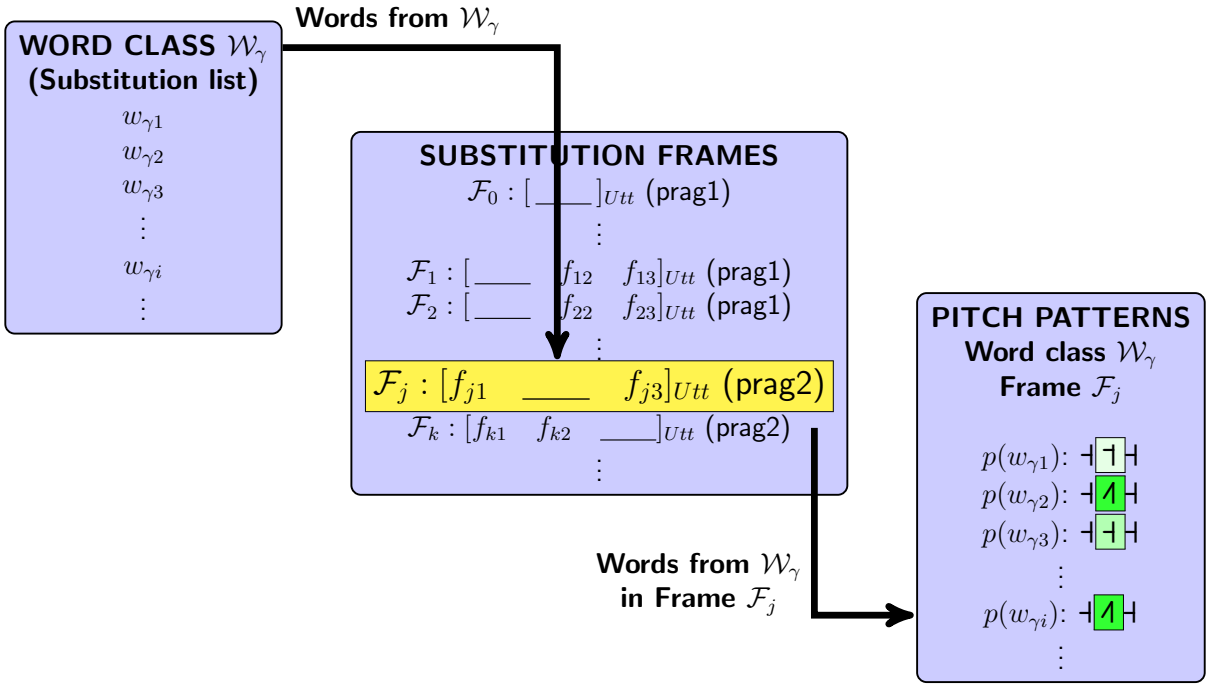


Figure 3: A schematic summarizing the second step in the discovery procedure for tonemes in Pike (1948, Ch. 4). This consists of eliciting items in a word class (substitution list) $\mathcal{W}_\gamma$ in a substitution frame $\mathcal{F}_j$ (boxed in yellow) among the substitution frames licit for $\mathcal{W}_\gamma$ and sorting the items by pitch pattern into tentative tone classes (tonemes). Frame $\mathcal{F}_j$ positions items utterance-medially in a particular pragmatic context $prag2$. The elicited pitch patterns (transcribed with IPA tone symbols) for $\mathcal{W}_\gamma$ in $\mathcal{W}_\gamma$ are clustered into tentative tone classes. For simplicity, we assume that the substitution items are monosyllabic and that there is one tone per syllable. Here, $w_{\gamma 2}$ and $w_{\gamma i}$ are proposed as members of a tentative tone class (colored bright green) since they share a pitch pattern in context $\mathcal{F}_j$ distinct from the pitch patterns in $\mathcal{F}_j$ for other words in $\mathcal{W}_\gamma$.

hypothesis at this point is:

Hypothesis 1 (Existence of tonal classes) *There are tonal classes in Kirikiri.*

Note that tonal classes are not directly observable: we can only indirectly observe the presence of tonal classes through the *dependent variable* of the pitch contour over the word, which we observe as we elicit each new item with our consultant. As is implicit in Pike's procedure, in order to proceed in discovering hidden structure which is not directly observable, we must make assumptions linking properties of speech that we can observe—*observed variables*—to unobservable, underlying tonal classes—*latent variables*. Such assumptions can be encoded in a *linking hypothesis*, which makes explicit assumptions about the relation between observed variables, e.g. pitch contours over words, and (unobserved) latent variables, e.g. tone classes:[5]

Hypothesis 2 (Linking hypothesis between lexical tone classes and pitch contours) *Lexical tone classes (tonemes) induce systematic variation in the pitch contours of words. Therefore, the unobservable cognitive concept of a toneme is observable via its influence on the pitch contours of words: if two words have different pitch contours, then they belong to different lexical tonal classes.*

This linking hypothesis in Hypothesis 2 should make the reader sputter with incredulity: what about all the other potential sources of influence on the pitch contours of words uttered during an elicitation that we've been discussing in §2.1? Suppose, for instance, that we elicited one word at the end of the utterance and another one in the middle of the utterance. If these two words were uttered with different pitch contours, and we concluded on the basis of Hypothesis 2 that the two words belonged to different tonal classes, we could be mistaken: perhaps the pitch differences were actually due to the difference in prosodic position between the two elicited utterances.

There is still a sense in which Hypothesis 2 is reasonable, though. LEXICAL TONE CLASS could certainly be *one* variable in a model explaining variability in the pitch contour over a word, and PROSODIC POSITION and any other of the variables mentioned in §2.1 could be *explanatory variables* as well—independent variables that are central to our research questions and hypotheses. There could be *many* explanatory variables in a model of pitch contour variation. In order to provide evidence that TONAL CLASS is one of these explanatory variables (Hypotheses 1 and 2), we follow the strategy below:

- Propose a set of independent variables to include in a model to explain pitch contour variation. This set does not include TONAL CLASS, since we are not yet at the stage where we can manipulate TONAL CLASS as an independent variable.

item—properties of the substitution frame, as well as properties of the target word, including the tonal class of the target word within the elicitation item. Throughout this paper, when we refer to any independent or dependent variable, we always mean to refer to the variable associated with the experimental unit of the elicitation item, e.g. "the pitch contour over a word" is short for "the pitch contour over a word embedded in an elicitation item uttered by the consultant in an elicitation session."

[5]Some of the earliest discussion of linking hypotheses in studying human behavior comes from the study of human vision (Brindley, 1960), in which linking hypotheses must be made about the mapping between perceptual and psychological states. Here, we're also making linking hypotheses about the mapping between perceptual (auditory) states (pitch contours over words) and psychological (cognitive) states (tonal concepts—tonemes). See Teller (1984) for a historical perspective and some philosophical discussion and Yurovsky et al. (2012) for current work on linking hypotheses.

- See how far the set of explanatory variables in our model goes towards explaining the variability in pitch contours of words from elicited utterances.
  - If the leftover unexplained variability is huge,[6] we suspect that we may have missed one or more important explanatory variables in our model, and if we have considered a sufficiently wide range of possible variables for our model, then we can conclude with some confidence that we need to have TONAL CLASS in the model to help explain the pitch contour variability. (Figure 4b) We can then proceed to check if we have a big gain in explained variance with TONAL CLASS in the model.
  - If the leftover unexplained variability is small, we have no compelling positive evidence for the language having lexical tonal classes. The small proportion of variability left unexplained is likely due to a constellation of secondary influences on pitch contour variation we've abstracted away from, e.g. various sources of elicitation session-to-elicitation session variability, as discussed further in §3.1. (Figure 4c)

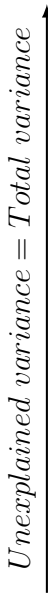
Under this strategy, we begin with a morass of variability in the pitch contour over a word (Figure 4a). None of this variability is explained, i.e. all of the variability is *unexplained*. Upon introducing variables into a model of pitch contour variation, we hope to be able to carve off some of that variability into *explained variability*, variability which is accounted for by our introduced explanatory variables. The better we understand what influences pitch contour variability, the more variability from the total variability that we are able to partition into *explained variability* and the less that remains *unexplained variability*. If our model does not do a good job of covering the sources of influence for pitch contour variability, the partition between explained and unexplained variability looks like the breakdown in Figure 4b, where most of the variability remains unexplained. If our model does in fact do a good job, then the partition looks like the decomposition in Figure 4c, where most of the variability is explained by the proposed explanatory variables.

At this point, our guiding research question is:

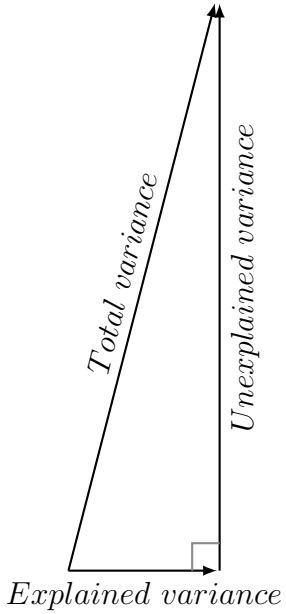
Question 1 (Explaining variability in pitch contours) *Can we explain part of the variability in the pitch contour over a word in an elicited utterance? How much of this variability can we explain?*

What explanatory variables should we include in a model of pitch contour variability? The answer from Pike's procedure is Hypothesis 3—the variables to include are in the list of variables we discussed in §2.1, which we summarize in Table 1. Unlike TONAL CLASS, we know enough about the variables in Table 1 to manipulate them as independent variables. For instance, we can explicitly manipulate the SYLLABIC LENGTH of the target word in an elicitation item to be set at different *levels*—possible instantiations of a variable in the experimental design, e.g. monosyllabic, disyllabic, etc.

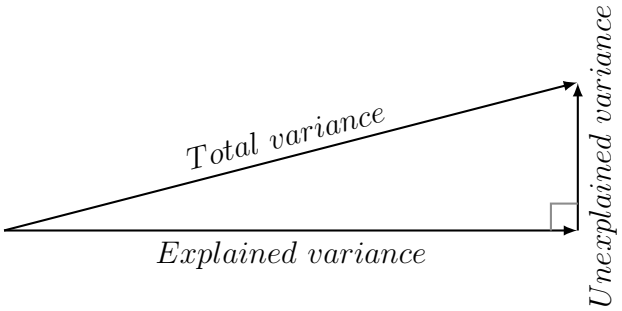
Hypothesis 3 (Explaining variation in pitch contours over words) *The primary influences on variation in pitch contours over words are the variables listed in Table 1.*



(a) No variability explained



(b) Most of the variability unexplained



(c) Most of the variability explained

Figure 4: A geometric summary of partitioning the total variability in the pitch contour over a word into two parts (Saville and Wood, 1986): (1) the variability explained, i.e., the variability accounted for by the independent variables given in Table 1 and (2) the residual (unexplained) variability, which must be attributed to the influence of variables not enumerated in Table 1—in particular, we hypothesize, the influence of the latent variable TONAL CLASS.

Variable	Examples of levels
Lexical class	noun, verb
Morphological complexity	simplex, complex
Syllabic length	monosyllabic, disyllabic
CV skeleton	CV, CVC, CCV, CV:
(Vowel length)	(short, long)
Segmental features	intial voiced stops, initial voiceless stops
Stress	Present, absent
Stress assignment	Initial, peninitial, 3rd syllable, penultimate
Prosodic position	utterance-initial, isolation
Pragmatic context	Out of blue focus, contrastive focus on first NP
Syntactic structure	Possessive prenominal phrase, relative clause

Table 1: Independent variables in Pike (1948)'s toneme discovery procedure. The table gives each variable name, and a couple examples of levels—possible instantiations of a variable in the experimental design, e.g. LEXICAL CLASS = noun; LEXICAL CLASS = verb. A horizontal line divides variables pertaining to the word (items in the substitution lists) and variables pertaining to the context (substitution frame). Vowel length is listed in parentheses since Pike mentions it explicitly, but it could also be subsumed under CV skeleton.

> We should not delude ourselves that the small list of variables in Table 1 exhaustively captures all aspects of the elicitation context in toneme discovery that may influence pitch contour variation over a word, since context is always unbounded (Is the consultant lethargic or excited? How far away are you sitting from the consultant? What were the last five words elicited—has the consultant started building a discourse context around them? What time of day is it? How much sleep did the consultant get last night?). However, we can try to capture the primary aspects of relevant context: we can tackle contextual variables that we know of and that we suspect may account for a large amount of variability in the pitch contour of a word in an elicited utterance. There'll always be variability left unexplained.[7]
>
> In sum, the initial stage in Pike's toneme discovery procedure can be cast in terms of seeing how far we can get in explaining variability in pitch contours of words without appealing to the hidden structure of tonal classes. We acknowledge every known source of influence, or at least all known primary sources of influence, on variability in the observable dependent variable, in an attempt to explain away all of the variability in the dependent variable. Any residual unexplained variability in the dependent variable must be due to a set of variables that has been overlooked. If there is a large amount of remaining unexplained variability, we hypothesize that we are failing to take into

[6] In statistical data analysis, what counts as "huge" is precisely defined in terms of the ratio of unexplained variability to total variability, but for application in fieldwork elicitation, an intuitive definition of "huge" is sufficient.

[7] It is typical in psycholinguistic studies for most of the variability in the dependent variable (which is frequently reaction time in a subject's response to some stimulus) to be due to variability among subjects and items: the explanatory power from the variables of interest is swamped by subject-to-subject and item-to-item variability. Baayen (2008, p. 281–282) gives an example where only 0.3% of the total variability explained by explanatory variables.

account the latent variable TONAL CLASS.

2.2.1 Example: the earliest stages in discovering Kirikiri tones

In this section, we illustrate the earliest stages of toneme discovery, when TONAL CLASS is treated as a latent variable in fieldwork on Kirikiri. We use data from the three earliest recorded elicitation sessions, `20111207-1-kiy-ap-wordlist`, `20111207-2-kiy-ap-framedwordlist`, and especially `20111208-6-kiy-ap-nps-vps` to demonstrate the process of exploring potential sources of unexplained variability in the dependent variable of the pitch contour over the word.

We begin with all the variability unexplained, the initial state illustrated in Figure 4a. A visualization of this state is shown in Figure 5, where we plot f0 contours over target words (substitution items) from all recorded elicitation items from elicitation sessions `20111207-1-kiy-ap-wordlist` and `20111207-2-kiy-ap-framedwordlist`. It looks like a mess.

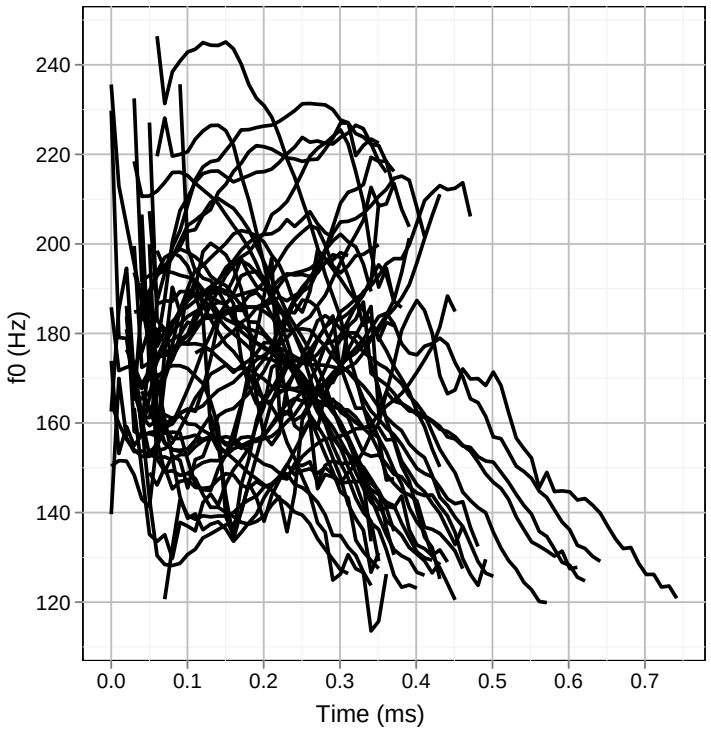


Figure 5: f0 contours from all recorded elicitation items from `20111207-1-kiy-ap-wordlist` and `20111207-2-kiy-ap-framedwordlist`.

With the data from `20111208-6-kiy-ap-nps-vps` (see Table **??** in §A for the full list of elicitation items), we also begin with all the variability in the pitch contour over the word unexplained. The plot of f0 contours for all substitution items for this elicitation session in Figure 6a is as much of a mess as Figure 5.

To work towards explaining some of the variability, we begin to partition the f0 contours by SUBSTITUTION FRAME (Table **??**) and properties of the substitution item (target word) given in Table **??**, such as LENGTH and LEXICAL CLASS. Across the sub-

Substitution context	Kirikiri	Gloss
Isolation	#___#	—
Adjective-black	____ koo	black ____
Adjective-small	____ soo	small ____
Adjective-female	____ kuu	female ____
VP-sleep	____ taru	____ is sleeping
VP-sound	____ kwaa zari	____ is making a sound

Table 2: A list of substitution frames from `20111208-6-kiy-ap-nps-vps`.

Target word	Gloss	Lexical class	Length (syll)
koo	ant	noun	1
foo	wallaby	noun	1
siji	pig	noun	2
nabij	dog	noun	2
kaza	gecko	noun	2
parai	bandicoot	noun	2

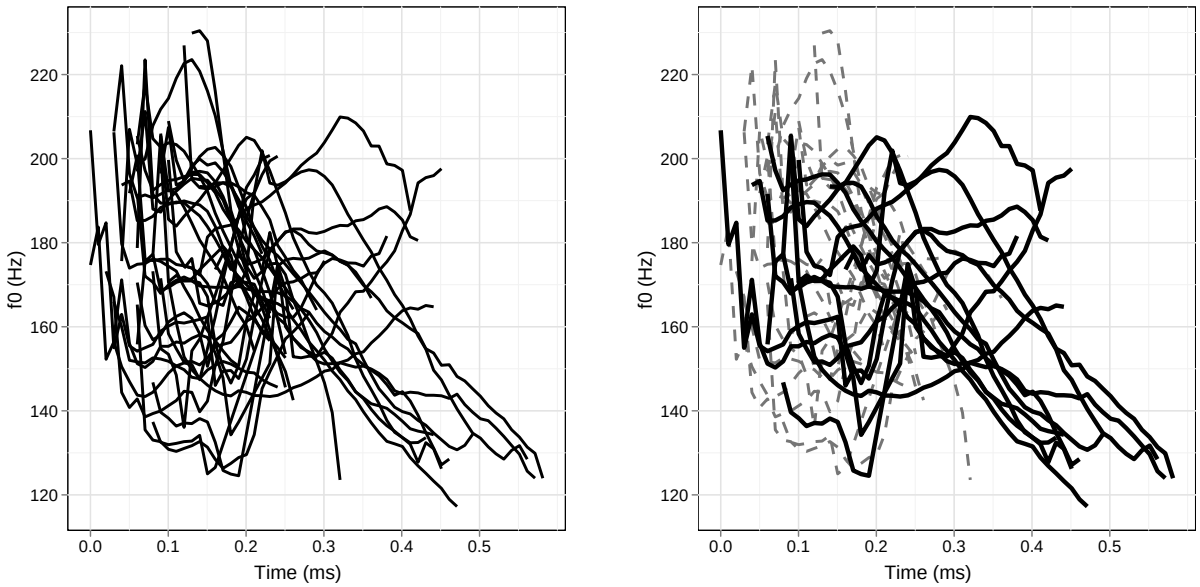
Table 3: A partial list of target words (substitution items) from `20111208-6-kiy-ap-nps-vps` and some of their properties.

stitution frames, PROSODIC POSITION is fixed to be initial, and SYNTACTIC STRUCTURE varies between (modified) NPs and VPs.

One immediate source of variability becomes clear when we differentiate between f0 contours uttered in isolation and all other f0 contours in the plots. Figure 6b shows that the f0 contours uttered in isolation, drawn as solid black lines, have durations roughly twice as long as f0 contours uttered in non-isolation contexts.

If we take a closer look at the non-isolation contexts, we find that further differentation between substitution frames in plotting f0 contours appears to reveal some additional structure in the set of f0 contours. Figure 7a shows f0 contours from non-isolation contexts, grouped by substitution frame. The f0 contours from VP frames include rises, while the f0 contours from adjective frames are falls.

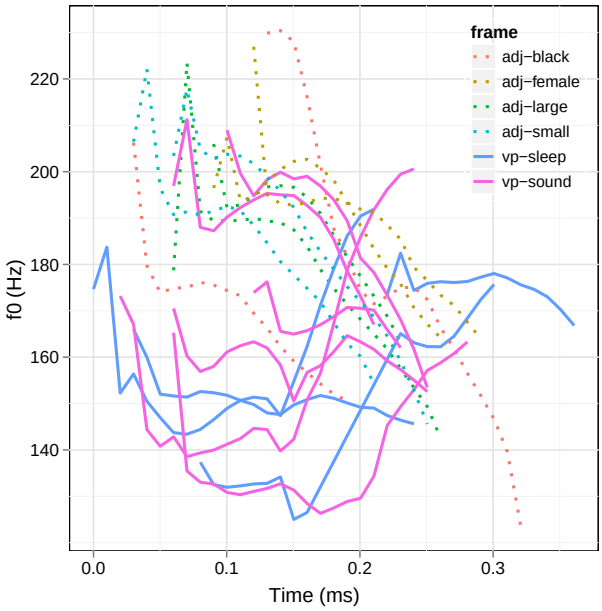
However, the relation between frame and f0 contour shape is likely spurious. In Figure 7b, we plot f0 contours for each target word in separate subplots, and the f0 contours within each subplot are very similar to one another despite being elicited in different substitution frames. It's clear from Figure 7b that the structural regularity in the f0 contours we found in Figure 7a is not due to the effect of different substitution frames, but rather, due to the effect of syllabic length of the target word: the f0 contours for the monosyllabic *foo* and *kee* fall steeply, while the f0 contours for the bisyllabic words show other patterns. It was an accident that only monosyllabic target words were elicited in adjective substitution frames.



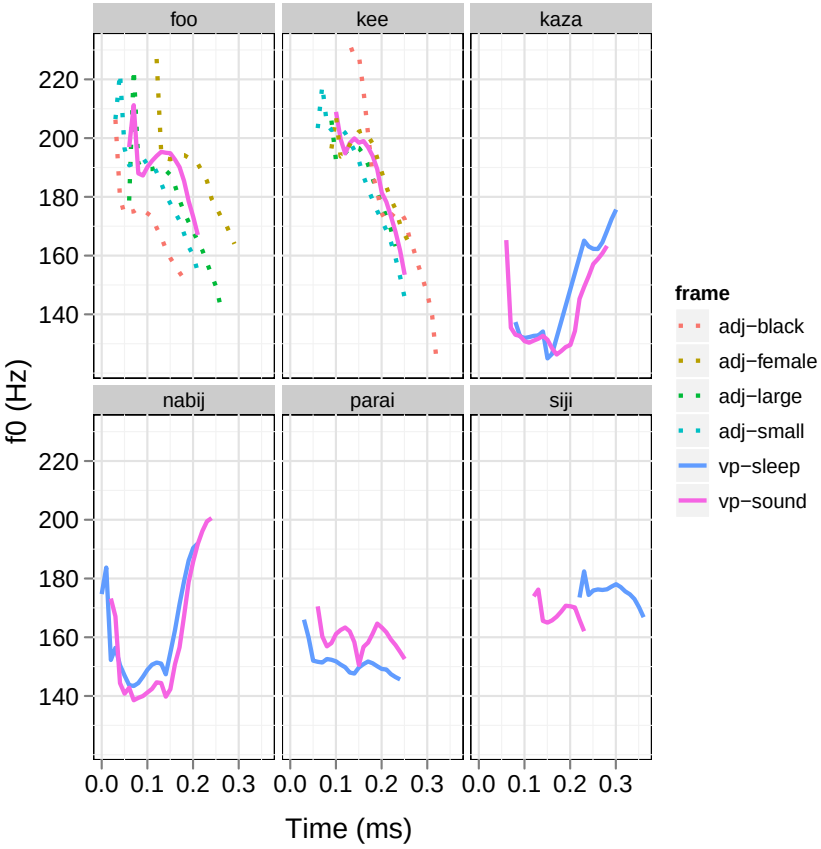
(a) All contours

(b) Contours from the isolation frame drawn in black solid lines; all other contours drawn in gray dashed lines.

Figure 6: f0 contours of target words from all recorded elicitation items in `20111208-6-kiy-ap-nps-vps`. When the contours from the isolation frames are differentiated from contours from other frames in Figure 6b, we move some of the variability in the f0 contours from being unexplained in Figure 6a to being explained.



(a) Contours for all words plotted together, with frames indicated by line type and color.



(b) Contours plotted separately for each target word, with frame indicated by color and line type.

Figure 7: f0 contours of the substitution items *foo* 'wallaby', *kee* 'ant', *kaza* 'gecko', *nabij* 'dog', *parai* 'bandicoot', and *siji* 'pig' from recorded elicitation items in `20111208-6-kiy-ap-nps-vps`.

The lack of structure in the set of f0 contours in Figure 7a in contrast to the orderliness of the set of f0 contours plotted by target word in Figure 7b suggests that the target word is a large source of variability in the pitch contour. However, the substitution frame and associated properties are not a large source of variability, since f0 contours elicited in different substitution frames for the same target word are very similar. A glance at the marked difference between f0 contours for *nabij* and *parai*, despite their sharing the same LENGTH adn LEXICAL CLASS, is strong evidence that we are missing some property of the target word, some latent variable, that is a large source of variability in the pitch contour: namely, TONAL CLASS.

2.3 Experimental design in later toneme discovery: tonemes as independent variables

In §2.2, our explanatory variables are a bunch of variables that we think could influence pitch contours over words within an elicitation item and thus obscure the relation between pitch contour variability and TONAL CLASS posited in the linking hypothesis, Hypothesis 2. Such variables are called *confounding variables*, and we performed a sleight of hand in §2.2 in treating these contextual variables as if they were explanatory variables.

Once we have enough experience with tonal classes to group words by their putative tonal classes, we can begin to treat (putative) TONAL CLASS (of the target word, the substitution item) as an *independent variable* to be systematically manipulated, alongside the set of independent variables given in Table 1. Our research question at this stage of toneme discovery becomes:

Question 2 (The effect of tonal class on pitch contours)
How does TONAL CLASS *affect the pitch contour of a word?*

Asking this research question introduces a new partition in our set of independent variables: the partition between *explanatory* and *confounding* variables (also called *extraneous* or *nuisance* variables). Since what we are interested in is the effect of TONAL CLASS on the pitch contour, TONAL CLASS is an explanatory variable. All the other independent variables (those enumerated in Table 1) are confounding variables. In the experimental design when TONAL CLASS was not yet on the table as an independent variable (§2.2), all the independent variables were treated as being explanatory variables to see if we could explain away all the variance in pitch contours over words with them, without resorting to the latent variable of TONAL CLASS.

In this new experimental design, we treat TONAL CLASS differently from the confounding variables, as it is the sole explanatory variable, and compare the proportion of variability explained with TONAL CLASS included as an independent variable to the proportion explained when it is not included. We deal with the confounding variables with strategies following the classic work of Fisher (1925, 1935).

Fisher proposed three strategies for reigning in the effect of confounding variables on the dependent variable: blocking, replication, and randomization.

2.3.1 Blocking and replication

Pike's strategy of dividing elicitation items into uniform groups is an example of *blocking* confounding variables and was conceived as such:

> In order to be significant the tonal contrasts must be found in words which are sufficiently similar to rule out interference from nonpitch characteristics, and they must occur in contexts which cannot cause the observed pitch differences. (Pike, 1948, p. 48)

For instance, elicitation items are split into homogenous groups by the blocking variable LEXICAL CLASS of the word: one block consists of only nouns, another block of only verbs, etc. Within each of these blocks, the level of LEXICAL CLASS is held constant, and each elicitation item is assigned to one level of the explanatory variable TONAL CLASS, e.g. Tone 1. Within an elicitation session, elicitation items are organized by these blocks: all items within one block are elicited before moving onto the next block.[8]. All the variables in Table 1 are treated as blocking variables. Although there are multiple blocking variables, we can create a single aggregate blocking variable called BLOCK subsuming all the confounding variables in Table 1.[9] Thus, a block consists of a homogenous group of elicitation items, matched for every variable listed in Table 1, e.g. one level of the BLOCK variable might be the group of items specified by the fixed levels in Table 4. This is like a generalized substitution frame, a complete fixed aggregate specification of context for an elicitation item.

Let's expand on this notion of a generalized substitution frame: while Pike refers to only the substitution frame as "context" in the quote about blocking above, we extend "context" to refer to the morphophonological properties of the words from the first step in the procedure as well: since we would like to assume that any variation in pitch patterns between words is due solely to tonemic contrast, *any* aspect of the elicitation situation that affects the pitch of a word, but which isn't tonal class, is part of the "context". What is of primary scientific interest here is tonemic contrast as the source of pitch variation between words. Thus, we make the methodological abstraction of setting aside sources, other than tonal class, that would potentially generate pitch variation between words.[10]

By blocking, we can take advantage of our knowledge of some of the sources of variability in the pitch contour of a word. Rather than leaving the variability induced by those sources as unexplained, we parcel that variability out between blocks, i.e., as variability explained by the blocking variable, thus adding to the proportion of variability explained. Previously, when we treated TONAL CLASS as a latent variable, explained variability consisted only of variability explained by variables aggregated in BLOCK. Now, the explained variability consists of variability explained by TONAL CLASS as well as variability explained by variables aggregated in BLOCK.

Blocking reduces the noise in the observed effect of TONAL CLASS on pitch contours. Across blocks, the effects of TONAL CLASS on pitch contours might look quite different, so that the particular pitch contours induced over words by different tonal classes may

[8] A powerful feature of Toney (see this volume) is to provide an easy way to perform post-hoc blocking, i.e. blocking of items from an elicitation session after the session is completed. Even if blocking was not imposed during the session, Toney can extract clips of items and play them such that one can hear all the items in a post-hoc block one after the other.

[9] There are more sophisticated ways to incorporate multiple blocking variables into an experimental design, such as the Latin square designs commonly used in psycholinguistics (Montgomery, 2005, p. 136–142), but it's sufficient for our purposes here to conceptually discuss a single aggregate blocking variable.

[10] The strategy of making methodological abstractions in the process of groping toward scientific understanding—setting contextual factors aside (for the moment) to hone in on what is of primary interest—is an old one. Plato (360 B.C.E) described it as carving nature at its joints.

Variable	Fixed levels
Lexical class	noun
Morphological complexity	simplex
Syllabic length	disyllabic
CV skeleton	CVCV
(Vowel length)	(short)
Segmental features	initial voiceless obstruent
Prosodic position	utterance-initial
Pragmatic context	Out of blue focus
Syntactic structure	Possessive prenominal phrase

Table 4: An example of a block specification aggregated over the confounding variables in Table 1. Each of the confounding variables are fixed at the levels stated in the table, e.g. LEXICAL CLASS in the block is fixed to be "noun". We might call this particular specification of the confounding variables *Block 1*, one of multiple levels of the aggregate blocking variable BLOCK. Another block, say *Block 2*, might have identical specifications other than that LEXICAL CLASS is "verb" rather than "noun" in *Block 2*.

be quite different. But within a homogeneous block, systematic variability in pitch contour induced by TONAL CLASS is much clearer since variability in pitch contours within a block due to factors other than TONAL CLASS is small. What is uniform across blocks is not the particular pitch contours induced by a given tonal class, but the unified, group behavior of members of a tonal class within each block in patterning as a unit.

An additional benefit of blocking is that each block serves as a *replication* of the experiment testing to confirm the way that TONAL CLASS affects pitch contours of target words. Converging evidence from each replication about systematic variability in pitch contours due to tone classes boosts our confidence about our posited tonal contrasts.

However, sometimes fixing a confounding variable at a constant level for the whole experiment, i.e. running just a single block, is also a good option. (Sometimes one can also run just a subset of multiple levels among all the possible levels, but this is less common than running either all of the levels or just one, since it's hard to justify why one picked some levels but not others.)

For instance, one might consider the variable SONORANCY (of all consonantal segments in the target word), with the two levels [+sonorant], [-sonorant]. One could vary between these two levels between blocks, but one could also fix SONORANCY to be [+sonorant]. This is very common in intonational experiments, since working exclusively with [+sonorant] segments can help reduce segmental perturbations to the pitch contour when tone-segment interactions are not of interest (see §4.3). Abstracting away from [-sonorant] consonants has the disadvantages that we miss the opportunity to: (1) replicate our elicitation experiment to build our confidence in our conclusions about toneme discovery, and (2) study interactions between levels of SONORANCY and TONAL CLASS. But by fixing SONORANCY at [+sonorant], we have the advantages of: (1) reducing noise in the relation between TONAL CLASS and the pitch contour, thus making it easier to uncover tonemes, and (2) reducing the number of items in the

elicitation so that the elicitation isn't too long and grueling.

Whether it's best to hold a confounding variable fixed or to vary it depends entirely on the research question. We provide some examples of making this decision in §3 and §3.3, but it's really a case-by-case decision so it's difficult to give general guidelines. One rule of thumb for factors to consider in the decision, if there are no other compelling reasons based on the research question, is to consider either choosing the "vanilla" level or choosing one or more extreme levels.[11] By "vanilla" level, we mean, roughly speaking, the most unmarked level. For instance, for SYLLABLE STRUCTURE, this might be CV rather than CCV, CVV, or CVC, depending on the word-prosody of a language. Picking a "vanilla" level can be a good option when you are wholly uninterested in the variable and simply need to choose something in order to set up your experiment. Picking extreme levels can be a good option when you are more interested in the interaction between your explanatory variable(s) and the confounding variable at hand—perhaps you're worried about generalizing your conclusions across the levels of the confounding variable, so performing replications at one or more extreme levels provides converging evidence for your conclusions and/or a challenging test of your hypotheses under worst case limiting conditions.[12]

2.3.2 Randomization

Some confounding variables are not amenable to blocking. Consider the order of elicitation items:

> The mere fact that one word is necessarily said before the other in repetitions by the informant will frequently cause sandhi changes, or phrasal conditioning, or intonational modifications of one of the words. To check on this possibility the investigator should (1) reverse the order in which the items are repeated, and (2) have the informant make a marked pause before each item. (Pike, 1948, p. 54)

As Pike states, it is unavoidable that the elicitation context sets up a discourse context in which the discourse extends beyond single elicitation items. Thus, prosodic marking of prominence and demarcation (phrasing) due to the imposition of prosodic structure and/or the particular pragmatic context introduced in the current discourse context may induce variation in pitch contours. Particularly dire is if there is a bias in where elicitation items of a given tonal class appear in the order of items. Suppose that the tonemes of a tone language include a falling tone as well as a low tone. Now suppose that within each block, we have one item exemplifying each toneme, and that the low tone items are always elicited at the end of the block. Since low tones often fall utterance-finally due to the interaction of tonal and intonational effects on the pitch contour, we might not be able to distinguish the true falling tone and the low tone.

What would be the levels of an independent variable ORDER (of tonal class)? Suppose we had five tonal classes, and one exemplar of each tone class to use as a substitution item within a block such as the block specified in Table 4. With only five items, we would have 5! = 5 * 4 * 3* 2 * 1 = 120 possible orders of the items within the block!

[11] Thank you to Pat Keating for teaching me this.

[12] The strategy of testing extreme cases is common for error checking and debugging in mathematics and programming, too.

Adding elicitation order as a blocking variable would add 120 times as many blocks. Another problem with blocking by ORDER is that it's not clear that this is the way we would want to define elicitation order—perhaps all that matters is what TONAL CLASS level is last in a block. Then we would only have five levels of ORDER, one level for each of the five elicitation items occupying the last slot in the block. Or alternatively, perhaps what matters is the identity of the TONAL CLASS of two elicitation items that occur next to one another in the elicitation sequence.

An alternative to blocking is *randomization*:

> Randomisation properly carried out ... relieves the experimenter from the anxiety of considering and estimating the magnitude of the innumerable causes by which his data may be disturbed. (Fisher, 1935, p. 44)

Rather than blocking by ORDER, rather than attempting to make sure that there are no biases in ordering with respect to TONAL CLASS by painstakingly fiddling with orders by hand, we *randomize* order of elicitation items within a block to eliminate bias. One way to conceptualize this is the following: for each block, suppose we label each elicitation item within that block with an integer, e.g. 1, 2, 3, ..., 24, 25 if there are 25 items in the block. We then write that number on a slip of paper and put all of the slips in a jar and mix them up. Each time we're ready to elicit a new item within a block, we draw a slip from the jar and read the number written on it. The number on the slip tells us which item to elicit. We then put the slip in the recycling bin (not back in the jar) and elicit the corresponding item.

This concludes our discussion of the second stage in toneme discovery, where (putative) TONAL CLASS is treated as an independent variable. In §2.3.3 below, we illustrate an example of manipulating TONAL CLASS as an independent variable in Kirikiri, focusing on blocking and replication as strategies. In the tutorial [SEE TUTORIAL], we show ways to perform simple randomizations of elicitation order with examples from Kirikiri elicitations.

2.3.3 Example: tonal class as an independent variable in Kirikiri

In §2.2.1, we gave an example of discovering TONAL CLASS as a latent source of variability in the pitch contour in Kirikiri. In this section, we fast forward a bit to one of the first elicitation sessions where we explicitly treated TONAL CLASS as an independent variable and systematically ran exemplars of different proposed tonal classes through a series of substitution frames in N_1N_2 prenominal possessive constructions. This is elicitation session `20111213-kiy-ap-1-framedwordlist`.

The substitution frames are chosen to treat TONAL CLASS of the substitution frame, a single word, as a blocking variable, so each of the five proposed tonal classes for the substitution frame serve as a block for replication for testing the proposed tonal classes of the substitution items (target words). In fact, the experimental design has two experiments in one: we can consider either N_1 to be the substitution item and N_2 to be the substitution frame, or vice versa. When the target word is N_1, it's utterance-initial, and when it is N_2, it is utterance-final, so we have experimental replications over different prosodic positions, as well as replications over different flanking tonal classes. The experimental design for this session, taking N_1 to be the substitution item is given below. The experimental design taking N_2 to be the substitution item is nearly

identical: wherever N_1 occurs in the description of the experimental design, replace it with N_2.

- Research question: How does TONAL CLASS affect pitch contour over a word?
- Strategy: Manipulate TONAL CLASS as an independent variable in different substitution frames.
- Research hypothesis: There are five tonal classes in Kirikiri.
- Linking hypothesis: We assume that distinct pitch contours imply distinct levels of TONAL CLASS, as in Hypothesis 2. Moreover, we assume that in one or more substitution frames, the pitch contour for a proposed tonal class is distinct from pitch contours for other proposed tonal classes.
- Experimental unit: individual elicitation items
- Explanatory variables: TONAL CLASS (of target word N_1), with levels $T1$, $T2$, $T3$, $T4$, $T5$
- Confounding variables
  - WORD LENGTH (of target word, in syllables): 2 syllables (fixed at this level)
  - SYNTACTIC STRUCTURE: prenominal possessive phrases (fixed)
  - TONAL CLASS OF FRAME WORD N_2, with levels $T1$, $T2$, $T3$, $T4$, $T5$ ... (blocking variable)
  - PROSODIC POSITION (of substitution item): utterance-initial (fixed; utterance-final if N_2 is the target word)
  - WORD LENGTH (of substitution frame in syllables): 2 syllables (fixed)
- Dependent variable: pitch contour over the target word N_1

A list of the five different blocks for each of the two sub-experiments, one with N_1 as the target word, and one with N_2 as the target word, is given in Table 5. The full layout of tonal sequences in the experimental design is given in Table 6. The levels of the independent variables N_1 TONE and N_2 TONE are fully cross-classified. We'll see another way to think about these kinds of fully cross-classified designs in §3.3.2. The exemplar words chosen for each TONAL CLASS for N_1 and N_2 are given in Table 7.

In Figure 8, we show f0 contours for all words, whether they were utterance-initial (N_1) or utterance-final (N_2). All f0 contours plotted in this section are *time-normalized*. For each f0 contour, mean f0 was extracted from 30 evenly spaced frames over the word, so, for instance. The effect of this normalization is to improve comparability of the shape of f0 contours over words that may have different durations and/or have been uttered at different speech rates. When we differentiate the f0 contours by TONAL CLASS, it is clear that there is structure in the mess of f0 contours shown in the plot.

When we separately plot time-normalized f0 contours for N_1 and N_2, see Figures 9a and 9b, the huge effect that TONAL CLASS has on variability in the f0 contour is quite apparent. Even though we haven't separated f0 contours by SUBSTITUTION FRAME, the contours within each tonal class in each figure have very similar shapes and pitch ranges. However, the f0 contours for the same TONAL CLASS across substitution frames can be quite different. For instance, T2 in N_1 shows a large fall at the end of the word, but the same tone in N_2 has have a large pitch rise to the end of the word, and T3 in N_1 has a large early peak, while T3 in N_2 has no large peak.

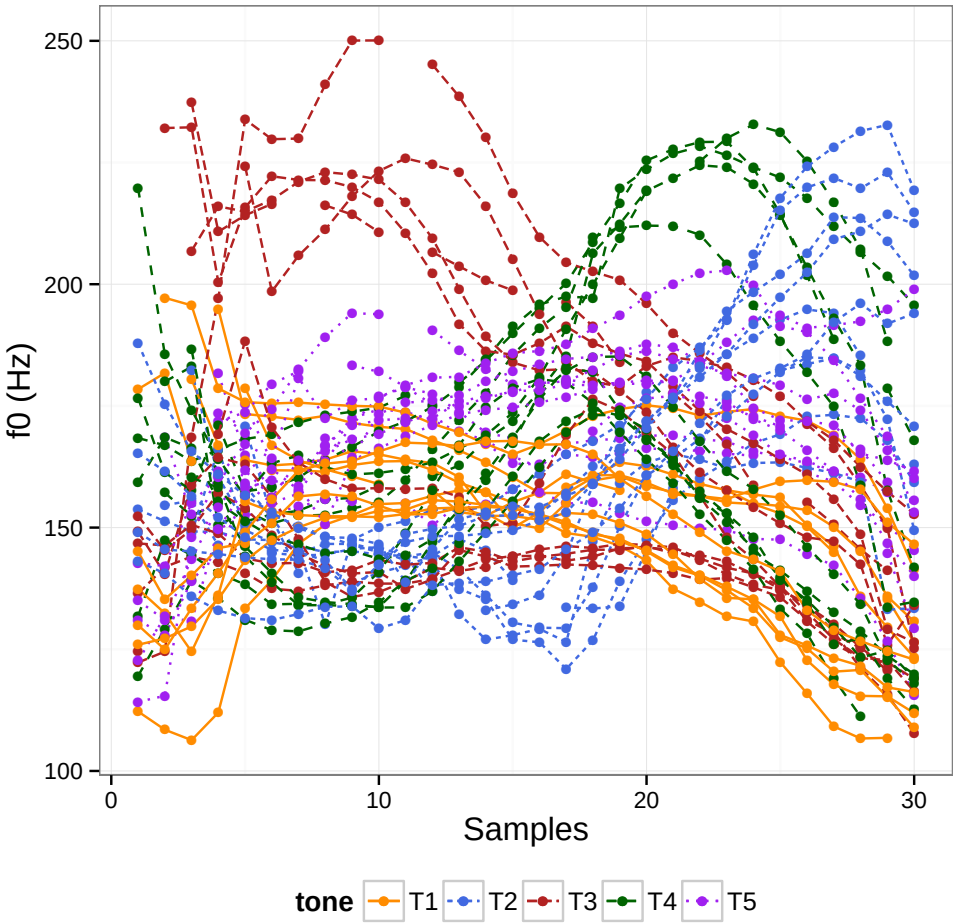
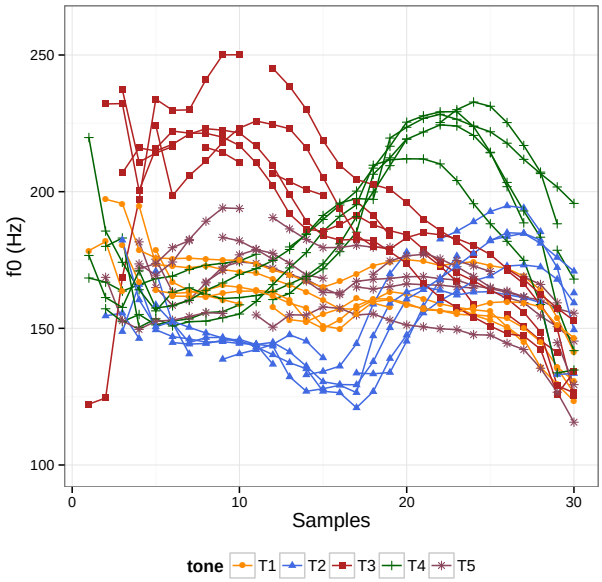
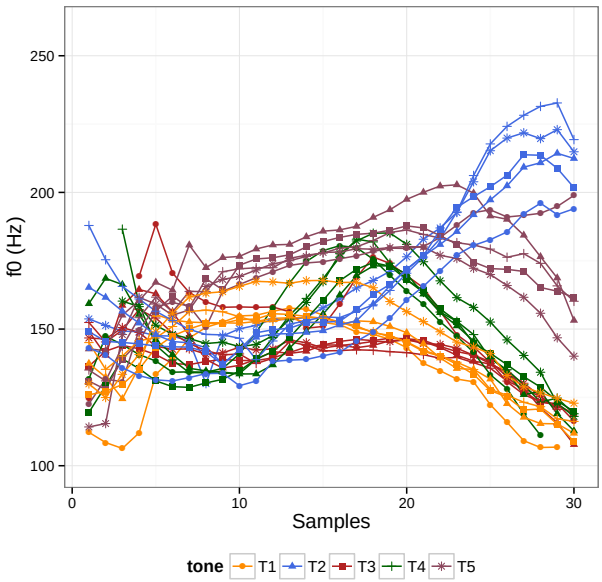


Figure 8: A plot of time-normalized f0 contours from all target words elicited in the session `20111213-1-kiy-ap-framedwordlist` for both N_1 and N_2. Tone class is indicated by color and line type. The x-axis indicates *samples*, not an absolute time scale since the f0 contours are time-normalized. For each word, 30 evenly spaced samples were taken from the f0 contour.



(a) Time-normalized f0 contours for N_1 only.



(b) Time-normalized f0 contours for N_2 only.

Figure 9: Time-normalized f0 contours plotted separately for N_1 and N_2.

Block	IV	Levels
N_2: T1	N_1 TONE	T1, T2, T3, T4, T5
N_2: T2	N_1 TONE	T1, T2, T3, T4, T5
N_2: T3	N_1 TONE	T1, T2, T3, T4, T5
N_2: T4	N_1 TONE	T1, T2, T3, T4, T5
N_2: T5	N_1 TONE	T1, T2, T3, T4, T5
N_1: T1	N_2 TONE	T1, T2, T3, T4, T5
N_1: T2	N_2 TONE	T1, T2, T3, T4, T5
N_1: T3	N_2 TONE	T1, T2, T3, T4, T5
N_1: T4	N_2 TONE	T1, T2, T3, T4, T5
N_1: T5	N_2 TONE	T1, T2, T3, T4, T5

Table 5: Manipulated independent variables (IVs) in the experimental design for elicitation session `20111213-1-kiy-ap-framedwordlist`. TONAL CLASS for N_1 (N_2), the target word, is varied over the 5 putative tonal classes within each block of varying the tonal class of the substitution frame word, N_2 (N_1).

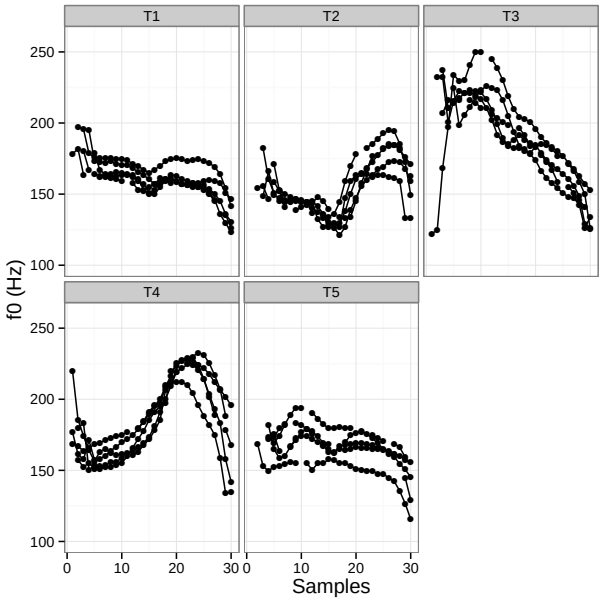
	+T1	*+T2*	*+T3*	*+T4*	*+T5*
T1	T1 + T1	T1 + T2	T1 + T3	T1 + T4	T1 + T5
T2	T2 + T1	T2 + T2	T2 + T3	T2 + T4	T2 + T5
T3	T3 + T3	T3 + T2	T3 + T3	T3 + T4	T3 + T5
T4	T4 + T1	T4 + T2	T4 + T3	T4 + T4	T4 + T5
T5	T5 + T1	T5 + T2	T5 + T3	T5 + T4	T5 + T5

Table 6: Manipulation of TONAL CLASS as an independent variable in a sequence of two tones. The levels of the independent variables N_1 TONE and N_2 TONE are cross-classified so that all possible combinations of tones ($5 \times 5 = 25$ in total) are included.

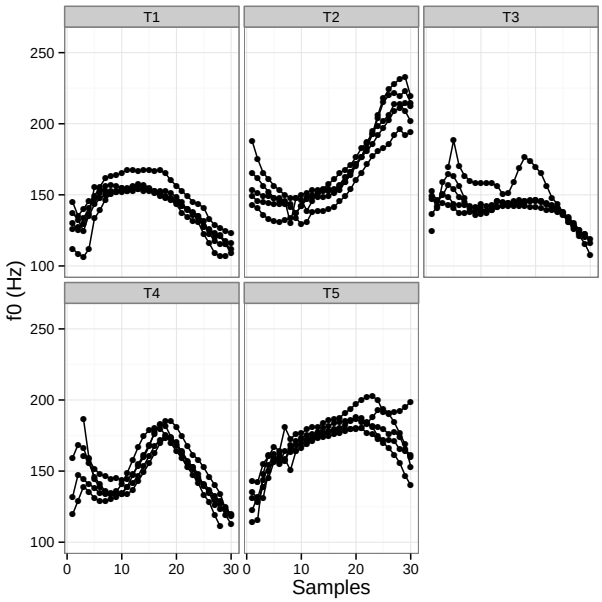
Tone	Noun	Word
T1	N1	parai ‘bandicoot’
	N2	giru ‘elbow’
T2	N1	kaza ‘gecko’
	N2	ora ‘tongue’
T3	N1	fivaa ‘snail’
	N2	faro ‘groin’
T4	N1	naraa ‘wasp’
	N2	kwawaa ‘chin’
T5	N1	tava ‘catfish’
	N2	koree ‘string’

Table 7: Wordbank for $N_1 + N_2$ data set from `20111213-1-kiy-ap-framedwordlist`

Finally, Figures 10a and 10b emphasize just how much variability in the pitch contour over the word can be attributed to TONAL CLASS. Within each tonal class, f0 contours for N_1 or N_2 are remarkably similar.



(a) f0 contours for N_1 by tonal class



(b) f0 contours for N_2 by tonal class

Figure 10: Time-normalized f0 contours plotted separately for N_1 and N_2 for each putative tonal class.

3 Using experimental design principles to generalize beyond Pike (1984)

In §2.2 and §2.3, we've seen how Pike's toneme discovery procedure can be recast in terms of general principles of experimental design. In this section, we show how understanding Pike's procedure in terms of experimental design allows us to relate Pike's elicitation methodology to other methods in tonal elicitation, via: (1) applying the same principles to a different set of variables (§3.1), including the special case of considering alternative ways of defining variables within Pike's procedure (§3.2), and (2) applying the same principles to different research questions (§3.3).

3.1 Generalizing the set of independent variables

One simple generalization beyond Pike's procedure is the consideration of additional independent and dependent variables. If we keep our research question about toneme discovery, then any additional independent variables must all be confounding variables, since our only explanatory variable is TONAL CLASS for toneme discovery. Measuring additional dependent variables for toneme discovery presupposes additional linking hypotheses between tonemes and phonetic parameters, such as aspects of voice quality like parameters indexing creakiness and breathiness. We consider such dependent variables beyond pitch, as well as pitch-based variables, in §4. In this section, we focus on expanding the set of independent variables considered.

Another independent variable is ELICITATION SESSION. Sometimes speakers may produce different pitch contours for the same target word in the same context (as captured by our set of confounding variables) from elicitation session to elicitation session. Perhaps the consultant was tired during one of the sessions and misunderstood what was intended to be elicited; perhaps intervening sessions affected the discourse context of the elicitation session–who knows? Some sources of variability in pitch contour generation we might never be able to get a grip on, but we may still move some variability from the *unexplained* partition to the *explained* variability partition by blocking by aggregate confounding variables like ELICITATION SESSION. ELICITATION SESSION acts as a cover term for the factors in play that vary over time between elicitation sessions, and by repeating elicitations across sessions, we add additional replication into the experimental design.

We enumerate some other potential confounding variables within a subject for the effect of TONAL CLASS on pitch contours over a word in Table 8. We provide a couple of example levels for each variable, as well as sample references which either manipulated the variable or discuss the relation between the variable and variability in the pitch contour. The listed confounds range from physiological factors to phonological and morphosyntactic grammatical factors to discourse and pragmatic factors.

It may seem counterintuitive to consider grammatical factors such as CASE to be confounding, since the underlying source of pitch variability is grammatical tone marking. But if the research question is to uncover how tonemic contrast—how *lexical* TONE CLASS—affects the pitch contour, then grammatical tone marking is indeed a confound. Take case-marking in Maasai (Kenya), e.g. èndérònì 'rat.NOM' and èndèrónì 'rat.ACC' (Hyman, 2011, p. 203). Suppose you had been eliciting simple determiner phrases, by coincidence in nominative case, e.g. in isolation or in response to "What saw the

man?", and mixed in determiner phrases in accusative case in the same block—without including CASE as a confounding variable—e.g. in response to the question "What did the man see?". In that case, the variability between èndérònì 'rat.NOM' and èndèrónì 'rat.ACC' would be bundled into unexplained variability and cloud the relation between the putative toneme sequence LHLL and the pitch contour over the word. If one were instead investigating the relation between CASE and the pitch contour over the word, then TONE CLASS would be a confound. What is an explanatory variable and what is a confounding variable always depends on the research question. To abuse the old saw about junk and treasure, "one experiment's confounding variable is another experiment's explanatory variable."

3.2 Parametrizing variables

A special case of generalizing the set of variables under consideration (§3.1) is when we don't add any new variables, but we change the definition of some of the variables. There are two main ways in which we might do this:

- Coarsening the partition of the possible instantiations of the variable
- Refining the partition of the possible instantiations of the variable

When we coarsen the partition, we reduce the number of levels for a variable, merging levels with one another. When we refine the partition, we increase the number of levels for a variable, splitting levels.

One example of coarsening the partition is given in Table 9. This is adapted from the study of contextual tonal variation in Mandarin in (Xu, 1997, p. 70). The explanatory variable of TONAL CLASS of the target syllable had 4 levels, following the 4-way tonal contrast in Mandarin (abstracting away from the neutral fifth tone). However, the explanatory variable of the PRE-TARGET TONE, i.e. the tone class of the syllable preceding the target syllable, collapsed the 4-way distinction for Mandarin tonal classes into a 2-way distinction based on the tonal offset: both Tone 1, a high tone, and Tone 2, a rise, were classified as having a high offset, and both Tone 3 (low) and 4 (fall) were classified as having a low offset. In autosegmental-theoretic terms (Goldsmith, 1990, 1976), one might say that the new variable assumes that contour tones are treated as tonal sequences (fall = HL, rise = LH) and pays attention only to the tone at the right edge of the syllable.

Another example of coarsening the partition defined by a variable comes from Keating et al. (2011), a cross-linguistic study of the acoustic parameters involved in distinguishing phonation types. Here, for the purposes of standardization for comparison with other languages, the 7-way tonal contrast in the White Hmong data from Esposito et al. (2009); Esposito (2012) was coarsened into a 3-way contrast capturing the rough location of the tone within the pitch range—either high, mid, or low—for the same data, for the purposes of the cross-linguistic comparison in Keating et al. (2011), see Table 10.

An example of refinement of the partition induced by a variable would be the reverse of the mapping for Mandarin tones in Table 9: rather than collapsing a 4-way distinction into a 2-way distinction, one would refine a 2-way distinction into a 4-way distinction. Another instance of refinement would be the addition of a level for TONAL CLASS upon the discovery of evidence for a new tonal class in the course of fieldwork.

Variable	Example levels	Reference
Speech rate	slow, fast	Fougeron and Jun (1998); Gandour et al. (1999); Kuo et al. (2007)
Speaker pitch range	small, large	Baken and Orlikoff (2000a, p. 175)
Background noise	absent, present	Zhao and Jurafsky (2009)
Consonant voicing	Voiced, unvoiced	Hombert (1978)
Vowel quality	High, low	Connell (2002); Hombert (1978)
Prosodic position	Utterance-final, phrase-medial	Maddieson (1978, p. 45-46), Hayes (1989), Shattuck-Hufnagel and Turk (1996), Zsiga and Nitisaroj (2007), (Gussenhoven, 2004, Ch. 6)
Declination	Phrase-initial, phrase-final	(Gussenhoven, 2004, Ch. 6)
Downstep	Post-trigger	(Gussenhoven, 2004, Ch. 6)
Tonal coarticulation and sandhi	preceding H, following L	Xu (1997); Chen (2000); Kuo et al. (2007)
Lexical frequency	low, high	Zhao and Jurafsky (2009)
Person	1sg, 3sg	Hyman (2011, p. 203)
Tense/aspect	present, past	Hyman (2011, p. 203)
Negation	present, absent	Hyman (2011, p. 203)
Case	nominative, accusative	Hyman (2011, p. 203)
Emphasis level	1 (mumble), 10 (shout)	Liberman and Pierrehumbert (1984)
Focus	Contrastive focus, out-of-blue focus	Katz and Selkirk (2011), Eady et al. (1986), Jun (2005)
Giveness	Given, new	Katz and Selkirk (2011)
Speech style	spontaneous, reading	Fernald and Simon (1984), Baken and Orlikoff (2000a, p. 175-176)
Phonetic accommodation	pre-exposure, post-exposure	Babel and Bulatov (2012); Babel (2012)

Table 8: Some additional examples of potential confounding variables in toneme discovery, arranged roughly in order from variables related to: physiology and the speech signal, phonetics and phonology, morphosyntax, and discourse and pragmatics.

Tone	Old level	Mapping	New level
Tone 1	High	High $\mapsto$ H	H
Tone 2	Rise	Rise $\mapsto$ H	H
Tone 3	Low	Low $\mapsto$ L	L
Tone 4	Fall	Fall $\mapsto$ L	L

Table 9: A coarsening of the levels for TONAL CLASS in Mandarin from Xu (1997).

Tone	Old level	Mapping	New level
b-tone	High-rising	High-rising $\mapsto$ H	H
null-tone	Mid	Mid $\mapsto$ M	M
s-tone	Low	Low $\mapsto$ L	L
j-tone	High-falling	High-falling $\mapsto$ H	H
v-tone	Mid-rising	Mid-rising $\mapsto$ M	M
m-tone	Low-falling	Low-falling $\mapsto$ L	L
g-tone	Mid-falling	Mid-falling $\mapsto$ M	M

Table 10: Coarsening of levels for TONAL CLASS in White Hmong. Note: the g-tone is high-falling for females, but we abstract away from that here. Old levels come from Esposito et al. (2009); Esposito (2012). New levels come from cross-linguistic study of phonation contrasts in Keating et al. (2011)

Those two examples of refinement both involve increasing the number of levels to some finite number, but refinements may also involve mapping from a set of a finite number of distinctions, e.g. 4 levels, to a set of potentially infinitely many distinctions, i.e. the set of *real numbers*.[13] An example of this kind of refinement would be changing a LENGTH variable from counting syllables, e.g. 1 syllable, 2 syllables ..., to measuring absolute time, e.g. 343.25 milliseconds, 692.11 milliseconds. This kind of refinement might seem intuitively more drastic than refining a 2-way tonal distinction into a 4-way one, and it is: it's a change in variable type (Stevens et al., 1937), similar to a change in type in type-theoretical semantics (Gamut, 1992, Ch. 4),(Carpenter, 1997, Chs. 2, 3).

3.3 Generalizing to different research questions

A more drastic way to generalize beyond Pike's procedure is to apply principles of experimental design to other research questions. In this section, we give examples of research questions: (1) treating tone as a dependent variable rather than an independent variable (§3.3.1), (2) exploring the mapping from underlying tonemes to surface tones in Thlantlang Lai as described in Hyman (2007) (§3.3.2), (3) examining phonetic tonal sandhi, i.e. tonal coarticulation in White Hmong (§3.3.3), and (4) uncovering evidence for a tonal case marker in Samoan (§3.3.4).

3.3.1 Example: tone as a dependent variable

Thus far in this paper, we've considered TONAL CLASS as an independent variable manipulated by the fieldworker, but we've never considered TONAL CLASS as a dependent variable. This is not because TONAL CLASS cannot be treated as a dependent variable, but simply due to the nature of our research questions—we've focused on making hypotheses about possible TONAL CLASSES and their reflexes in the pitch contour and refining these hypotheses.

There are two main situations in which tone might appear in the dependent variable:

- in explorations of how tonal contrast is produced and perceived

[13]The set of *real numbers* contains numbers like 3.0, 1.542, π, 2.9, 2.99, 2.999, 2.9999999999999999....

- in explorations of phonological allophony and alternation

Some example research questions about exploring the dimensions of tonal contrast are:

- What effect does TONAL CLASS have on the pitch contour over a word?
- What parameters in the speech signal are available for discriminating different tonal classes?
- What cues in the speech signal do listeners use to identify tones?

Some examples of work along these lines appear in Connell (2000) (perception), Khouw and Ciocca (2007) (perception), and DiCanio (2009) (production).

In explorations of phonological allophony and alternation, tone makes an appearance in the dependent variable because the mapping between underlying tonemes and surface tones (or between surface tones) is of primary importance. UNDERLYING FORM is manipulated as an explanatory variable, and the dependent variable is the surface form. Note that there must be a linking hypothesis about the mapping from observables (perhaps the pitch contour over a word) to surface tones in such an elicitation experiment. We present an example of the tonal fieldwork exploring phonological allophony in the next section, §3.3.2.

3.3.2 Example: tonotactics in Thlantlang Lai (Hyman, 2007)

In tonal fieldwork, as soon as the fieldworker is far along enough in toneme discovery to manipulate TONAL CLASS as an explanatory independent variable, a natural set of follow-up research questions is to pursue further detail in understanding tonal allophony and alternation. Here, we take Hyman (2007)'s work on Thlantlang Lai (Tibeto-Burman Kuki-Chin, Myanmar) tonal allophony as an example. Hyman (2007) examined the surface tone sequences of prenominal possessive phrases, e.g. râal ràng 'enemy's horse'.

We can cast this in terms of the following experiment:

- Research question: Are there tonotactic n-gram restrictions in Thlantlang Lai?[14]
- Strategy: Control any variables suspected to induce variation in surface realization of underlying tones. (Tonotactic restrictions, perhaps formulated as constraints, are not included in this set of variables, since whether or not Thlantlang Lai has such restrictions is yet unknown).
- Research hypothesis: There are tonotactic bigram restrictions in Thlantlang Lai.
- Linking hypothesis: We assume some mapping between the pitch contour over the word and the surface tones.
- Experimental unit: n-gram over Thlantlang underlying tones
- Explanatory variables: UNDERLYING TONAL CLASS of each noun: N_1 TONE, N_2 TONE, ..., N_n TONE, with levels H, L, HL
- Confounding variables
  - n-GRAM LENGTH: bigrams, trigrams, 4-grams, ... (blocking variable)

[14] An *n-gram* is a sequence of discrete units of length n, e.g. a 2-*gram* or *bigram* is a sequence of length 2.

  - SYNTACTIC STRUCTURE: prenominal possessive phrases (fixed at this level)
  - WORD LENGTH (syllables): 1 syllable (fixed at this level)
  - PROSODIC POSITION (of n-gram): isolation (fixed)
- Dependent variable: surface tone sequence of the n-gram

This design should look familiar. First, we exploit the same strategy here for tonotactic constraint discovery as we did for toneme discovery when we treated TONAL CLASS as a latent variable in Kirikiri in §2.2: we try to control for everything we suspect may affect the dependent variable and then if we still have residual variance left over, we attribute that to hidden structure—in this case, tonotactic constraints. Thus, we have a number of blocking variables, most which we fix to a single level, e.g. SYNTACTIC STRUCTURE is held constant to be prenominal possessive phrase.

Like in Pike's experimental design in§2.3, an explanatory variable is (underlying) TONAL CLASS. However, here we have multiple explanatory variables, since we are interested in sequences of tones—we have one explanatory variable per noun, e.g. for bigrams, we have two explanatory variables, each with three levels: HL, H, and L. Table 11 shows the explanatory variables for each n-gram block and Table 12 shows the words used to create the bigram sequences. Each n-gram block provides a replication for discovering bigram tonotactic restrictions.

Block	IV	Levels
$N_1 + N_2$	N_1 TONE	HL, H, L
	N_2 TONE	HL, H, L
$N_1 + N_2 + N_3$	N_1 TONE	HL, H, L
	N_2 TONE	HL, H, L
	N_3 TONE	HL, H, L
$N_1 + N_2 + N_3 + N_4$	N_1 TONE	HL, H, L
	N_2 TONE	HL, H, L
	N_3 TONE	HL, H, L
	N_4 TONE	HL, H, L
⋮	⋮	⋮

Table 11: Independent variables for experimental design for tonotactics in Thlantlang Lai (Hyman, 2007, (18)). Abbreviations: IV = (explanatory) independent variable, N = noun, H = high, L = low.

Because there are multiple explanatory variables in this design, another parameter of the design is how to investigate the interactions of the explanatory variables. The usual choice in fieldwork in this situation is to investigate all possible ways the explanatory variables may vary together—to choose a *factorial* design, where each level of each explanatory variable is cross-classified with another, e.g. since there are three tonal classes in Thlantlang Lai, there are $3 \times 3 = 9$ possible factor combinations (*treatments*) for a bigram, as shown in Table 14.

Hyman (2007)'s experimental design for Thlantlang Lai tonotactics is not at all unusual in tonal fieldwork. For instance, Hyman (1985) examined the same kinds of constructions with the same kind of experimental design in Bamileke-Dschang (Niger-Congo Bantoid, Cameroon); see Table 1 in Hyman (1985), which lists bisyllabic N_1+N_2

Tone	Noun	Word
HL	N1	râal ‘enemy’
	N2	zôong ‘monkey’
H	N1	kóoy ‘friend’
	N2	vók ‘pig’
L	N1	bòoy ‘chief’
	N2	ràng ‘horse’

Table 12: Wordbank for N_1 + N_2 data set for Thlantlang Lai tonotactics (Hyman, 2007, (19))

	$+HL$	$+H$	$+L$
HL	HL $+HL$	HL $+H$	HL $+L$
H	H $+HL$	H $+H$	H $+L$
L	L $+HL$	L $+H$	L $+L$

Table 13: Factorial design for bigrams N_1N_2 in Thlantlang Lai (Hyman, 2007, (18)). The levels of the two explanatory variables, N_1 TONE and N_2 TONE, are cross-classified so that all possible combinations of tones ($3 \times 3 = 9$ in total) are included. The levels of N_2 TONE are italicized so that they are distinguishable from those of N_1 TONE. Abbreviations: N = noun, H = high, L = low.

associative constructions, e.g. èfɔ̀ màndzwì ‘chief of leopards’. The factorial design in testing all possible sequences of n-grams for some n is used all the time in production experiments on phonetic and phonological tonal sandhi, e.g. Xu (1997).

In fact, *any paradigmatic elicitation typically has a factorial design.* Consider the elicitation of verbal morphology. Possible explanatory factors and their corresponding levels might be TENSE (past, present, future), PERSON (first, second, third), and NUMBER (singular, plural). To fill out the paradigm for a verb, we'd do a factorial experiment of TENSE × PERSON × NUMBER, for $3 \times 3 \times 2 = 18$ possible treatments, one for each table cell in the factorial experimental design shown in Table 14.

3.3.3 Example: tonal realization in White Hmong

A similar factorial design for tonal bigrams, but for studying the phonetic realization of tones in White Hmong (Hmong-Mien, China) is given below. Since the focus is the acoustic variation induced by tonal classes, there is a large and detailed set of acoustic dependent variables.

- Research question: How are tones in White Hmong acoustically realized?
- Strategy: Control some known sources of variability in tonal realization and manipulate others to study a selected range of tonal variability.
- Research hypothesis:
- Linking hypothesis: Acoustic dimensions relevant for tonal discrimination in the production of White Hmong tones include f0-based parameters and various spectral parameters.

Past tense		
	Number	
	Singular	*Plural*
1st	1sg.past	1pl.past
2nd	2sg.past	2pl.past
3rd	3sg.past	3pl.past
Present tense		
	Number	
	Singular	*Plural*
1st	1sg.pres	1pl.pres
2nd	2sg.pres	2pl.pres
3rd	3sg.pres	3pl.pres
Future tense		
	Number	
	Singular	*Plural*
1st	1sg.fut	1pl.fut
2nd	2sg.fut	2pl.fut
3rd	3sg.fut	3pl.fut

Table 14: Factorial experimental design in verbal morphology elicitation for the set of explanatory variables TENSE (past, present, future), PERSON (first, second, third), and NUMBER (singular, plural). The design for the variable interaction is factorial since the variables are fully cross-classified and we have 3 (TENSE) × 3 (PERSON) × 2 (NUMBER) = 18 elicitation items.

- Experimental unit: elicited sentences
- Explanatory variables
  - N_1 TONE: b, n, s, j, g, m, v
  - N_2 TONE: b, n, s, j, g, m, v
- Confounding variables
  - PROSODIC POSITION: isolation, sentence-medial
  - CARRIER PHRASE: fixed with two phrases, with target words randomly assigned to one of the two phrases
  - SEGMENTAL FEATURES OF WORDS: fully [+sonorant] (fixed)
  - CV SKELETON: CVV (fixed)
  - PRAGMATIC CONTEXT: out of the blue (fixed)
- Dependent variables
  - mean fundamental frequency
  - syllable onset fundamental frequency
  - syllable offset fundamental frequency
  - mean spectral tilt
  - mean harmonic-to-noise ratio

Block	IV	Levels
Isolation	N_1 TONE	b, n, s, j, g, m, v
	N_2 TONE	b, n, s, j, g, m, v
Sentence-medial	N_1 TONE	b, n, s, j, g, m, v
	N_2 TONE	b, n, s, j, g, m, v

Table 15: Experimental design for acoustic study of tones in White Hmong.

3.3.4 Example: tonal case marking in Samoan

Moving back upwards towards the morphosyntax-prosody interface, this section gives an example of a 2×2 factorial design examining the effect of the interaction of CASE-MARKING PATTERN (absolutive-oblique, ergative-absolutive) and WORD ORDER (VSO, VOS) on the f0 contour in Samoan (Polynesian, Samoa), with the goal of examining the hypothesis that there is a high tone at the left edge of absolutive arguments. The experimental design involves minimal sets of sentences, keeping segmental material in test sentences constant except for the segmental case markers for ergative and oblique case. Some factors are controlled for optimizing our chances of observing prosodic realization realized in the f0 contour. First, words are fully sonorant so that the f0 contours is free from segmental perturbation. Secondly, arguments of long length, i.e. many words, are used to allow plenty of segmental material for intonational tonal events to be realized (Bruce, 1977).

- Research question: Does Samoan have a high tone at the left edge of absolutive arguments?
- Strategy: Control any variables suspected to induce variation in surface realization of underlying tones and vary CASE-MARKING PATTERN and WORD ORDER. To support our hypothesis, we must find an *interaction* effect on the intonational realization in the sentence, such that the presence of high pitch peak at the left edge of the second argument occurs when the levels of the two factors interact such that the second argument has absolutive case.
- Research hypothesis: Samoan has a high tone at the left edge of the absolutive argument.
- Linking hypothesis: A high tone in Samoan is realized as pitch peak realized at the edge of a prosodic word.
- Experimental unit: elicited sentence
- Explanatory variables
  - CASE-MARKING PATTERN: absolutive-oblique, ergative-absolutive
  - WORD ORDER: VSO, VOS
- Confounding variables
  - CONSTITUENT LENGTH: long (fixed)
  - COORDINATION: absent (fixed)
  - SEGMENTAL FEATURES OF WORDS: fully [+sonorant] (fixed)
  - STRESS PATTERN: primary stress on penultimate mora (fixed)

  - CV SKELETON: CVCVCV (fixed)
  - PRAGMATIC CONTEXT: out of the blue
- Dependent variable: presence of high pitch peak at the left edge of the second argument

	erg-abs	abs-obl
VSO	V-erg-abs	V-abs-obl
VSO	V-abs-erg	V-obl-abs

Table 16: 2x2 factorial design for Samoan absolutive case marking elicitation experiment, with CASE-MARKING PATTERN (erg-abs, abs-obl) fully crossed with WORD ORDER (VSO, VOS). Abbreviations: erg = ergative, abs = absolutive, obl = oblique, V = verb, S = subject, O = object.

4 The dependent variable: pitch and beyond

We've been vague about the nature of the dependent variable for the most part in this paper, describing it as something like the pitch contour over the word. Yet, the dependent variable is our window into the hidden structure of tonal concepts, so all our conclusions rely on our assumptions about it. In this section, we focus on becoming precise about the dependent variable. We'll consider questions like: what's the difference between pitch and fundamental frequency (f0)? What can examining f0 off of recordings get you beyond what you get from a tonal transcription and why is it that sometimes what you hear doesn't seem to match what you see on the computer's pitch track? How might we refine our linking hypothesis mapping TONAL CLASS to the pitch contour over a word?

First we review some fundamentals about pitch (§4.1) and discuss the role of transcriptions and recordings in measuring the dependent variable in §4.2. Then we discuss some fundamentals of how pitch tracking algorithms work and how to use pitch tracks from audio recordings as a tool to aid elicitation alongside the perceptions of the ear in §4.3. We close by focusing on ideas for refining linking hypotheses mapping TONE CLASS to f0- and pitch-based parameters and voice source parameters beyond f0, including parameters related to phonation (§4.4).

4.1 The basics of fundamental frequency and pitch

Fundamental frequency and pitch are often used interchangeably, but if we are being precise about terminology as we'll be here, they are distinct terms. *Pitch* refers to an auditory percept closely related to the rate of vocal fold vibration, or equivalently, the glottal pulse rate, which is called the *fundamental frequency*, f0.[15] The measurement

[15]Sometimes f0 is capitalized as F0, but keeping it lower case helps reinforce the fact that the physics of fundamental frequency has very little to do with the physics of *formants*, vocal tract resonances critical in the description of vowel quality, which are standardly abbreviated as F1, F2, etc. for first formant, second formant.

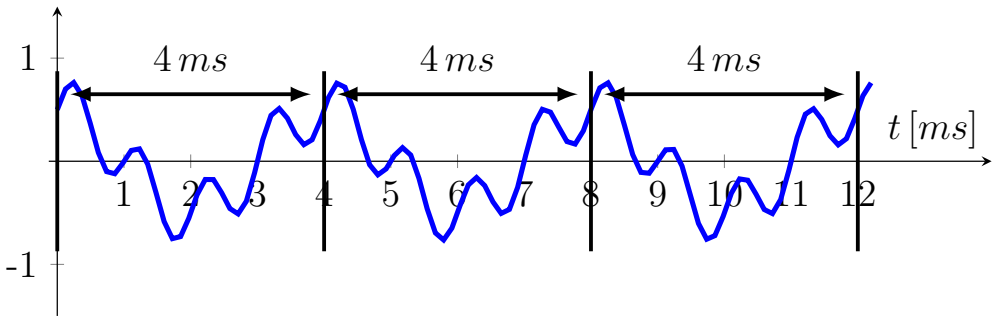


Figure 11: Measuring fundamental frequency (f0) from the speech signal waveform. The horizontal x-axis shows time (t), in milliseconds (ms) and the vertical y-axis shows the amplitude of the waveform, which is related to how loud the sound is. Here, there are two three of the waveform shown, one from 0 to 4 ms, another from 4 to 8 ms, and the third from 8 to 12 ms. The waveform cycle repeats every 4 ms, so the *period*, the length of time of one cycle for the waveform (or equivalently, the duration of one glottal pulse), is 4 ms. The fundamental frequency is the rate of repetition, $\frac{1\,cycle}{4\,ms} \times \frac{1000\,ms}{1\,second} = 250\,cycles/s$. We call the unit [cycles/s] *Hertz* (Hz) so f0 is 250 Hz.

of f0 from the acoustic speech signal is shown in Figure 11, which shows three cycles of a train of repeating glottal pulses with cycle lengths or a *period* of 4 milliseconds.

Pitch is not something that can be directly extracted from a recording of speech because it is an interpretation of speech sounds that is mediated by the auditory pathways of the nervous system. What we *can* extract from recorded speech is fundamental frequency, as it is an acoustic parameter of the speech signal. Although we might refer to the timecourse of fundamental frequency estimated from a recording as the *pitch contour*, this is strictly a colloquial term for what is more accurately referred to as the *fundamental frequency contour* or *f0 contour*. When we write down tonal transcriptions during an elicitation session, we are transcribing the perceived pitch contour, but when we examine the output of an f0 estimation algorithm on a computer, we're examining the f0 contour extracted from the speech signal, not the pitch contour.

The time course of fundamental frequency over a word is certainly informative about the time course of pitch over a word and vice versa, but the relation between the two is not transparent. This is because: (1) language experience tunes the way the auditory system processes f0 and (2) the relevant context for interpreting pitch in speech is not fully understood.

The effect of language experience on pitch perception in human speech

Evidence supporting that language experience tunes pitch perception comes from both infant and adult studies. A large body of work in infant speech development supports the idea that infants are born as "universal citizens", able to discriminate the speech sound contrasts of any of the world's languages . They then undergo a perceptual reorganization in the first year of the life and develop language-specific speech perception in response to ambient language input, e.g. see Figure 1 in Kuhl (2004). This tuning of the auditory system has been studied for tonal contrasts: in

Mattock and Burnham (2006); Mattock et al. (2008), both infants exposed to tone languages (Mandarin, Cantonese) and non-tone languages (French, English) as their native language showed behavioral evidence that they can discriminate Thai low vs. rising lexical tones at 4 and 6 months. However, by 9 months, French and English infants did not discriminate the tones, while Mandarin and Cantonese infants still did.

A more direct source of evidence for language-dependent pitch processing comes from electrophysiological studies studying the brainstem's frequency following response—very roughly speaking, the human neural pitch tracking machinery (see Krishnan and Gandour (2009) for a review). This body of studies found that Mandarin speakers showed higher-fidelity and more robust encoding of pitch contours than English native speakers when listening to Mandarin tonal stimuli (Krishnan et al., 2005). However, this Mandarin speaker pitch processing advantage was absent when listeners were presented with pitch contours that were linear ramps—straight line segments—rather than the curvilinear trajectories found in natural tone languages (Xu et al., 2006). Together, these studies suggest that Mandarin speakers have an advantage in pitch processing of speech relative to English speakers, but only for pitch contours within their language experience.

The implication of language-dependent pitch perception for tonal fieldwork is that fieldworkers cannot assume that their perception of pitch in an elicited utterance is representative of the consultant's or anyone else's perception.

The role of context in pitch perception

In addition to pitch being in the ear of the beholder, pitch perception is also integrated with perception of other qualities of the speech signal. This means that we miss important information when we factor out pitch perception as an isolated process in perceiving the speech signal. For instance, intensity and spectral properties of sound affect pitch perception (Baken and Orlikoff, 2000b, p. 146). Moreover, Rose (1988); House (1990) found evidence that pitch perception in speech is strongly influenced by segmental context. House, for instance, proposed a model of pitch perception in speech where processing resource constraints are in play as a listener tracks both spectral characteristics of speech, e.g. vowel quality, as well as pitch movements. In areas of rapid change in intensity and spectral composition, such as at the onset of a vowel, a falling f0 contour could be perceived as a level contour, but during stable regions of intensity and spectral composition, such as in the middle of a vowel, the same falling f0 contour could be perceived as falling—in short, the alignment of the f0 contour to the segmental string matters for pitch perception. This is an example of "vertical" or paradigmatic context for pitch perception.

One example of "horizontal" or syntagmatic context for pitch perception comes from Wong and Diehl (2003), which showed the dramatic effect of information about the pitch range in preceding context on biasing the identification of high, mid, and low tones in Cantonese. By raising and lowering the pitch of a mid tone preceding the target tone to be identified, Wong and Diehl (2003) was able to flip listeners between perceiving a high tone (when the preceding mid tone was lowered), mid tone (when f0 of the preceding mid tone was left almost unaltered) or low tone (when the preceding mid tone was raised), even though f0 of the target tone remained unchanged throughout.

In sum, although we can extract f0 from the speech signal independently from other

acoustic parameters, we cannot factor out pitch perception in speech independently from the perception from concurrent perception of other properties of the speech signal , since perception of these properties is integrated. We can also measure f0 at a particular time, without reference to the past or future. But we cannot understand the perception of pitch in speech at some point in time without taking into account other information from the past (or even the future!) that may provide a background against which pitch at the current instant is relativized.

The gap between f0 and pitch might lead the reader to wonder, if what we're interested in understanding is how tonal contrast is defined in a particular language to speakers of that language, and if pitch is what is speakers experience—not f0—then what good is the acoustic data that can be provided by recordings in tonal fieldwork?

4.2 The role of recordings in tonal fieldwork

Acoustic data from recordings can still be very valuable in tonal fieldwork, even if the acoustic record itself does not tell us exactly what native speakers of the language actually perceive. Acoustic data provides an objective, lossless record of elicitation items. By *objective*, we mean that recordings are not filtered through anyone's biased ears—unlike tonal transcriptions, they are unbiased by language experience and assessments of the relevant context for perception. By *lossless*, we mean that the acoustic record preserved in a recording captures all the information available at the time of elicitation; in comparison, a tonal transcription of an elicitation item does not preserve the full detail of even the f0 contour.[16] Because acoustic data is objective, recordings preserve elicitation data in a raw format unbiased by the fieldworker's ear which can be readily compared to other data, perhaps with the application of standardizing transformations. Because acoustic data is lossless, perceptual information can still be generated from recordings long after the recorded elicitation took place. In fact, acoustic data can be used to generate stimuli for valuable perceptual experiments during elicitation [SEE TUTORIAL]. The visualization of f0 contours from recordings can also aid the fieldworker in learning how to produce and perceive tones in the language of study. Visualization of f0 contours has been shown to be valuable in training deaf speakers and second language learners in learning prosody, see Hermes (1998); Hardison (2004) and references within. Thus, recordings can offer an important supplement to tonal transcriptions, especially now that: (1) recording technology is relatively inexpensive and easily portable, and (2) speech analysis software is readily available. [SEE TUTORIAL]

The acoustic record is not a substitute for the fieldworker's ear, however. Pike cautioned against over-reliance on recordings for pitfalls closely related to the very reasons we gave for recordings being useful in tonal fieldwork:

> Various instruments which record speech—for example, the phonograph, the dictaphone, and magnetic-wire or magnetic-tape recorders—can be of considerable aid to the investigator, since they permit the repetition of phrases. The use of such machines, however, has two grave dangers:

[16] If we are being ruthlessly precise, we should say that recordings provide a *near*-lossless record of an elicitation item, since (digital) recorders collect (i.e. sample) information at some finite sampling rate, such that information between the samples is lost. Moreover, recordings don't preserve potentially relevant information for tone perception such as the visual context.

> (1) The investigator tends to deprive himself of hearing the natural range of key and free variation which comes in repetition by the informant and many, therefore, record as different some utterances of tonemes which are functionally the same in spite of temporary slight, free, pitch divergences. Sufficient recordings of repetitions by the informant himself would overcome this danger.
> (2) The investigator is tempted to be too "accurate," that is, to transcribe (just because he can find them with instruments) details which do not reflect the system, but are changes within tonemes. Here the danger can be avoided if the investigator uses such data to describe the tonal variants but for publication of grammatical and phonetic studies uses a written transcription which records only the significant tone units (tonemes). (Pike, 1948, p. 44)

Both dangers are consequences of the objectivity and losslessness of recordings—precisely the properties that make recordings invaluable as an archival tool. Asking a consultant to repeat an elicitation item many times, especially across sessions, is different than replaying a recording of an elicitation item many times. Each time an elicitation item is elicited, its elicited in a different context—each instance of elicitation is a separate replication. In contrast, each instance of playing the recording is a repetition of the same item in the same context. Exposure to different replications and the variability across them helps tune the ear to which dimensions are relevant or irrelevant for tonemic contrast: it's been shown that such variability can improve learning to generalize over speech sound categories in infants and adult second language learners (Rost and McMurray, 2009), (Lively et al., 1993). Moreover, tonal transcription may aid the process of generalization, since lossy tonal transcriptions encode hypotheses about the relevant dimensions of tonemic contrast, abstracting away from hypothesized irrelevant dimensions while preserving hypothesized relevant ones.

Since Pike's day, recording technology has improved dramatically, so there's little barrier to collecting many recordings of different elicitation instances, as he suggests in his first point about the fieldworker exposing him/herself to sufficient variability in the consultant's pronunciations. With such exposure to variability during elicitation sessions and afterwards in reviewing recordings, the fieldworker can move from "universal citizen"-like tone perception, i.e. from being "too accurate", as Pike states in his second point, towards homing in on the primary dimensions of tonal contrast. Thus, as long as the fieldworker continues to rely on his/her own ears while working with recorded data, Pike's dangers are quite surmountable.

There are two other dangers Pike doesn't quite mention which we address in the next section and in supplemental tutorials. First, if the fieldworker blindly accepts the f0 contours provided by speech analysis software as the truth, not only might he/she be "too accurate" in transcription off of a recording, but even worse, he/she may be misled. To help the reader avoid this, the next section, §4.3, dispells some of the magic in computational f0 estimation and provides some background on how f0 detection algorithms work and pointers on interpreting f0 tracks, especially when they can be misleading.

The second danger is the amount of labor involved in processing and analyzing recorded data. Recordings are the most useful when they are processed and analyzed, but a ten minute long recording can easily take a few hours to split into smaller files and segment into words and sounds.[17] Fortunately, there are scripts available to help

[17]See Turk et al. (2006) for an introduction to segmentation in prosodic research.

streamline processing of recorded files [SEE TUTORIAL], and one can also improve efficiency by having an appropriate file management system [SEE TUTORIAL]. One can also cut down on the amount of data from the outset by recording summary, capstone elicitations, which are set up at the end of a chunk of exploratory work in addition to recordings of the entire elicitation session.

4.3 Interpreting f0 contours

In this section, we give a quick overview of methods for *f0 estimation* (also called pitch estimation, pitch tracking, pitch detection, f0 tracking, f0 detection) and give an intuitive explanation of one of the most common computational algorithms used. Then, we catalog some situations in which tone-segment interactions and voice quality can cause deviations from an idealized smooth f0 contour.

4.3.1 f0 estimation

There are many excellent extant overviews of f0 estimation. Ladefoged (2003, Ch. 3) has a short, friendly introduction to pitch analysis geared towards fieldwork. Baken and Orlikoff (2000b, p. 153–167) gives a very readable overview of f0 estimation methods.[18] Hess (1983) is a more technical, classic compendium of f0 estimation algorithms.

Here, we seek only to provide an intuitive explanation of a common component of f0 estimation algorithm to help illuminate when and why f0 tracks can be misleading. *Short-time autocorrelation* is the heart of many f0 estimation algorithms, including one in Praat (Boersma, 1993; Boersma and Weenink, 2010) and closely related methods are used in other widely used algorithms like the RAPT algorithm (Talkin, 1995), which is commonly used in benchmarking other f0 estimation algorithms and used in ESPS XWaves, Wavesurfer, and Toney (this volume).

From Figure 11, it may seem quite easy to pick out how one cycle of the waveform and the period of the waveform. Our eyes are excellent pattern detectors! What the computer sees as the representation of the waveform, though, is a string of numbers, e.g. 0.83, 0.91, 0.93, 0.90, 0.82, ... indicating sound pressure levels over time. One way the computer can detect repeating patterns is to compute the *autocorrelation* of the waveform, a measure of the similarity between waveforms, as we explain below. The basic idea is to see how well the waveform shape correlates with itself as we shift a copy of a chunk of it incrementally forwards in time. We look for the shift at which the self-correlation is highest and take this as the period of the waveform.

Let's take a simple waveform, a square wave (Figure 12), for illustrative purposes. How do we find the period of this waveform with autocorrelation?

First, we select a little chunk of the waveform, boxed in Figure 12 and shown in Figure 13. Then, we incrementally shift this little chunk rightward, sliding along the waveform and measure the amount of overlap between the chunk and the waveform (Figure 14). The fact that we work with small time chunks of the speech signal is why we call the procedure *short-time* autocorrelation. In real speech signals, waveforms hardly demonstrate the perfect repeatability exhibited by the square wave in Figure

[18] Some of the methods described prior to p. 161 are described as analog methods which use electronic circuitry hardware built to estimate f0. While f0 estimation nowadays is done digitally via computational algorithms written in software, the principles described in the analog methods are still widely used in digital methods today.

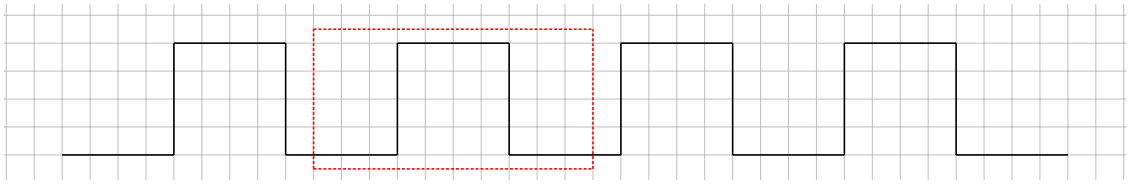


Figure 12: A portion of a square wave. The period of the wave is the amount of time between the repeated squares, which is 8 units. (Each grid square is 1 unit long). We select a chunk of the waveform, indicated in the dashed box.

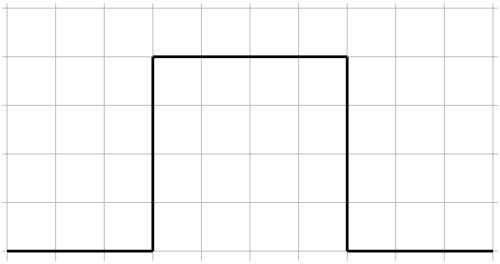


Figure 13: The chosen chunk of the waveform, of length 10.

12, yet f0 is only well-defined on such signals with perfect repeatability. Thus, we work with small time chunks of speech, since we may reasonably assume approximately perfect repeatability within a short time window.

In Figure 15, we display the series of incremental shifts and the amount of overlap between the chunk and the waveform. Since we've used a square wave, the amount of overlap is easy to count up as the number of shaded squares. There is an overlap of 0 through a shift of 4 units, shown in Figure 15a. At a shift of 4 units, the squares begin to overlap by one unit so there is an overlap of 4 units (Figure 15b), and the overlap increases through Figures 15c and 15d until a shift of 8 units, when the squares are superimposed on one another (Figure 15e). From a shift of 9 to 12 units, the overlap in the squares decreases until the overlap vanishes (Figures 15g through 15i).

In Figure 16, we plot the amount of overlap—the autocorrelation—as a function of the shift size. Note that the peak occurs at a shift size of 8 units. This gives us an estimate of 8 units for the period, which is equivalent to an f0 of 1/8 Hz if we take each unit to be a second. This estimate is exactly what we expect from

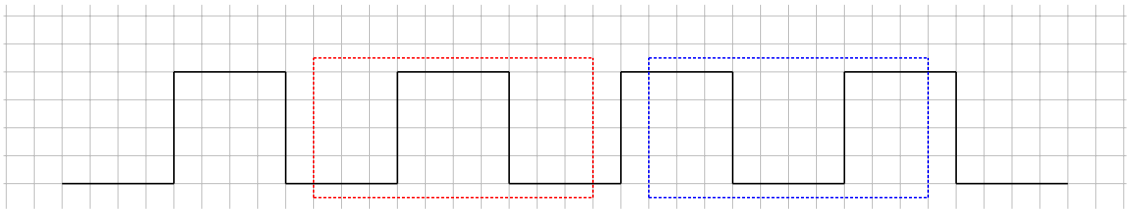
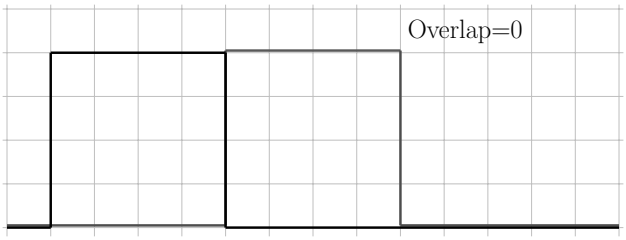
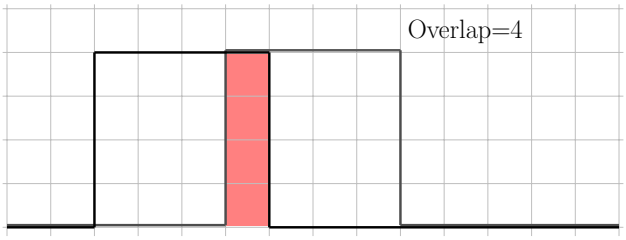


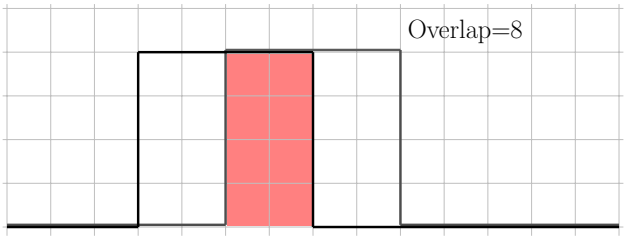
Figure 14: The chosen chunk of length 10 (the left dashed box) is incrementally shifted rightward one unit at a time. Here, we'll shift it 12 times, up to 12 units to the right, to where the right dashed box is. At each shift, the amount of overlap between the chosen chunk and the original waveform is computed as a measure of their similarity.



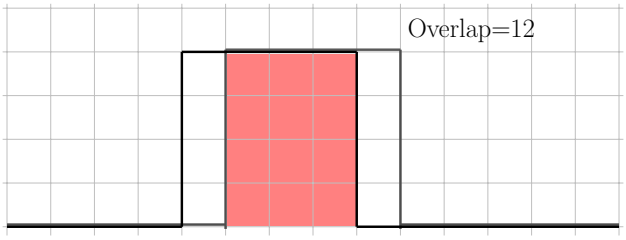
(a) Shifted 4 units rightward



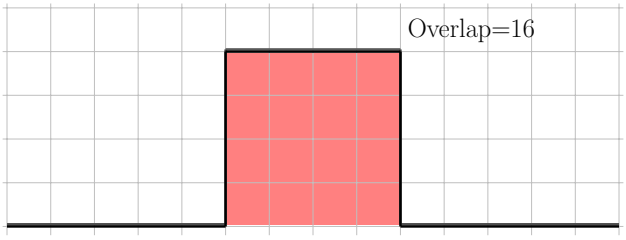
(b) Shifted 5 units rightward



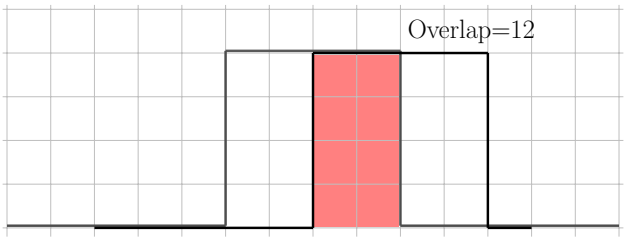
(c) Shifted 6 units rightward



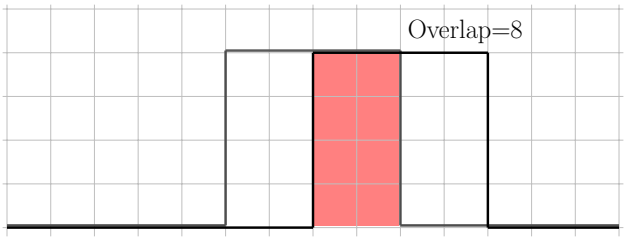
(d) Shifted 7 units rightward



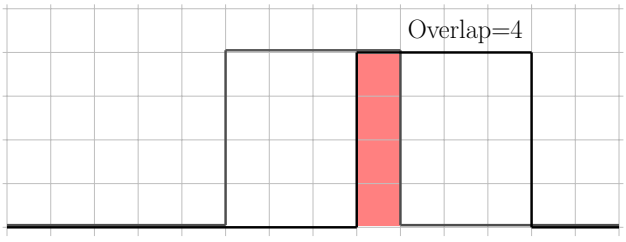
(e) Shifted 8 units rightward



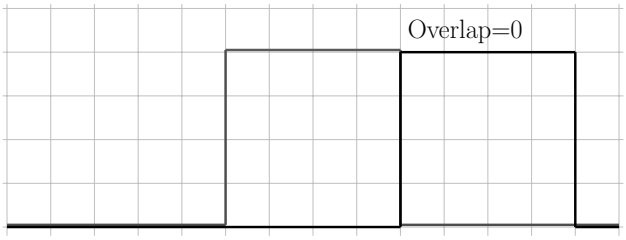
(f) Shifted 9 units rightward



(g) Shifted 10 units rightward



(h) Shifted 11 units rightward



(i) Shifted 12 units rightward

Figure 15: Shifts rightward of the waveform chunk, by increments of 1 unit, from 4 units to 12 units. Overlap occurs from shift sizes of 5 units to 11 units. The waveform chunk being shifted is drawn in black. The original waveform it's being compared to is drawn in gray.

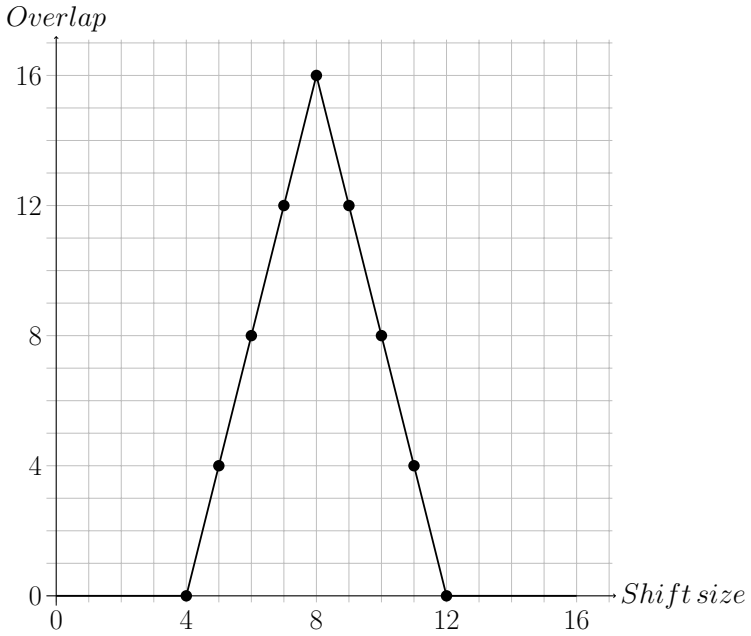


Figure 16: A graph of the number of squares in overlap between the compared waveforms for different shift sizes. The maximum overlap of 16 squares occurs at a shift 8 units rightward from the starting point for shifting. This is the peak in the plot and indicates the shift size where the compared waveforms are the most similar, and thus the short-time autocorrelation estimate of the period, 8 units, and thus a f0 of 1/8 units.

our original description of the waveform in Figure 12. These steps of (1) chunking (windowing), (2) shifting, (3) measuring overlap, and (4) finding the shift size that yields the greatest overlap, are the essentials of autocorrelation-based f0 estimation and related algorithms.

Note that if we continued to shift rightwards and measure overlap, we'd find that the triangle shape in Figure 16 continues to repeat every 8 units, as shown in Figure 17, giving us estimates of multiples of 8, i.e. $8 \times 2 = 16, 8 \times 3 = 24, 8 \times 4 = 32, \ldots$ for the period, equivalent to f0 estimates of 1/16, 1/24/, 1/32, ..., i.e 1/2, 1/3, 1/4, ... of the autocorrelation f0 estimate. These are called *sub-harmonics* of f0, derived from multiples of the period of the waveform. The *sub-* prefix indicates that these frequencies are lower than the true f0. It's typical to see autocorrelation peaks corresponding to sub-harmonics. If a waveform repeats every 8 units with perfect regularity, it'll also repeat at any multiple of 8 units with perfect regularity, so we pick the peak at the smallest non-zero shift size (in the case of the square wave, the peak at 8 units) to be the period. In real speech signals, in part since waveforms rarely repeat with perfect regularity, autocorrelation peak heights at shift sizes corresponding to sub-harmonics are typically—but not always—lower than those corresponding to true f0. We'll see in the next section, §4.3.2, that strong sub-harmonics can pose thorny issues for f0 estimation.

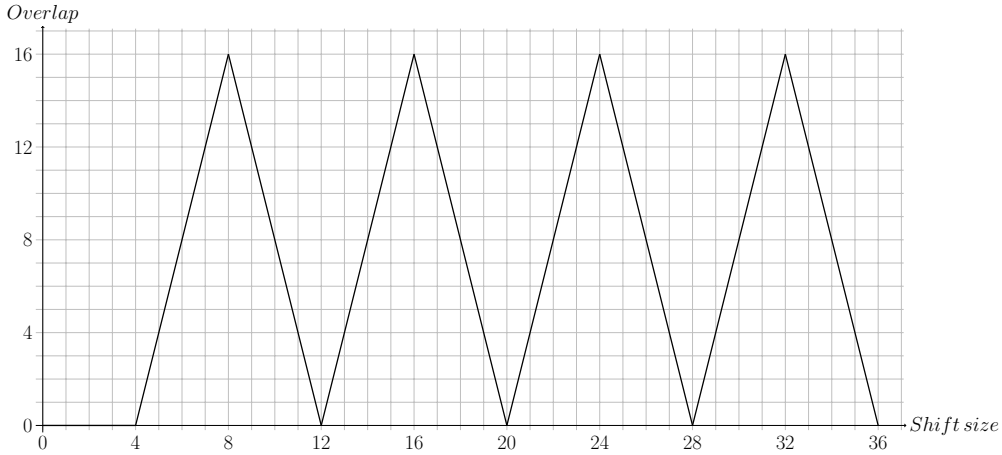


Figure 17: A graph of overlap as a function of shift size as we continue to shift rightward past a 12-unit shift. The triangle pattern repeats every 8 units, i.e. it has a period of 8 units, same as our estimated period of 8 units in Figure 16.

4.3.2 When fundamental frequency is ill-defined

Now that we've discussed what happens in computational f0 estimation, we can discuss how and when f0 contours from recordings can be misleading. There are two ways in which this can happen: (1) errors in some component of the f0 estimation algorithm, and (2) properties of the speech signal which make f0 and thus f0 estimation ill-defined.

A major source of errors comes from consonant-tone interaction, particularly voicing. Speech signals aren't simple shapes repeated with perfect regularity like the square wave we've been working with in §4.3.1. For one, the rate of vocal fold vibration is only well-defined when the vocal folds are vibrating, i.e. when the segment being uttered is voiced. Thus, f0 is not defined over voiceless intervals and a standard component of computational f0 detection in addition to a short-time autocorrelation algorithm is also a voicing detector. This detects whether an interval in the speech signal is voiced or not, and consequently, whether or not an estimate of f0 should be attempted for that interval. As a component of computational f0 estimation, voicing detection can be an occasion for errors, as reviewed by Gussenhoven (2004, p. 6). For instance, oscillations in the waveform during voiceless fricatives might be mistaken for regular vocal fold vibration, or laryngealized intervals with irregular spaced glottal pulses might be mistaken as voiceless.

The other main source of errors comes from properties of the voice source such as non-modal phonation, e.g.:

> Even if the detection of voicing and voicelessness is correct, the pitch tracker may fail to analyse the voiced signal correctly. When the voice becomes creaky, as it often does at lower pitches, the algorithm may be confused by peaks in the signal that do not correspond to the vibratory action of the vocal folds. (Gussenhoven, 2004, p. 6)

If the f0 estimator returns a sub-harmonic, say, 60 Hz, as its estimate for an f0 of 120 Hz, there may be an abrupt halving of f0 in the f0 track called *pitch halving* or *period doubling*. This could be considered an error if there is a clearly perceived pitch corresponding to an f0 of 120 Hz rather than 60 Hz. One case in which this error could occur is if the f0 estimation routine was fed in 100 Hz as the maximum f0 to be considered as a possible f0 estimate by the software user, so that 120 Hz would be out of consideration and 60 Hz chosen as the f0 estimate.

However, not all cases of pitch halving are tracking errors, and more generally, not all cases of discontinuities in the f0 contour are tracking errors, although its common to lump them all together as such. The issue is that the definition of fundamental frequency in phonetics assumes *(quasi-) perfect temporal regularity* in the occurrence of glottal pulses as the vocal folds vibrate. Large deviation from regularity implies that f0 is ill-defined. There are canonical situations when this is the case which occur regularly in tonal fieldwork, as enumerated by Pierrehumbert (1990, p. 387) and Talkin (1995, p. 500).

First, in the rapid transition of f0 in a steep tonal rise or fall, the duration from one glottal pulse cycle to the next can vary wildly, violating the assumption of regularity in the duration of cycle-to-cycle periods. While there may be a clear percept of a rapid pitch transition, e.g. the percept of a steep rise, there may be no clear percept of individual absolute pitch estimates at points during the transition.

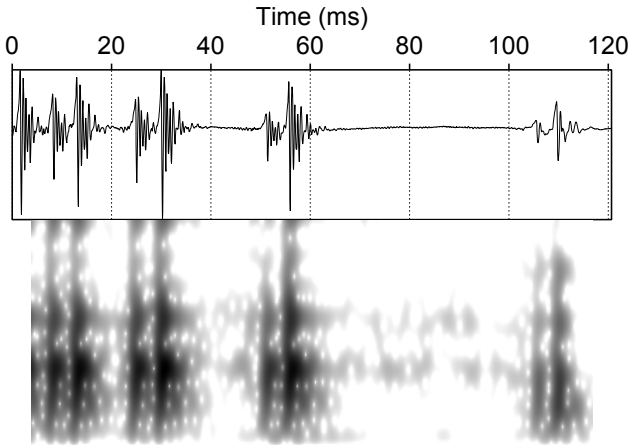


Figure 18: Waveform and spectrograph displaying vocal in Mandarin Tone 4 (5-level scale: 21(3)), the dipping or low tone. The distance between glottal pulse is irregular and as long as 50 ms. Vocal energy immediately drops to nothing after each pulse.

Second, non-modal voice quality can cause irregularity in glottal pulses, violating the assumption of quasi-periodicity in the speech signal. In such cases, f0 is not well-defined, and there may not be a clear pitch percept, either. *Vocal fry* is a well-defined laryngealized voice quality contingent on an extremely low f0 in the range from 7 to 78 Hz (Gerratt and Kreiman, 2001). The glottal pulses in vocal fry are impulse-like, similar to the sound of an explosion or gunshot—a sharp noise burst that rapidly decays into silence—and are typically widely variable in duration from pulse to pulse, as shown in the waveform and spectrogram of the Mandarin tone displayed in Figure 18. Thus, there is typically no regular repetition of the glottal pulse cycles in vocal fry and no well-defined f0 or absolute pitch percept.

In another kind of laryngealized voice quality, *period doubling* occurs in the waveform, in which pairs of glottal cycles alternate in duration and/or amplitude. Such a waveform has multiple, distinct repeating cycles rather than a single repeating cycle (which occurs in periodic waveforms) or no repeating cycle at all (which occurs in aperiodic waveforms). The percept of such a waveform can be bitonal and varies depending on the relative strength of sub-harmonics.[19] In Figure 19, we show an example of long-short alternating glottal cycles from a laryngealized low falling tone in Cantonese.

The key point for tonal fieldwork is to be aware that there are instances in which f0 estimation algorithms genuinely make errors and there are instances in which f0 itself and possibly also the pitch percept are not well-defined. In the case of genuine f0 tracking errors, the fieldworker can avoid being misled by paying attention to what his/her ears tell him/her. In the case of ill-defined f0, even what the ear tells us can defy easy representation as a transcription of the pitch contour, since there may be no clear pitch percept: irregularities in the f0 tracks in such cases reveal the limits of

[19]Sun and Xu (2002) has implemented a f0 estimation algorithm that pays attention to relative subharmonic strength in estimating f0.

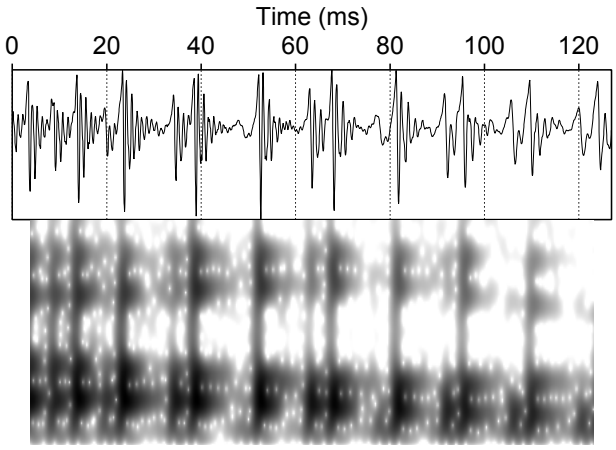


Figure 19: Waveform and spectrograph displaying period doubling in Cantonese Tone 4 (5 level scale: 21), the low falling tone. The period for glottal cycles is quite regular, but there are actually two repeating patterns, a long cycle and a short cycle, as can be seen by the pattern of closely spaced double striations followed by widely spaced striations in the spectrogram.

standard pitch-based tonal transcriptions. Such cases invite careful consideration of how to represent the dependent variable in toneme discovery. How should we describe the pitch contour over a word as a dependent variable? And is it sufficient to consider only pitch as the dependent variable? We close our explication of experimental design in elicitation with a re-examination of our linking hypotheses in toneme discovery.

4.4 Fundamental frequency and beyond

When we first began to describe elicitation in terms of experimental design in §2.2, we started with the linking hypothesis Hypothesis 2, repeated here:

Hypothesis 2 (Linking hypothesis between lexical tone classes and pitch contours) *Lexical tone classes (tonemes) induce systematic variation in the pitch contours of words. Therefore, the unobservable cognitive concept of a toneme is observable via its influence on the pitch contours of words: if two words have different pitch contours, then they belong to different lexical tonal classes.*

But how can we tell if two words have different pitch contours? What do we really mean by the pitch contour of a word? What are the relevant dimensions of tonal contrast that we might try to encode in a tonal transcription during toneme discovery? We've stressed that the exact choice of tonal transcription is not critical in initial stages of clustering tonal classes (§2.1), when the focus is on sorting elicitation items into the same tonal class or different tonal class. But we still need to be aware of what acoustic and perceptual dimensions may be relevant, even if we don't take much stock in trying to represent them.

Moreover, as we become more confident in the hypothesized tonal classes, it is typical to settle into using tonal transcriptions, so what's (implicitly or explicitly)

encoded as working hypotheses about relevant dimensions of tonal contrast in these transcriptions is important to think about. As mnemonic devices, transcriptions play a role in tuning the ear. In the course of toneme discovery, the choices of what and what not to include in a transcription can strongly bias what the fieldworker pays attention to going forward in fieldwork. If, for example, one never transcribed properties of the voice source beyond f0 like details of phonation from the outset, one is unlikely to continue to pay attention to them, although they may play a role in the grammar that is consequently missed later on.

In the rest of this section, we review acoustic and perceptual dimensions of tonal contrast that have been studied in a number of tone languages to provide a baseline for considering linking hypotheses mapping from tonemes to dimensions of contrast.

In Table 17, we catalog dimensions of tonal contrast that have been proposed in the literature, with one row per language per study. When we write *f0*, we mean absolute f0 height; when we write $f0'$, we mean f0 movement/velocity (see Figure 22); when we write $|f0'|$ we mean magnitude of f0 change, and $sgn(f0')$ indicates direction of movement. *Contouricity* is a binary parameter in Brunelle (2009) with two levels, simple (no zero-crossing in f0') and complex (zero-crossing in f0') that we notate as $|\{roots(f0')\}| > 0$. The parameter $sgn(f0')$ (direction) in (Gandour, 1983; Gandour and Harshman, 1978) showed a distribution suggestive of a binary-valued parameter $sgn(f0') > 0$ (rise or not rise). The feature $|rel(ative)f0_{fin}| > T$, where T is some threshold, is meant to represent Gandour and Harshman (1978)'s dimension "extreme endpoint", in which tones with a high or low endpoint contrast with those with a mid endpoint.

Note several different variations on f0 features: f0, average f0, relative/normalized f0, $f0_{fin}$. If just f0 is written, this means the study did not specify detail or was not designed in such a way to be more specific about how f0 height was meant. Average f0 means f0 averaged over the syllable or vowel. Relative f0 means f0 expressed in relative terms, such as "high" or "low" or f0 offset between two points. $f0_{fin}$ means f0 at offset of target syllable.

Language	Useful features	Materials/methods	Reference
Cantonese	avg. f0, $sgn(f0')$, $\|f0'\|$	Isol. monosyll., acoustics/perception/discriminant	Khouw and Ciocca (2007)
Cantonese	avg. f0, $sgn(f0')$	Synthetic isol. monosyll. perception, MDS	Gandour (1983)
Cantonese	avg. f0, $sgn(f0')$, $\|f0'\|$	Reanalysis of Fok (1974)	Gandour (1981)
Cantonese	avg. f0, $sgn(f0')$, $\|f0'\|$	Synthesized monosyllables continuum, perception	Vance (1977)
Cantonese	rel. f0, $sgn(f0')$, $\|f0'\|$	Isol. monosyll., acoustics/perception	Fok (1974)
Hmong (Green)	H1-H2, jitter, shimmer, f0 quadratic polynomial coeff.	Monosyll. in carrier phrase, perception/ANOVA	Andruski (2006)
Hmong (Green)	f0 quadratic polynomial coeff.	Monosyll. in carrier phrase, acoustics/discriminant	Andruski and Costello (2004)
Hmong (Green)	f0, H1-H2, V dur; VOT (all normalized), jitter, shimmer	Monosyll. in carrier phrase, acoustics/discriminant	Andruski and Ratliff (2000)
Mambila	f0	Synthesized sawtooth in natural carrier phrase, perception	Connell (2000)
Mandarin	avg. f0, $f0_{fin}$, f0'	Isol. monosyll., EEG/MMN, MDS	Chandrasekaran et al. (2007)
Mandarin	Syll. duration, creaky voice	Isol. monosyll., resynthesis	Liu and Samuel (2004)
Mandarin	Amp., 50 Hz $< \Delta Amp <$ 500 Hz, V duration	Isol. monosyll., Various resynthesis, acoustic/perception	Fu and Zeng (2000); Fu et al. (1998)
Mandarin	Amplitude	Signal-correlated noise, perception	Whalen and Xu (1992)
Mandarin	Syll. duration	?, Acoustic, perception	Gårding et al. (1986)
Mandarin	avg. f0, $sgn(f0')$	Synthesized isol. monosyll., perception, MDS	Gandour (1983)
Mandarin	f0, f0'	Isolated monosyll., acoustic	Howie (1976)
Taiwanese	avg. f0, $sgn(f0')$	Synthesized isol. monosyll., perception, MDS	Gandour (1983)
Thai	avg. f0, $sgn(f0')$	Synthesized isol. monosyll., perception, MDS	Gandour (1983)
Thai	avg. f0, $sgn(f0')$, $\|f0'\|$	Synthesized monosyllables, perception, MDS	Gandour (1979)
Thai	avg. f0, $sgn(f0')$, duration, $\|f0'\|$, $\|rel.\ f0_{fin}\| > T$	Synthesized isol. monosyll., perception, MDS	Gandour and Harshman (1978)
Thai	f0, $\|f0'\|$, $sgn(f0')$	Synthesized monosyllables continuum, perception	Abramson (1978)
Thai	Something not f0	Isolated monosyllables, Whispered speech, perception	Abramson (1972)
Vietnamese	f0, f0', phonation, $\|\{roots(f0')\}\| > 0$	Resynthesized isolated monosyllables, perception	Brunelle (2009)
Vietnamese	phonation (only)	isolated monosyllables, acoustic	Pham (2003)
Yoruba	avg. f0, $sgn(f0')$	Synthesized isol. monosyll., perception, MDS	Gandour (1983)
Yoruba	avg. f0, $sgn(f0')$, duration, $\|f0'\|$, $\|rel.\ f0_{fin}\| > T$	Synthesized isol. monosyll., perception, MDS	Gandour and Harshman (1978)
Yoruba	avg. f0, relative f0, $sgn(f0')$, $\|f0'\|$	Isolated disyllables, perception/MDS	Hombert (1976)

Table 17: Dimensions proposed to be useful for tonal classification in the linguistic literature. Note that in general, the parameter set proposed for a given paper is not intended to be exhaustive. For instance, in some cases, non f0-based parameters were proposed/shown to be useful but not f0-based parameters because the studies were not designed to address the usefulness of f0-based parameters.

We can draw several generalizations from Table 17. Suppose we take the set of all useful dimensions for discriminating tones proposed in some study for each language.[20] Keeping in mind that these studies were for citation form or citation form-like tones, then we can make the following generalizations:

Based on the survey, dimensions of tonal contrast that the fieldworker should be ready to listen for include dimensions falling roughly into three categories: (1) static f0 dimensions, (2) dynamic f0 dimensions, and (3) dimensions of voice quality beyond f0:

Static f0 features:

- f0
- average f0
- relative/normalized f0
- f0 at offset ($f0_{fin}$)
- extreme endpoint ($|rel.\ f0_{fin}| > T$)

Dynamic f0 features:

- f0 slope/velocity ($f0'$)
- f0 direction ($sgn(f0')$)
- f0 slope magnitude ($|f0'|$)
- contouricity ($|\{roots(f0')\}| > 0$)
- f0 quadratic polynomial coefficients (a and b from $ax^2 + bx + c$)

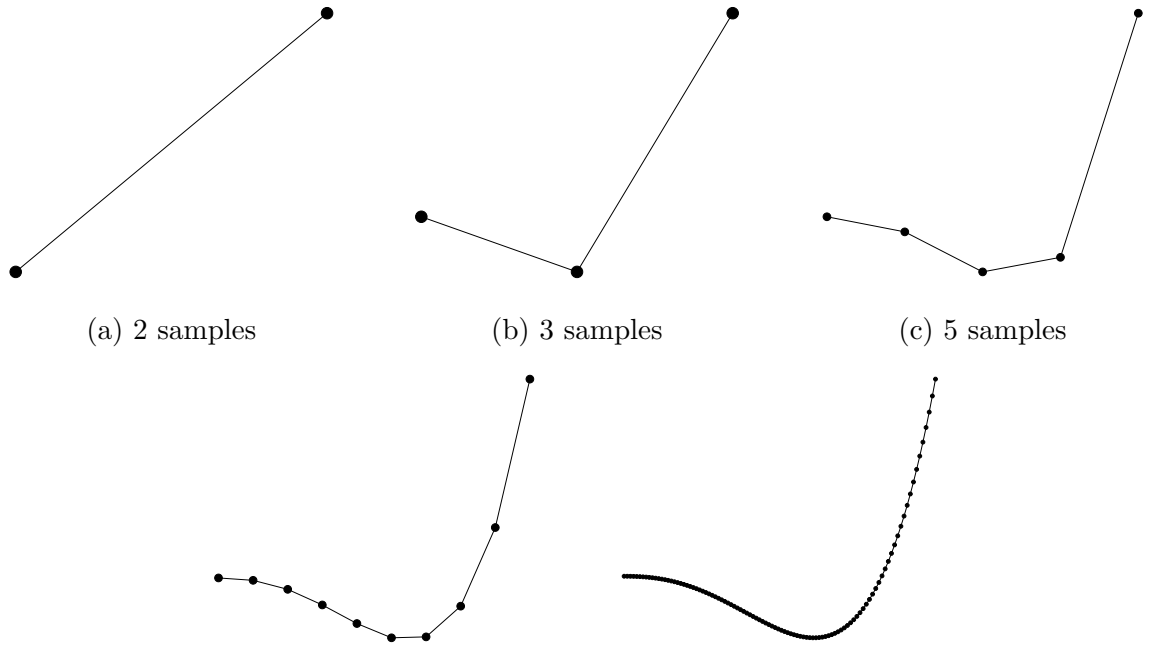
Voice quality features:

- amplitude
- amplitude fluctuations (50 Hz $< \Delta Amp <$ 500 Hz)
- jitter, shimmer (cycle-to-cycle variability in period and amplitude, respectively)
- duration
- phonation (modal/{creaky/glottalized/laryngealized})
- spectral measures: H1-H2

The presence of dimensions of voice quality beyond f0/pitch in the survey is a reminder to us to remember to pay attention to other properties of the speech signal than f0/pitch in toneme discovery. As we saw in the discussion of cases in which f0 and even pitch is ill-defined (§4.3.2), even in tone languages in which f0-based dimensions may be sufficient for tonal discrimination and other dimensions of voice quality may provide secondary cues for tonal perception, these other dimensions of voice quality can interact strongly with f0 and pitch percepts.

Setting aside dimensions of voice quality beyond f0, even within f0-based dimensions of contrast, the variety of f0-based parameters included in Table 17 merits discussion. Going into depth about f0-based dimensions is beyond the scope of this paper, but we'll give a sketch of the following issues which arise in Table 17 for linking hypotheses involving the dependent variable of f0 and pitch in toneme discovery:

[20] In general, for the survey, we take positive results most seriously. That there is, for instance, no evidence in Mambila that $f0'$ is a useful feature for tonal categorization should not be taken to mean that $f0'$ is not a useful feature, since the study was not designed to test that.



(a) 2 samples (b) 3 samples (c) 5 samples

(d) 10 samples (e) 100 samples

Figure 20: A schematic comparison of increasingly fine-grained temporal resolution of an f0 contour. It takes at least 3 samples to capture the dip in the contour and at least 5 to capture the slight bend near the beginning, but just 5 points are sufficient to capture all the points of inflection in the contour—the shape of the contour captured by 5 points, though crudely drawn, is the same as the shape of the contour captured by 100 samples.

1. different choices for the temporal resolution in the contour
2. different choices for where to measure f0/pitch in the syllable
3. the description of contour shapes in terms of mathematical functions
4. transforms of f0/pitch, calibrated with respect to some scale
5. dimensions based purely on f0/pitch movement rather than levels

These choices for linking hypotheses are relevant for both f0- and pitch-based dependent variables. For f0, the choices are about what the relevant acoustic dimensions for tonal contrast are. For pitch, the choices are about what the relevant perceptual dimensions of tonal contrast are.

Temporal resolution and the criteria for sampling points By *temporal resolution*, we mean how fine-grained the timecourse of f0/pitch variation is followed, as illustrated in Figure 20, where we increase the number of points sampled from a schematic f0 contour from 2 samples all the way to 100 samples.

Note that tonal transcriptions typically implicitly encode linking hypotheses about both the relevant degree of temporal resolution and what defines critical points of the f0 contour for tonal contrast. Chao, who introduced the iconic tone letters (Chao, 1930) used in the International Phonetic Alphabet for representing linguistic tone as

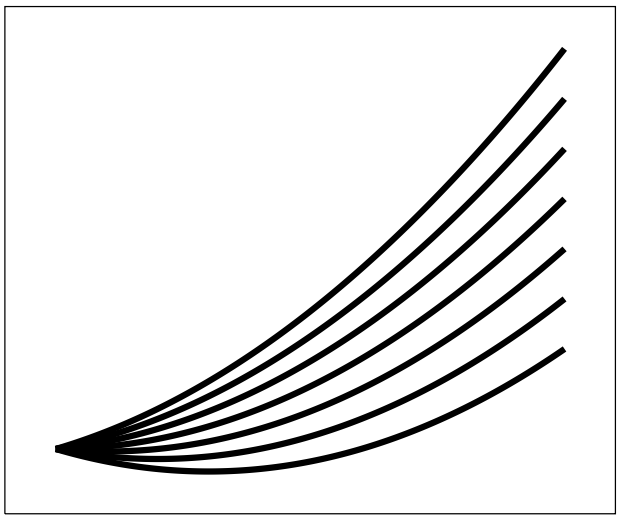


Figure 21: A family of quadratic polynomial functions of the form $x^2 - bx$.

well as the 5-point numeric scale, wrote: "the exact shape of the time-pitch curve, so far as I have observed, has never been a necessary distinctive feature, given the starting and ending points, or the turning point, if any, on the five-point scale" (Chao, 1968, 25). Tone letters are typically understood to have up to 3 samples, e.g. ˩˧, and linking hypotheses that are encoded in tonal transcriptions regarding temporal resolution in pitch contours typically assume very coarse-grained resolution. Linking hypotheses regarding what constitutes a critical point in the contour in tonal transcriptions typically follow the guidelines given by Chao above: starting and ending points and any turning points—points at which the direction of pitch movement changes.

Functions indexing shape Another line of thought that has been studied for linking hypotheses for tonal classes is parametrizing contours in terms of a restricted set of mathematical functions. This kind of linking hypothesis places the focus not on particular important points in the f0 contour, but on the overall shape of the contour. One restricted set of functions that has been proposed is polynomials of small degree, e.g. Andruski and Costello (2004); Kochanski et al. (2005). In Figure 21, we show an example of some members of a family of polynomials of degree 2, i.e., quadratic polynomials. A possible revision of our linking hypothesis would be to refine the criteria for a difference between pitch contours to be more explicit about the range of deviation permitted in considering two contours to be the same and thus within the same tonal class, which could be stated in terms of families of polynomials.

Transforms In §4.1, we introduced the role that phonetic context plays in relativizing pitch perception. Rather than relying on phonetic information from the surrounding context to relativize f0 features, another strategy which is used, sometimes in addition to including contextual information, is transformation of the distance metrics for the features. In addition to logarithmic-like f0 scaling transformations, e.g. semitones, erbs, bark-scales, a common transformation is expressing f0 variation in terms of z-scores (Rose, 1987) to make elicitation items uttered in different pitch ranges within or across speakers comparable:

$$f0_{norm} = (f0 - \overline{f0})/s \quad (1)$$

where s is the distance of one standard deviation from the mean and $\overline{f0}$ is some mean f0 value. This could be the mean f0 value for a speaker, for all speakers in a dataset, or some other reference value for calibration. In tonal transcriptions, linking hypotheses assuming similar transforms are encoded in the normalization of contours to the 5-point scale (Fon and Chiang, 1999) or to three tonal levels, H, M, L.

F0 and pitch movement The case of ill-defined absolute f0 estimates during rapid f0 transitions (§4.3.2) is also suggestive of f0 movement and pitch movement being a relevant dimension of contrast. Dynamic f0 and pitch parameters have a long history in the discussion of tonal representation as well. Uniformity in contour tone representation was a matter of debate as early as the days when autosegmental theory was in its infancy (Leben, 1973); Clark (1978) discussed such an idea from a phonological point of view; Gandour and Harshman (1978) i.a. found perceptual evidence for the relevance of pitch movement in dimensions of tonal contrast, and Gauthier et al. (2007) presented computational modeling work exploring this idea.

Extracting dynamic f0 parameters or listening for pitch movement-based percepts is also performing a kind of transform. In Figure 22, we show a series of f0 contours for rises, falls, and level tones. The contours for the rate of change of f0, the f0 velocity, are distinct between the rises, the falls, and the level tones. However, all the rises have the same rate of change and thus share the same f0 velocity contour; all the falls have the same rate of change and thus share the same f0 velocity contour, and all the level tones have no change and thus share the same constant f0 velocity of 0. Thus, extracting f0 velocity contours or listening for pitch movement collapses differences between f0 contours that are purely shifts upwards and downwards in pitch range and could make tones uttered in different pitch ranges, possibly different speakers, comparable (see Gauthier et al. (2007)).

5 Conclusion

The experimental state of mind in fieldwork elicitation is not a rigid prescription for standardized laboratory protocols, but rather, a framework for thinking that can help linguists navigate the iterative cycle of hypothesis generation and testing in discovering how a language works. We've seen in this paper that the experimental state of mind has long since been a part of methodology in tonal fieldwork, as explicated in Pike's classic toneme discovery procedure.

Here, all we've done is to acknowledge the application of experimental design principles in tonal fieldwork very explicitly. We've shown how the recognition of experimental design principles at work in elicitation brings a uniformity to elicitation methodology used for a wide array of linguistic questions in tonal fieldwork, from discovering the tonemes of a language, to studying the phonetic realization of tones, to studying tonal alternation and allophony and the syntax-prosody interface. We've also seen how explicit statement of components of experimental design, like the statement of linking hypotheses and implementation of strategies for reigning in confounding variables, can help bring precision to elicitation design and inspire further hypotheses.

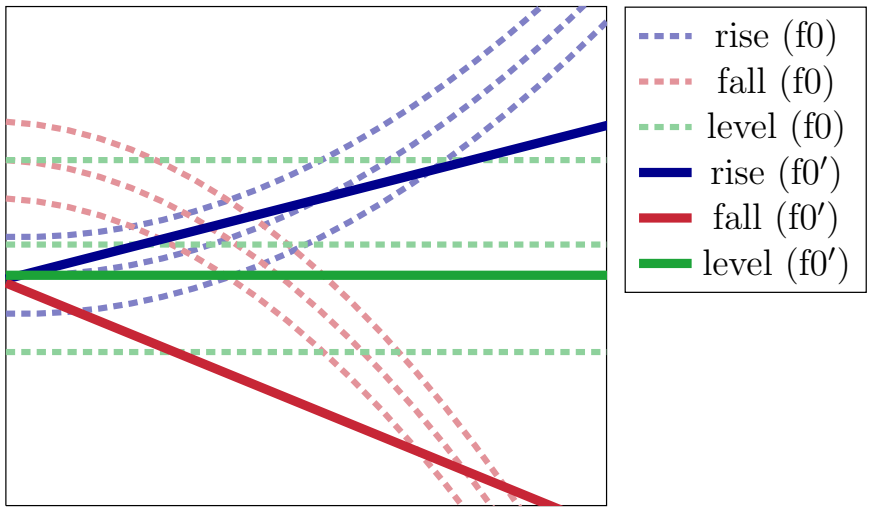


Figure 22: A comparison of f0 and f0 velocity contours for rises, falls, and level tones. Rises are shown in navy, falls in red, and level tones in green; f0 contours are dashed and f0 velocity contours are solid. All three level tones share the same f0 velocity contour; all three rises have the same f0 velocity contour, and all three falls have the same f0 velocity contour. But the f0 velocity contours are distinct between the rises, falls, and level tones.

If the principles of experimental design presented here seem very familiar, it is because the fieldworker's state of mind has always been an experimental state of mind. It's no accident that the essential qualities of the fieldworker's state of mind of the love of discovery of language-particular uniqueness and dedication to "whole language" highlighted in Hyman (1985, p. 29-30) are also essential qualities of the experimental state of mind. The somewhat paradoxical combination of meticulous attention to detail and embrace of the big picture is requisite for operationalizing a research question in terms of explanatory and confounding variables, linking hypotheses, and the like.

References

Abramson, Arthur S. 1972. Tonal experiments with whispered Thai. In *Papers on linguistics and phonetics in memory of Pierre Delattre*, ed. A. Valdman, 31–44. The Hague: Mouton.

Abramson, Arthur S. 1978. Static and dynamic acoustic cues in distinctive tones. *Language and Speech* 21:319–325.

Andruski, Jean E. 2006. Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics* 34:388–404. URL `http://www.sciencedirect.com/science/article/B6WKT-4KSD811-1/2/7f4f8c907c4266b9719184d321c12fc9`.

Andruski, Jean E., and James Costello. 2004. Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association* 34:125–140.

Andruski, Jean E., and Martha Ratliff. 2000. Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association* 30:37–61.

Baayen, R. H. 2008. *Analyzing linguistic data: a practical introduction to statistics*. Cambridge University Press.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40:177–189. URL `http://www.sciencedirect.com/science/article/pii/S0095447011000763`.

Babel, Molly, and Dasha Bulatov. 2012. The role of fundamental frequency in phonetic accommodation. *Language and Speech* 55:231–248. URL `http://las.sagepub.com/content/55/2/231`.

Baken, R. J., and Robert F. Orlikoff. 2000a. *Clinical measurement of speech and voice*. San Diego, CA: Singular Thomson Learning, 2nd edition.

Baken, R. J., and Robert F. Orlikoff. 2000b. *Clinical measurement of speech and voice*, chapter Vocal fundamental frequency, 145–223. 6. San Diego, CA: Singular Thomson Learning, 2nd edition.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences, Amsterdam*, volume 17, 97–110.

Boersma, Paul, and David Weenink. 2010. Praat: doing phonetics by computer (version 5.1.32) [computer program]. `http://www.praat.org`.

Brindley, G.S. 1960. *Physiology of the retina and the visual pathway*. London: Edward Arnold Ltd.

Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: CWK Gleerup.

Brunelle, Marc. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37:79–96.

Carpenter, Bob. 1997. *Type-logical semantics*. Cambridge, Massachusetts: MIT Press.

Chandrasekaran, Bharath, Jackson T. Gandour, and Ananthanarayan Krishnan. 2007. Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology & Neuroscience* 25:195–210.

Chao, Yuen-Ren. 1930. A system of tone-letters. *Le Maître Phonétique* 45:24–27.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, Matthew Y. 2000. *Tone sandhi*. Cambridge University Press.

Clark, Mary Morris. 1978. *A dynamic treatment of tone with special attention to the tonal system of Igbo*. Doctoral Dissertation, University of New Hampshire.

Connell, Bruce. 2000. The perception of lexical tone in mambila. *Language and Speech* 43:163–182. URL `http://www.ingentaconnect.com/content/king/ls/2000/00000043/00000002/art00002`.

Connell, Bruce. 2002. Tone languages and the universality of intrinsic f0: evidence from africa. *Journal of Phonetics* 30:101–129. URL `http://www.sciencedirect.com/science/article/B6WKT-45F4DH1-4/2/b256d873dfeb4047c8c6176fd5f2a3da`.

DiCanio, Christian T. 2009. The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association* 39:162–188.

Eady, Stephen J., William E. Cooper, Gayle V. Klouda, Pamela R. Mueller, and Dan W. Lotts. 1986. Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29:233–280.

Esposito, Christina M. 2012. An acoustic and electroglottographic study of white

hmong tone and phonation. *Journal of Phonetics* 40:466–476. URL http://www.sciencedirect.com/science/article/pii/S0095447012000174.

Esposito, Christina M., Joseph Ptacek, and Sherrie Yang. 2009. An acoustic and electroglottographic study of white hmong phonation. *The Journal of the Acoustical Society of America* 126:2223. URL http://link.aip.org/link/?JAS/126/2223/1.

Fernald, Anne, and Thomas Simon. 1984. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology* 20:104–113.

Fisher, Ronald A. 1925. *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd.

Fisher, Sir Ronald A. 1935. *The design of experiments*. New York: Hafner Publishing Company, (9th edition 1971) edition.

Fok, C.Y.Y. 1974. *A perceptual study of tones in Cantonese*. Number 18 in Occasional Papers and Monographs. Hong Kong: University of Hong Kong, Centre of Asian Studies.

Fon, Janice, and Wen-Yu Chiang. 1999. What does Chao have to say about tones. *Journal of Chinese Linguistics* 27:13–37.

Fougeron, Cécile, and Sun-Ah Jun. 1998. Rate effects on french intonation: prosodic organization and phonetic realization. *Journal of Phonetics* 26:45–69. URL http://www.sciencedirect.com/science/article/B6WKT-45J4YMN-8/2/9470a0273f434170254bcf4dc9ab8ec7.

Fu, Qian-Jie, and Fan-Gang Zeng. 2000. Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing* 5:45–57.

Fu, Qian-Jie, Fan-Gang Zeng, Robert V. Shannon, and Sigfrid D. Soli. 1998. Importance of tonal envelope cues in chinese speech recognition. *The Journal of the Acoustical Society of America* 104:505–510. URL http://link.aip.org/link/?JAS/104/505/1.

Gamut, L.T.F. 1992. *Logic, language and meaning*, volume 2. Chicago: Chicago University Press.

Gandour, Jack. 1981. Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics* 9:20–36.

Gandour, Jack. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11:149–175.

Gandour, Jack, Apiluck Turntavitikul, and Nakarin Satthamnuwong. 1999. Effects of speaking rate on Thai tones. *Phonetica* 56:123–134.

Gandour, Jackson T. 1979. Perceptual dimensions of tone: Thai. *South-east Asian Linguistic Studies* 3:277–300.

Gandour, Jackson T., and Richard A. Harshman. 1978. Crosslanguage differences in tone perception: a multidimensional scaling investigation. *Language and Speech* 21:1–33.

Gauthier, Bruno, Rushen Shi, and Yi Xu. 2007. Learning phonetic categories by tracking movements. *Cognition* 103:80–106.

Gerratt, Bruce R., and Jody Kreiman. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* 29:365–381. URL http://www.sciencedirect.com/science/article/B6WKT-457CJ0P-J/2/a40000ac94fd4ba475653abd2ec493db.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Basil Blackwell.

Goldsmith, John Anton. 1976. *Autosegmental phonology*. Doctoral Dissertation, Mas-

sachusetts Institute of Technology.

Gårding, Eva, Paul Kratochvil, and Jan-Olof Svantesson. 1986. Tone 4 and Tone 3 discrimination in modern Standard Chinese. *Language and Speech* 29:281–293.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge University Press.

Hardison, Debra M. 2004. Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology* 8:34–52.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In *Rhythm and meter*, ed. Paul Kiparsky and Gilbert Youmans, 201–260. Orlando, FL: Academic Press.

Hermes, Dik J. 1998. Auditory and visual simularity of pitch contours. *Journal of Speech, Language and Hearing Research* 41:63–72.

Hess, Wolfgang. 1983. *Pitch determination of speech signals: algorithms and devices*. Springer-Verlag.

Hombert, Jean-Marie. 1976. Perception of tones of bisyllabic nouns in Yoruba. *Studies in African Linguistics* Supplement 6.

Hombert, Jean-Marie. 1978. Consonant types, vowel quality, and tone. In *Tone: a linguistic survey*, ed. Victoria A. Fromkin, 77–111. Academic Press.

House, David. 1990. *Tonal perception in speech*. Lund, Sweden: Lund University Press.

Howie, John Marshall. 1976. *Acoustical studies of mandarin vowels and tones*. Cambridge University Press.

Hyman, Larry M. 1985. Word domains and downstep in bamileke-dschang. *Phonology Yearbook* 2:47–83. URL `http://www.jstor.org/stable/4419952`.

Hyman, Larry M. 2001. Fieldwork as a state of mind. In *Linguistic fieldwork*, ed. Paul Newman and Martha Ratliff, 15–33. Cambridge, UK: Cambridge University Press.

Hyman, Larry M. 2007. Elicitation as experimental phonology: Thlantlang lai tonology. In *Experimental approaches to phonology*, ed. Maria-Josep Solé, Patrice Speeter Beddor, and Manjari Ohala, 7–24. Oxford; New York: Oxford University Press.

Hyman, Larry M. 2011. Tone: Is it different? In *The handbook of phonological theory*, ed. John Goldsmith, Jason Riggle, and Alan C. L. Yu, 197–239. Wiley-Blackwell. URL `http://onlinelibrary.wiley.com/doi/10.1002/9781444343069.ch7/summary`.

Jaeger, T. Florian, Elisabeth Norcliffe, and eds. Alice C. Harris. 2014. Laboratory in the field.

Jun, Sun-Ah. 2005. Prosodic typology. In *Prosodic typology*, ed. Sun-Ah Jun, chapter 16. Oxford University Press.

Katz, Jonah, and Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in english. *Language* 87:771–816. URL `http://muse.jhu.edu/journals/language/v087/87.4.katz.html`.

Keating, Patricia, Christina Esposito, Marc Garellek, Sameer Khan, and Jianjing Kuang. 2011. Phonation contrasts across languages. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1046–1049. Hong Kong, China.

Khouw, Edward, and Valter Ciocca. 2007. Perceptual correlates of Cantonese tones. *Journal of Phonetics* 35:104–117. URL `http://www.sciencedirect.com/science/article/B6WKT-4MFJ1RT-2/2/ac7c29117f08b9e603ecf7a47463e86e`.

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

Krishnan, Ananthanarayan, and Jackson T. Gandour. 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language* 110:135–148. URL http://www.sciencedirect.com/science/article/B6WC0-4W2M6N5-1/2/7f9877125ad283b8ff5b5d991e1d7d7f.

Krishnan, Ananthanarayan, Yisheng Xu, Jackson Gandour, and Peter Cariani. 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research* 25:161–168.

Kuhl, Patricia K. 2004. Early language acquisition: cracking the speech code. *Nat Rev Neurosci* 5:831–843.

Kuo, Yu-ching, Yi Xu, and Moira Yip. 2007. The phonetics and phonology of apparent cases of iterative change in Standard Chinese. In *Tones and tunes: Experimental studies in word and sentence prosody*, ed. Carlos Gussenhoven and Tomas Riad, volume 2, 211–235. Berlin, Germany: Mouton de Gruyter.

Ladefoged, Peter. 2003. *Phonetic data analysis*. Blackwell Publishing.

Leben, William. 1973. *Suprasegmental phonology*. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Liberman, Mark, and Janet Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In *Language sound structure*, 157–233. The MIT Press.

Liu, Siyun, and Arthur G. Samuel. 2004. Perception of mandarin lexical tones when f0 information is neutralized. *Language and Speech* 47:109–138. URL http://las.sagepub.com/cgi/content/abstract/47/2/109.

Lively, Scott E., John S. Logan, and David B. Pisoni. 1993. Training japanese listeners to identify english /r/ and /l/. II: the role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America* 94:1242–1255. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3509365/, PMID: 8408964 PMCID: PMC3509365.

Maddieson, Ian. 1978. The frequency of tones. *UCLA Working Papers in Phonetics* 41:43–52.

Mattock, Karen, and Denis Burnham. 2006. Chinese and english infants' tone perception: Evidence for perceptual reorganization. *Infancy* 10:241.

Mattock, Karen, Monika Molnar, Linda Polka, and Denis Burnham. 2008. The developmental course of lexical tone perception in the first year of life. *Cognition* 106:1367–1381. URL http://www.sciencedirect.com/science/article/B6T24-4PG2KY6-2/2/7ab0bcca467169a4878b39309c0e1d15.

Montgomery, John C. 2005. *Design and analysis of experiments*. John Wiley & Sons, Inc., 6th edition.

Pham, Andrea Hoa. 2003. The key phonetic properties of Vietnamese tone: reassessment. In *15th ICPhS Barcelona*, 1703–1706.

Pierrehumbert, Janet B. 1990. Phonological and phonetic representation. *Journal of Phonetics* 18:375–394.

Pike, Kenneth L. 1948. *Tone languages*. University of Michigan, Ann Arbor.

Plato. 360 B.C.E. *Phaedrus*. URL http://classics.mit.edu/Plato/phaedrus.html.

Rose, Phil. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication* 6:343–352. URL http://www.sciencedirect.com/science/article/B6V1C-48V218W-BS/2/

6d65fd3baa7708bedca1444f65d43b91.

Rose, Philip John. 1988. On the non-equivalence of fundamental frequency and pitch in tonal description. In *Prosodic analysis and Asian linguistics: to honour R. K. Sprigg*, Pacific Linguisics, C-104, 55–82. Australian National University: School of Pacific and Asian studies.

Rosenbaum, Paul R. 1999. Choice as an alternative to control in observational studies. *Statistical Science* 14:259–304. URL http://projecteuclid.org/euclid.ss/1009212410.

Rost, Gwyneth C., and Bob McMurray. 2009. Speaker variability augments phonological processing in early word learning. *Developmental Science* 12:339–349.

Saville, D. J., and G. R. Wood. 1986. A method for teaching statistics using N-Dimensional geometry. *The American Statistician* 40:205–214. URL http://www.jstor.org/stable/2684537.

Shattuck-Hufnagel, Stefanie, and Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25:193–247. URL http://dx.doi.org/10.1007/BF01708572.

Stevens, S. S., J. Volkmann, and E. B. Newman. 1937. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America* 8:185–190. URL http://link.aip.org/link/?JAS/8/185/1.

Sun, Xuejing, and Yi Xu. 2002. Perceived pitch of synthesized voice with alternate cycles. *Journal of Voice* 16:443–459.

Talkin, David. 1995. A robust algorithm for pitch tracking (RAPT). In *Speech coding and synthesis*, ed. W. B. Kleijn and K. K. Paliwal, 495–518. Elsevier Science Inc.

Teller, Davida Y. 1984. Linking propositions. *Vision Research* 24:1233–1246. URL http://www.sciencedirect.com/science/article/pii/0042698984901780.

Turk, Alice, Satsuki Nakai, and Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: a practical guide. In *Methods in empirical prosody research*, ed. Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter, and Johannes Schließer, 1–28. Walter de Gruyter.

Vance, Timothy J. 1977. Tonal distinctions in Cantonese. *Phonetica* 34:93–107.

Whalen, D. H., and Yi Xu. 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49:25–47.

Wong, Patrick C. M., and Randy L. Diehl. 2003. Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language & Hearing Research* 46:413–421. URL http://jslhr.asha.org/cgi/content/abstract/46/2/413.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25:61–83.

Xu, Yisheng, Ananthanarayan Krishnan, and Jackson T. Gandour. 2006. Specificity of experience-dependent pitch representation in the brainstem. *NeuroReport* 17:1601–1605.

Yurovsky, Daniel, Shohei Hidaka, and Rachel Wu. 2012. Quantitative linking hypotheses for infant eye movements. *PLoS ONE* 7:e47419. URL http://dx.doi.org/10.1371/journal.pone.0047419.

Zhao, Yuan, and Dan Jurafsky. 2009. The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics* 37:231–247. URL http://www.sciencedirect.com/science/article/pii/S0095447009000175.

Zsiga, Elizabeth, and Rattima Nitisaroj. 2007. Tone features, tone perception, and peak

alignment in thai. *Language and Speech* 50:343–383. URL `http://las.sagepub.com/cgi/content/abstract/50/3/343`.

A Elicitation item lists for Kirikiri

item	target	utterance	gloss	lex.class	frame	len.syll
1	koo	koo	black	adj	iso	1
2	kee	kee koo	black ant	n	adj-black	1
3	foo	foo koo	black wallaby	n	adj-black	1
4	kuu	kuu	female	adj	iso	1
5	kee	kee kuu	female ant	n	adj-female	1
6	foo	foo kuu	female wallaby	n	adj-female	1
7	soo	soo	small	adj	iso	1
8	kee	kee soo	small ant	n	adj-small	1
9	foo	foo soo	small wallaby	n	adj-small	1
10	kee	kee	ant	n	iso	1
11	foo	foo	wallaby	n	iso	1
12	fijree	fijree	large	adj	iso	2
13	kee	kee fijree	large ant	n	adj-large	1
14	foo	foo fijree	large wallaby	n	adj-large	1
15	siji	siji taru	the pig sleeps	n	vp-sleep	2
16	nabij	nabij taru	the dog is sleeping	n	vp-sleep	2
17	parai	parai taru	the bandicoot is sleeping	n	vp-sleep	2
18	foo	foo taru	the wallaby is sleeping	n	vp-sleep	1
19	kaza	kaza taro	the gecko is sleeping	n	vp-sleep	2
20	siji	siji	pig	n	iso	2
21	nabij	nabij	dog	n	iso	2
22	parai	parai	bandicoot	n	iso	2
23	foo	foo	wallaby	n	iso	1
24	kaza	kaza	gecko	n	iso	2
25	siji	siji kwaa zari	the pig is making a sound	n	vp-sound	2
26	nabij	nabij kwaa zari	the dog is making a sound	n	vp-sound	2
27	parai	parai kwaa zari	the bandicoot is making a sound	n	vp-sound	2
28	foo	foo kwaa zari	the wallaby is making a sound	n	vp-sound	1
29	kaza	kaza kwaa zari	the gecko is making a sound	n	vp-sound	2
30	kruuw	kruuw kwaa zari	the frog is making a sound	n	vp-sound	1
31	kee	kee kwaa zari	the ant is making a sound	n	vp-sound	1
32	fuu	fuu kwaa zari	the honey bee is making a sound	n	vp-sound	1
33	tidu	tidu kwaa zari	the kingfisher is making a sound	n	vp-sound	2
34	kruuw	kruuw	frog	n	iso	1
35	fuu	fuu	honey bee	n	iso	1
36	tidu	tidu	kingfisher	n	iso	2

Table 18: List of elicitation items for `kiy-20111208-ap-nps-vps`.

item	kirikiri	gloss	tone1	tone2	frame	word1	word2
1	paRai giRu	bandicoot's elbow	1	1	11	paRai	giRu
2	kaza giRu	gecko's elbow	2	1	21	kaza	giRu
3	fivaa giRu	snail's elbow	3	1	31	fivaa	giRu
4	naraa giRu	wasp's elbow	4	1	41	naraa	giRu
5	tava giRu	catfish's elbow	5	1	51	tava	giRu
6	paRai ora	bandicoot's tongue	1	2	12	paRai	ora
7	kaza ora	gecko's tongue	2	2	22	kaza	ora
8	fivaa ora	snail's tongue	3	2	32	fivaa	ora
9	naraa ora	wasp's tongue	4	2	42	naraa	ora
10	tava ora	catfish's tongue	5	2	52	tava	ora
11	paRai faRo	bandicoot's groin	1	3	13	paRai	faRo
12	kaza faRo	gecko's groin	2	3	23	kaza	faRo
13	fivaa faRo	snail's groin	3	3	33	fivaa	faRo
14	naraa faRo	wasp's groin	4	3	43	naraa	faRo
15	tava faRo	catfish's groin	5	3	53	tava	faRo
16	paRai kwawaa	bandicoot's chin	1	4	14	paRai	kwawaa
17	kaza kwawaa	gecko's chin	2	4	24	kaza	kwawaa
18	fivaa kwawaa	snail's chin	3	4	34	fivaa	kwawaa
19	naraa kwawaa	wasp's chin	4	4	44	naraa	kwawaa
20	tava kwawaa	catfish's chin	5	4	54	tava	kwawaa
21	paRai koRee	bandicoot's string	1	5	15	paRai	koRee
22	kaza koRee	gecko's string	2	5	25	kaza	koRee
23	fivaa koRee	snail's string	3	5	35	fivaa	koRee
24	naraa koRee	wasp's string	4	5	45	naraa	koRee
25	tava koRee	catfish's string	5	5	55	tava	koRee

Table 19: List of elicitation items for `kiy-20111213-1-kiy-ap-framedwordlist`.