Frequency effects in Southern Min syllable contraction

James Myers and Yingshing Li

Graduate Institute of Linguistics
National Chung Cheng University
Minhsiung, Chiayi 62102
Taiwan

Accepted by *Journal of Phonetics*
February 24, 2009

Corresponding author:
James Myers
Graduate Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chia-Yi 62102
Taiwan
Email: Lngmyers *at* ccu.edu.tw

Abstract

Taiwan Southern Min syllable contraction, like many lenition processes, has long been thought to be influenced by lexical frequency. This has never been investigated quantitatively, however, because frequency estimates are not readily available. In this study, frequency estimates were derived from a small spoken corpus and from subjective judgments of both uncontracted and contracted forms. A production experiment was then run, with speakers shadowing randomly ordered disyllabic stimuli that eliminated contextual predictability. The dependent measures were acoustic correlates of intersyllabic segment reduction and tonal merger, and independent measures included lexical frequency, production latency, duration, and other phonetic and phonological factors. Regression analyses showed that both segment reduction and tonal merger correlated with lexical frequency, independently of all other factors. Further analyses suggested that despite the traditional description of syllable contraction in terms of categorical representations, the relationship between frequency and the degree of syllable contraction is gradient, with no evidence for an alternation between fully contracted and fully uncontracted forms.

Keywords: lexical frequency; lenition; syllable contraction; Southern Min; Taiwanese

1. Introduction

It has often been noted that higher-frequency lexical items show a higher probability, or greater degree, of phonetic lenition. This includes deletion or reduction of consonants (Cacoullos & Ferreira, 2000; Bybee, 2001, 2002; Patterson & Connine, 2001; Jurafsky, Bell, Gregory, & Raymond, 2001; Jurafsky, Bell, & Griand, 2002; Ranbom & Connine, 2007), weakening of vowels (Fidelholtz, 1975; Hooper, 1976; van Bergem, 1995; Berkenfield, 2001; Jurafsky et al., 2001; Jurafsky et al., 2002; van Son, Bolotova, Lennes, & Pols, 2004; Munson & Solomon, 2004; Munson, 2007), shortening of the duration of words, morphemes, or segments (Wright, 1979; Kawamoto, Kello, Higareda, & Vu, 1999; Berkenfield, 2001; Jurafsky et al., 2001; Jurafsky et al., 2002; van Son et al., 2004; Munson & Solomon, 2004; Pluymaekers, Ernestus, & Baayen, 2005; Gahl, 2008; but cf. Geffen & Luszcz, 1983; Cohn, Brugman, Crawford, & Joseph, 2005), coarticulation (Scarborough, 2004; Tseng, 2005b; Ernestus, Lahey, Verhees, & Baayen, 2006; Wong, 2006), and avoidance of stress clash (Hammond, 1999).

The apparent influence of lexical frequency on lenited productions has attracted much interest because the phenomenon raises the possibility that articulation is influenced by fine detail stored in long-term memory (Pierrehumbert, 2001, 2002, 2003; Bybee, 2001, 2002; Ernestus & Baayen, 2003; Beckman & Pierrehumbert, 2003; Au, 2005; Wedel, 2006; Yu, 2007). Nevertheless, a number of empirical questions remain open. This paper addresses three of them with underused methods and a novel data source.

First, is frequency correlated directly with lenition, or only indirectly via the mediation of more general aspects of speech production? For example, it is well known that frequency affects ease of access of lexical items in production (Balota & Chumbley, 1985; Jescheniak & Levelt, 1994; Bonin & Fayol, 2002), and some have argued that apparent frequency effects in phonetics are actually access effects that have nothing to do with articulation itself (e.g., Geffen & Luszcz, 1983; Jescheniak & Levelt, 1994). Others have pointed out that frequency is correlated with predictability; speakers need not expend as much effort on clear articulation in words that they know listeners are expecting them to say anyway (e.g., Jurafsky et al., 2001; van Son et al., 2004). Even if frequency affects articulation directly, it may do so in a minimal way, such as by adjusting a single speaking rate parameter, rather than influencing language-specific or word-specific articulatory details.

Our methodological response to these challenges is to adopt an experimental regression-based approach. Collecting speech production data experimentally makes it possible both to eliminate the influence of discourse context and to measure (and then factor out) response times as an index of ease of access. Using a regression-based design then allows us to factor out many other potential influences on lenition that would be impossible or impractical to control in a traditional factorial design.

The second question we address concerns the structure of the correlation between frequency and lenition. In particular, is a gradient increase in frequency correlated with a gradient increase in the degree of lenition? Alternative possibilities include a correlation in which frequency affects not the degree of contraction, but merely the probability that a categorically reduced alternate form will be produced. In order to address these kinds of issues, both frequency and lenition must be measured on continuous scales, and the shape of the correlation between them must be examined. Previous studies generally lack one or more of these conditions. For example, the dependent measure in Bybee (2002) is the presence or absence of word-final alveolar stops in English, so the study does not in fact demonstrate phonetic gradience. Several studies have used continuous dependent measures, but almost all of them, following traditional factorial designs, treat frequency as a nominal or, at best, an ordinal factor (e.g., Wright, 1979; Geffen & Luszcz, 1983; Kawamoto et al., 1999; Berkenfield, 2001; Munson & Solomon, 2004; Cohn et al., 2005; Munson, 2007). Only rarely are both frequency and the dependent measure treated as continuous variables in a regression-based approach (e.g., Jurafsky et al., 2001, 2002; van Son et al., 2004; Gahl, 2008), and it's even rarer to examine the shape of the correlation (e.g., Pluymaekers et al., 2005, Ernestus et al., 2006).

The third question we address concerns the phonological status of frequency-sensitive lenition. Generally speaking, the literature shows a split between two types of analyses, namely those using categorical dependent measures to examine phenomena traditionally considered "phonological" (e.g., flapping in English: Patterson & Connine, 2001; word-final alveolar deletion in English: Bybee, 2002; assibilation in Finnish: Anttila, 2006; variation between labiodentals and bilabials in Spanish: Cacoullos & Ferreira, 2000), and those using gradient dependent measures to examine phenomena traditionally considered "phonetic," with particular focus on duration (e.g., Wright, 1979; Kawamoto et al., 1999; Jurafsky et al., 2002; Munson & Solomon, 2004; van Son et al., 2004; Pluymaekers et al., 2005; Ernestus et al., 2006, looked at the degree of voice assimilation, a parameter that has also interested phonologists). It thus remains unclear to what extent frequency influences the fine phonetic detail of many patterns that have been studied almost exclusively within the tradition of theoretical phonology.

The particular frequency-influenced phonological phenomenon examined in this paper is syllable contraction in Taiwan Southern Min. In addition to helping to address the above three questions, our study also responds to another limitation in the current literature, namely its heavy emphasis on well-studied European languages. Indeed, Southern Min syllable contraction, along with the associated phenomenon of tonal merger, has no exact equivalent in familiar languages like English. The study of frequency effects in Southern Min syllable contraction also faces a methodological challenge common in research on less-studied languages, namely the lack of large corpora. We show how reliable frequency estimates can nevertheless be obtained through the subjective estimation of frequency by native speakers.

2. Syllable contraction in Taiwan Southern Min

In the Taiwan variety of the Southern Min branch of the Sinitic language family (informally known as Taiwanese), a sequence of two syllables (within or across words) may be reduced into one, a phenomenon variously known as syllable contraction, syllable fusion, or syllable merger (Cheng, 1985; Chung, 1996, 1997; Tseng, 1999; Hsiao, 1999, 2002; Hsu, 2003). Syllable contraction has also been studied in Mandarin (Hsiao, 2002; Cheng, 2004; Tseng, 2005ab; Chung, 2006) and Cantonese (Wong, 2006). The phenomenon is similar to the reduction of syllable sequences found in many languages, such as English *they'll* or *gonna*, but it is far more productive and pervasive in Sinitic languages.

Taiwan Southern Min syllable contraction has attracted much attention among Chinese phonologists (cited above) due to its productivity, systematicity, and dependence on sonority and syllable structure. A few examples are illustrated in Table 1 (note that digits represent tone levels, where 1 = lowest and 5 = highest, and underlined digits represent shortened tones). In syllable contraction, intervocalic segments are elided (in particular the first syllable coda and second syllable onset), edge segments (including the first syllable onset and second syllable coda) are preserved, the tonal contours of the two segments are merged into a single monotonic contour (rising, falling, or level), edge portions of tonal contours (initial point of the first tone and final point of the second) are generally preserved (though exceptions to this generalization are common, as in (j)-(n)), and the output respects the sonority profile of a single syllable. The preservation of vowels in the contracted syllable also depends on sonority (Hsu, 2003), with vowels of higher sonority favored over glides or vowels of lower sonority. Observations like these have been translated into generative formalism (autosegmental and/or Optimality-Theoretical) by Chung (1996, 1997), Hsiao (1999, 2002), and Hsu (2003).

[INSERT TABLE 1 ABOUT HERE]

Note that contraction can occur both within words (e.g., (d)) and between words (e.g., (n)), that contraction affects only two syllables even if part of a longer unit (e.g., (c)), and that the output of contraction may be variable, both across and within individual speakers (e.g., (e)). Hsu (2003) even provides an appendix noting disagreements between her judgments and those of Chung (1996, 1997). Cross-item variation includes whether or not contraction occurs at all; only in a small number of cases is contraction obligatory (e.g., /gua/ 'I' + /n/ 'plural' > [guan] 'we'). According to Tseng (1999), the probability of contraction in fluent speech depends, among other things, on prominence (syllable contraction tends to target unstressed and short syllables), speaking rate (more likely in fast speech), prosodic boundaries (less likely across them), and segmental features (less likely the higher the sonority of the intervocalic consonants; see also Cheng, 1985).

Syllable contraction seems to be a "late" process in production, for a number of reasons. First, it "follows" tone sandhi; the non-final tones in the first column in Table 1 are all in context form, not citation form, and the merged tone is derived from this (Hsiao, 1999). Second, as observed by Hsu (2003), the output of contraction need not obey constraints otherwise operating in the language. For example, the contracted syllable [k^hiai^{21}] in (i) is not attested in the Taiwan Southern Min syllabic inventory, and the contracted tones in (m) and (n) are not found in the tonemic inventory. Third, at least for some items, contraction can fall along a phonetic continuum, from totally uncontracted to partially contracted to fully contracted (Tseng, 1999). Some of the variability noted in Table 1 may result from variation in the degree of contraction.

Nevertheless, there is evidence of lexical influence as well. As noted above, tone contours do not always preserve tonal edges, and this often seems to happen in order to force the output tone contour to conform to the tonemic inventory (e.g., examples (k)-(m)). Some alternative forms for contracted syllables seem to serve a similar purpose (e.g., the alternative contracted syllable in (i), [k^hai^{21}], which does exist in the syllabic inventory). The existence of obligatorily contracted forms shows that the process has been fully lexicalized in a few cases. Moreover, Hsu (2003) has argued that syllable contractions tend to be avoided if they would output an identical copy of one of the input syllables, suggesting that contraction is constrained by perception and intentional processes, rather than being purely articulatory. It has also been observed (e.g., Cheng, 1985; Tseng, 1999) that contraction is more common when function words or bound grammatical morphemes are involved. Finally, and of greatest relevance to the present study, it has long been noted, at least informally, that higher frequency items seem to be more likely to be contracted than lower frequency ones (e.g., Cheng, 1985; Tseng, 1999). Since the frequency effect has been claimed to hold even in phrases, the affected units are presumably stored as lexical chunks (for other examples of phrase frequency affecting lenition see Krug, 1998; Bybee & Scheibman, 1999; Bybee, 2002).

Note that most of the items in Table 1 are of relatively high frequency. Data sets like these are filtered through the acceptability judgments of native speakers, and thus the fact that they sometimes reflect the influence of syllabic or tonal constraints is not surprising. For the same reason, such data sets do not reflect the partial contractions confirmed by instrumental studies on spontaneous speech in Taiwan Southern Min (Tseng, 1999), as well as in Mandarin (Cheng, 2004; Tseng, 2005a), though as noted above, the listing of alternative forms for some items may be a sign that these items actually vary gradiently in the degree of contraction.

3. Estimating frequency

Before we can examine the effect of frequency on syllable contraction, we must first have reasonably reliable estimates of lexical frequency. In the case of Southern Min, this turns out to be hardly a trivial issue.

The lexical frequency of an item is generally taken to reflect the amount of prior experience that a typical speaker has had with this item. The standard way of estimating this experiential frequency is by means of a corpus. Because of the highly skewed nature of word frequency distributions, a corpus can never be "large enough" if the goal is to estimate the typical speaker's experiential frequency; a smaller corpus is always worse than a larger one (Baayen, 2001). Thus for practical reasons, corpora are generally based on written samples, even when used in speech research. This approach is potentially a serious problem for Sinitic languages, however. The holdover of classical Chinese writing styles causes dramatic differences in lexical frequencies across modalities, where different function words and word-size constraints apply (e.g., morphemes that are bound in speech are often free in writing).

For the vast majority of the world's languages, no large, well-balanced corpora of the appropriate modality are available, and Southern Min is not atypical in this regard. Moreover, not only is this language generally not written, but there is no generally agreed-upon orthographic system (Ang, 1995); ensuring cross-transcriber consistency in choice of Chinese characters is particularly difficult (romanization is much harder for native-speaking users of a Chinese corpus to read).

Currently the largest available corpus is the Southern Min Spoken Corpus (Myers & Tsay, 2003), consisting of transcriptions of spontaneous conversations on radio talk shows in Southern Taiwan. As of March, 2008, it contains about 450,000 word tokens (including morphologically complex words) transcribed in a standardized orthography (etymologically correct Chinese characters supplemented by romanization for words with no known Mandarin cognate). Thus not only is this corpus not as balanced as it could be (the speakers are mostly middle-aged or older and the variety of discussion topics is relatively limited), but it is rather small. It is not even half the size of the Brown corpus (Kucera & Francis, 1967), which by contemporary standards is itself rather small. The most commonly used corpora for major languages now contain tens if not hundreds of millions of tokens (e.g., the Academia Sinica Balanced Corpus of Modern Chinese; Chen, Huang, Chang, & Hsu, 1996). The World Wide Web provides an even larger and ever-expanding corpus for many languages with written traditions (Hundt, Nesselhauf, & Biewer, 2007), but unfortunately Web-based frequency estimation does not work for Southern Min, given the lack of a consistent orthography and the difficulty of distinguishing among Sinitic languages by orthography alone (many common morphemes in Southern Min have Mandarin cognates, written with the same character, that are quite rare).

Hence we were forced to supplement corpus-based frequency estimates with subjective frequency estimates from native speakers.

3.1 Subjective frequency estimates

Subjective estimates of frequency can predict behavioral measures in language processing as well, if not better, than estimates based on a corpus (Gernsbacher, 1984; Balota, Pilotti, & Cortese, 2001). Subjective estimates are not identical to corpus-based estimates, of course, since they are sensitive to other factors as well, such as meaningfulness (Balota et al. 2001), and they are also partially dependent on modality (i.e., listening, reading, speaking, writing; Gaygen & Luce, 1998). Nevertheless, the high correlation between subjective and corpus-based frequency estimates suggests that speakers "do have access to relatively pure information about frequency of occurrence" (Balota et al. 2001, p. 645). This is a highly welcome result for all researchers interested in studying frequency effects in languages where large corpora are not available.

In the context of a phonetic study, however, the modality effects on subjective frequency estimates raise an interesting methodological problem. While judges presented with acoustic stimuli may be estimating lemma frequency, that is, how often they have heard the items in any physical realization at all (lenited or non-lenited), they may also (in some trials, or to some degree in all trials) be judging them only in the specific physical realization in which they are presented. Given the latter possibility, there is the risk of circularity in using subjective frequency estimates to predict a phonetic phenomenon, since it could be that the subjective estimates are themselves influenced by this very phenomenon taking place in the stimuli used to elicit them.

Demonstrating such circularity would take positive evidence, however, since there is no a priori reason to expect that listeners asked to judge frequency would instead be giving judgments significantly affected by just the particular phonetic phenomenon that is under investigation. The simplest test for such circularity would be to confirm that subjective estimates are indeed positively correlated with corpus-based frequency estimates. An additional method to reduce the risk of circularity would be to attempt to predict responses, not from the subjective frequency estimates themselves, but rather from a "purified" form of these estimates after statistically removing from them the influence of the phonetic measures of reduction in the eliciting stimuli (we explain this procedure below).

3.2 Methods

3.2.1 Participants

Forty university students in southern Taiwan were paid to participate in the frequency estimation experiment. All of them were fluent native speakers of Taiwan Southern Min, and as is typical of their generation, they were also fluent in Taiwan Mandarin (the Southern-Min-influenced variety of Mandarin serving as a de facto norm in Taiwan). The average age was 21.3 years old (11 males and 29 females).

3.2.2 Materials and design

Materials consisted of 120 disyllabic words or phrases selected from the Southern Min Spoken Corpus (Myers & Tsay, 2003). Items were selected by listing all lexical items in the corpus in order of token frequency, and choosing an item from within each of 120 equally sized subdivisions, resulting in token frequencies ranging from 1 to 1207. Of the selected items, 80 contained no function morphemes and 40 did. There was no significant difference in mean log frequency (corpus-based estimates) for items with vs. without function morphemes (1.51 vs. 1.34, $t(118) = -1.09$, $p > .27$). All experimental items are listed in the Appendix.

The 120 uncontracted items were recorded by a phonetically-trained female native speaker of Taiwan Southern Min who was asked to produce each disyllable naturally but distinctly, without any contraction. To produce a matching set of contracted forms, the same speaker was given modified phonetic transcriptions of the same items, where intersyllabic consonants were deleted, vocoids of the two syllables were concatenated, and tones were merged into a single contour. No attempt was made to have the resulting syllables conform to the phonotactics of Southern Min, but otherwise the contraction process followed the generalizations observed in natural contractions by Hsu (2003). These contractions were not necessarily complete, since they sometimes contained sequences of non-high (mid or low) vowels or non-monotonic tone contours, which made it clear to listeners that they had been reduced from two full syllables (diphthongs in Southern Min must contain a high vowel, and syllable-internal tone contours must be monotonic). Both sets of recordings were digitized at a sampling rate of 22050 Hz as WAV files. Transcriptions of both uncontracted and contracted forms of each item are listed in the Appendix.

Each participant judged all 120 items. Half of the items presented to each participant was in uncontracted form and half was in contracted form, matched in frequency and in the proportion of items containing function morphemes. The items were presented in counterbalanced lists to two groups of twenty participants each.

3.2.3 Procedure

To maximize sensitivity, frequency judgments were measured using magnitude estimation (e.g., Stevens, 1957). This allowed judgments to range across a continuous scale rather than being limited to a small number of discrete values.

Participants sat before computer monitors wearing headphones (Gamma semi-open digital stereo LH-945). They were first familiarized with the concept of numerical magnitude by estimating the lengths of lines presented on the monitor. At the start of the training session, a so-called modulus line was presented for participants to rate with an arbitrarily chosen number. The lengths of all subsequent lines had to be rated proportional to the modulus. Next, the concept of judging line length was transferred to judging the frequency of ten spoken stimuli proportional to a spoken modulus item. After these practice sessions, the experiment proper began. The modulus stimulus was /tsai^{53} hue^{33}/ "good-bye", which was not used in the production experiment but had a corpus frequency roughly in the middle of the frequency continuum (according to the corpus). To help participants prepare to judge a list of items containing both contracted and uncontracted forms, the modulus item was presented in partially contracted form ([tsaiue^{53}]). Participants gave the modulus an arbitrary number representing how often they had heard this item before. Then the experimental 120 items were played in random order, and each had to be given a number representing its degree of frequency proportional to the modulus. No time limit was given to make these judgments. The experiment was controlled by E-Prime (Schneider, Eschman, & Zucolotto, 2005), which recorded the participants' typed responses. On average, the whole procedure took about 25 minutes per participant.

Scores were first converted to proportions by dividing out each participant's modulus score. For example, if a participant rated the modulus as 10, and gave 20 and 5 to two other items, then for this participant the two experimental items received the scores of 2 and 0.5, respectively. To serve as norms for our production experiment, scores were averaged across participants. Like the corpus-based frequency estimates, the distributions of frequency estimates were positively skewed, with a thinner tail towards the top of the scale. Following a standard procedure, we took the logarithms of the corpus-based and subjective frequency estimates, a transformation which tends to make the values more normally distributed (by spreading out lower values and compressing higher ones) and which may more accurately reflect how frequency is actually processed (Baayen, 2001).

3.3 Results and discussion

Partial correlations were computed for corpus-based morphological surface frequency estimates (corpus frequency, for short), subjective frequency estimates for uncontracted forms (uncontracted frequency), and subjective frequency estimates for contracted forms (contracted frequency). The partial correlation between log corpus frequency and log uncontracted frequency was positive, and statistically significant ($r(120) = .46$, $p < .0001$), indicating that around 21% (r^2) of variance in the uncontracted frequency was accounted for by corpus frequency even after contracted frequency was partialed out. The partial correlation between log corpus frequency and log contracted frequency was also positive and statistically significant ($r(120) = .30$, $p < .001$), though notably weaker ($r^2 = .09$). The partial correlation between the two subjective frequency estimates did not reach statistical significance ($r(120) = -.15$, $p > .1$, $r^2 = .02$).

The positive correlations between corpus-based and subjective frequency estimates suggest a relation between these measures. The lack of a significant correlation between the two subjective estimates, however, is difficult to interpret, as with any null result. One possibility is that the contracted frequencies are a noisier reflection of objective frequencies than uncontracted frequencies, perhaps partly because they show a somewhat narrower distribution (SD 0.37, versus SD 0.44 for uncontracted frequencies). A more important finding was that the mean score for contracted frequency (-0.10) was significantly different from that for uncontracted frequency (0.46) by a paired *t* test ($t(119) = -10.62$, $p < .001$). This suggests that listeners were taking phonetic detail into account in giving their frequency estimates, since if they were judging lemma frequency only, contraction should not have mattered. This is consistent with other studies that have found speakers to be reasonably accurate in subjectively estimating the rate of variable phonetic processes (e.g., estimating the rate of schwa deletion in a given French word; Racine & Grosjean, 2002).

These results suggest that the subjective estimates do reflect experiential (lexical) frequency, though by their nature they cannot tell us whether the relevant experience involves abstract lemmas as well as acoustic forms. Given that our goal is simply to explore the phonetic structure of frequency effects in syllable contraction, these estimates should suffice. Shortly we describe an additional measure taken to improve the validity of the subjective estimates as predictors for syllable contraction.

4. A production task

Recall that the three questions addressed in this study concern whether lexical frequency correlates with lenition independently of other factors, whether or not this correlation holds relatively consistently over the whole frequency range, and whether such effects can be found even in a process, like Southern Min syllable contraction, that has hitherto attracted more attention from phonologists than phoneticians. Addressing these questions required a multiple-regression approach with continuous measures for both frequency and syllable contraction.

In order to have fullest control over our materials, as well as to eliminate the influence of contextual predictability, we collected productions in a shadowing task involving isolated words and phrases. In a shadowing task, lexical access of both perceptual and production representations are involved, since it is known that both perceptual and production tasks show facilitative effects of frequency (e.g., for spoken word perception: Norris, 1986; Warren & Marslen-Wilson, 1987; Connine, Titone, & Wang, 1993; for spoken word production: Balota & Chumbley, 1985; Jescheniak & Levelt, 1994; Bonin & Fayol, 2002). Since it has been suggested that frequency effects in lenition may be due to ease of lexical access (Geffen & Luszcz, 1983; Jescheniak & Levelt, 1994), we estimated this factor by measuring production latency (reaction time, or RT).

We also measured utterance duration, partly because it could serve as a very crude measure of speaking rate. As noted by a reviewer, a more accurate measure of speaking rate would be the duration of realized segments (i.e., utterance duration divided by the number of realized segments). However, we decided that counting realized segments was not only infeasible given the large number of tokens, but also begged one of the key issues in this study, namely whether syllable contraction is always represented categorically. Utterance duration may also correlate with the acoustic features we use to measure syllable contraction, including the distance between syllable peaks and tonal slope (longer syllables tend to have flatter tonal contours). Including it as a separate predictor may give us more confidence that our measures of syllable contraction are not merely measuring overall word shortening.

A side-benefit of our regression-based approach is that we were able to include a wide variety of other covariates (described below). If lexical frequency correlates with our measures of contraction, even when these many "nuisance" variables are partialed out in regression analyses, this would answer our first research question. Nevertheless, this study differs from many regression-based studies in that we do not determine the independent variables empirically by comparing statistical models with various factors added or dropped. Instead, we fix the set of independent variables ahead of time on the basis of a priori expectations about which factors are most likely to be relevant. In particular, we do not consider all possible interactions, since the large number of a priori factors (sixteen) makes this impossible. Moreover, we use the same set of independent variables regardless of dependent variable (i.e., those measuring overall syllable contraction or just tonal merger).

4.1 Methods

4.1.1 Participants

Twenty university students in southern Taiwan were paid to participate in the shadowing experiment. All of them were fluent native speakers of Taiwan Southern Min as well as Taiwan Mandarin. The average age was 20.5 years old (6 males and 14 females). None participated in the frequency estimation experiment.

4.1.2 Materials

Stimuli were the same 120 auditory recordings of disyllabic uncontracted forms used in the collection of subjective frequency estimates (see Appendix). As described above, 40 of these contained function morphemes and 80 did not, and they varied widely across the frequency range, whether estimated by the corpus (log range: 0 to 3.08), by subjective estimates for uncontracted forms (log range: -0.51 to 1.36), or by subjective estimates for contracted forms (log range: -0.77 to 1.30). The mean duration was 864.65 ms (range: 566 to 1061 ms).

Since we have no a priori expectation about which of the three types of frequency estimates (i.e., corpus-based frequencies, subjective frequencies of uncontracted forms, and subjective frequencies of contracted forms) will be most appropriate, we would like to include all three in our statistical models. However, as we saw in 3.3, corpus frequency is correlated with both subjective estimates, complicating the interpretation of the regression results. Moreover, the subjective estimates were elicited using phonetic stimuli that, if at all natural, must themselves be contracted to a greater or lesser degree, posing the risk of circularity.

We dealt with these two problems in the same way, namely by replacing key variables in our statistical models with their orthogonalized versions, that is, the residuals left after factoring out their expected confounds. More precisely, we first fit linear regression models predicting log subjective uncontracted frequencies and log subjective contracted frequencies from properties of the stimuli used to elicit the estimates (uncontracted and contracted forms, respectively). These two phonetic properties, trough depth and tonal distance (defined below), are the dependent measures in our production experiment, as acoustic correlates of syllable contraction. The residuals of these regression models, which represent all the variation in the frequency estimates not accounted for by these phonetic measures, were then used as predictors instead of the original estimates, since they presumably represent "purer" information about frequency than the raw measures. Next, to deal with the potential confound between log corpus frequency and the two log subjective frequency estimates (now orthogonalized as just described), we fit another linear regression model with the former predicted by the latter, and again used the residuals of this model as predictor instead of corpus frequency itself. In the remainder of this paper, then, discussion of our three frequency estimates refers to these doubly orthogonalized values (although, as it happens, the analyses gave virtually identical results with the original non-orthogonalized values).

In addition to coding each item for lexical frequency and lexical category (content vs. function), a number of phonological properties were coded as well. These properties were chosen on the expectation that they may affect our dependent measures independently of frequency. Consonantal properties were the following: presence/absence of an obstruent coda in the first syllable, presence/absence of a nasal coda in the first syllable, presence/absence of an onset plosive in the second syllable, presence/absence of an onset fricative in the second syllable, and presence/absence of an onset sonorant in the second syllable. These factors attempt to encode the sonority of the intervening consonants; according to Cheng (1985) and Tseng (1999), higher consonantal sonority should be associated with more contraction. A vocalic property was the presence/absence of a decreasing nucleus height between two syllables, which attempts to encode the tendency for reduction to be affected by the relative sonority of the two syllable peaks (Hsu, 2003). We also counted the total number of segments in the citation forms, on the assumption that phonologically longer forms would be less likely to contract because they tend to have more complex codas or onsets intervening between the two syllable peaks. Not only are these factors associated with contraction (or its absence) independently of frequency, but some are also independently associated with our acoustic measures of contraction, in particular the relative smoothness of the intensity contour across a disyllabic target (as described below). For example, not only are obstruents associated with less contraction than sonorants, as just noted, but they are also acoustically quieter, resulting in a dip in the intensity contour even if some articulatory contraction is actually present.

Tonal properties included as independent variables were the presence/absence of a difference in register (adjacent pitch height) across the phonemic tones (adjusted for tone sandhi) on each of the two syllables, and the presence/absence of a difference in slope (rising, falling, level). For example, the tones 55 and 33 would be coded as showing a register difference but not a slope difference, and the tones 53 and 33 would be coded as showing a slope difference but not a register difference. Again, the purpose was to get some handle on factors that might influence our dependent measure (here, measures of tonal merger) independently of frequency or articulatory contraction.

There were three additional independent variables that depended not on the stimuli but on the individual tokens produced by the participants themselves, namely reaction time, utterance duration, and maximum intensity. As noted at the start of section 4, reaction times in a shadowing task reflects ease of access, and utterance duration reflects speaking rate as well as other global properties of the target and articulation. Similarly, maximum intensity may correlate with the shape of the intensity contour or stress.

Only one type of interaction was included in the regression model, namely that between lexical category (content vs. function) and each of the three frequency estimates. This was done in order to test if items containing function morphemes are affected by frequency more strongly than items containing only content morphemes.

Finally, we considered the possibility that the items listed in the appendix to Hsu (2003) might behave differently from other items, given that they have been recognized as phonologically interesting; in particular, some may be lexicalized as fully contracted or as involving variation between categorical alternates (fully contracted vs. fully uncontracted). Since we intentionally tried to avoid such apparently lexicalized items, only nine of the 120 items in our experiment also appear in Hsu's appendix. The imbalance in sample sizes makes it undesirable to include the property of being listed in Hsu (2003) as a factor in the regression model, but it does seem relevant to test if frequency effects are still found when these items are removed (these items are marked in the Appendix).

4.1.3 Procedure

A shadowing (repetition naming) task was used. Participants sat before computer monitors and were presented with the stimuli over headphones (Gamma semi-open digital stereo LH-945). The experiment was run using DMDX (Forster & Forster, 2003), which also measured production latency (reaction time, or RT) from the onset of the stimulus to the initiation of speech as measured by a microphone on a stand (CAROL dynamic unidirectional MUD-326, IMP. 600Ω±30%, SENS. -48±2dB with 0dB=1V/μPa at 1 kHz). The microphone was also used to record the utterances with a digital MD recorder (Sharp MD-MT831W-GL).

Before starting the experiment, participants heard a recorded demonstration in which a male native speaker quickly repeated back ten sample stimuli (not included in the experiment). The demonstrator's speech was quick in speaking rate and short in duration (average 652.70 ms, from 554 to 727 ms); when contraction occurred, it was natural and not intentionally produced. No explicit instruction was given to contract syllables.

In each trial, an asterisk flashed on the center of the monitor for 1 s, after which the auditory stimulus was played. Participants had to respond within 2 s. As soon as participants produced an utterance, their RT was displayed on the monitor, along with a question about whether they were able to understand the stimulus. They were given unlimited time to answer this question by pressing keys labeled either "yes" or "no". After pressing a key, the next trial would begin. Stimuli were presented in random order. It took each participant about 15 minutes to complete the 120 trials.

4.1.4 Data preparation

Errors were removed from the data set prior to analyses, resulting in a loss of 1.45% data points (35 items). Errors were defined as failure to respond within the time limit, inability to understand the stimulus (as indicated by the participant's key press response), and the production of non-speech sounds (e.g., hesitation sounds) and other utterances that couldn't possibly have been meant as attempts to shadow the targets.

Response utterances were digitized at a sampling rate of 22050 Hz and analyzed using Praat (Boersma & Weenink, 2008). Waveforms and spectral displays were used to measure duration from the initiation of the first syllable to the end of the second syllable. Praat's default algorithms for tracking intensity and pitch were used to generate intensity contours and f0 contours, respectively, for each utterance. Valid intensity contours could not be plotted for 6 tokens (0.25%), so these were removed from further analysis. For the analysis of tonal merger, a further 24 tokens (1%) were removed due to problems plotting f0 contours.

4.2 Results and discussion: Intersyllabic segment reduction

Since phonologists have sometimes analyzed intersyllabic segment deletion (or reduction) and tonal merger separately, we decided to analyze separate dependent measures for these two phenomena. Given our goals, each measure was intended to capture the gradient degree of phonetic contraction rather than the mere presence or absence of lenition. We describe analyses of overall syllable contraction in this section, and analyses of tonal merger in the next. In each case, we begin with analyses addressing our first question (the independent effect of frequency) and then turn to the second (the structure of the frequency-contraction correlation).

4.2.1 Dependent and independent measures for segment reduction

We measured the degree of syllable contraction with a value we call trough depth, representing the degree of intersyllabic segment reduction. To measure this, we followed the algorithm proposed by Mermelstein (1975) for automatically finding syllable boundaries. The algorithm starts by finding the maximum intensity in a given time window (in our case, a disyllabic token), and then draws a convex hull along the intensity contour inward towards the maximum (one of the syllable peaks); the hull is identical with the intensity contour where it is monotonically increasing towards the maximum, but remains constant (a horizontal line) when intensity decreases. Trough depth is the maximum difference in intensity (measured in dB) between the convex hull and the actual intensity level. Since intensity tends to peak at syllable peaks, this algorithm measures the depth in intensity at the boundary of two syllables. In actual practice, the algorithm sometimes had to be tweaked by hand, with reference to the spectrogram, so that it did not measure an irrelevant intensity trough near the edge of the disyllable due to irrelevant noise (for example, the release burst of a disyllable-initial plosive). Figure 1 presents a sample disyllable showing the maximum point (white dot), trough depth measure (vertical bar), and the portion of the convex hull that the trough depth is measured from (horizontal bar). Hence trough depth can vary from some positive number, for uncontracted or partially contracted forms, to zero, for more fully contracted forms (though not necessarily to the degree that a phonotactically valid single syllable is produced).

[INSERT FIGURE 1 ABOUT HERE]

Independent variables were the sixteen described in 4.1.2 above, namely the phonological variables of coda obstruent, coda nasal, onset plosive, onset fricative, onset sonorant, cross-syllable sonority rise, tonal register difference of target (sandhi) tones, tonal slope difference of target (sandhi) tones (coded with 1 = present, 0 = absent), and number of target segments; the trial-specific variables of RT, duration, and maximum intensity; lexical category (content vs. function, coded with 1 = content, -1 = function in order to facilitate the interpretation of interactions); and finally the lexical frequency estimates of orthogonalized log corpus frequency, orthogonalized log uncontracted frequency, and orthogonalized log contracted frequency, along with the interactions between lexical category and each of the three frequency estimates.

Since our design is not factorial, it is important to test whether our independent variables show collinearity, which would make it difficult to interpret the results of a regression analysis. We tested this by calculating the variance inflation factor (VIF) of the independent variables (i.e., $1/(1\text{-}R^2)$ for each variable predicted by the others) for all of the observations, ignoring grouping by participants and items (tolerance, another common measure of collinearity, is simply the inverse of VIF). The maximum VIF tolerance was 3.48 for onset plosive, below the conventional threshold of 5 (for all variables, mean VIF was 1.55, with SD 0.59). In particular, the VIF for corpus-based frequencies, subjective uncontracted frequencies, and subjective contracted frequencies were quite low (respectively 1.18, 1.50, and 1.13), reducing the risk that an apparent effect of frequency could actually be due to some other variable.

Incidentally, the linear regressions used to compute VIF, as well as analyses taking by-participant and by-item grouping into account (as described below) showed the familiar negative effects of frequency (as estimated by the three estimates) on RT (higher frequency meant faster responses) and duration (higher frequency meant shorter utterances). Other correlations with frequency in the materials included higher frequency in tokens with nasal codas and lower frequency in tokens with onset sonorants. None of these correlations raised a serious risk of collinearity (in particular, VIF was 1.19 for RT and 1.17 for duration), and there was no reason to consider these correlated factors as potentially measuring the same underlying entities (unlike the case with the frequency measures); hence we did not orthogonalize them.

4.2.2 Testing the independence of frequency effects on segment reduction

In order to maximize the power of our analysis, we entered all data points into a linear mixed-effects model (LME; Pinheiro & Bates, 2000; Baayen, 2008). Like repeated-measures regression (Lorch & Myers, 1990), LME treats participants as random, but it encodes participants as a single random variable rather than as a set of dummy variables, thereby improving sensitivity. LME also makes it possible to have more than one random variable at a time (e.g., both participants and items), and since a model containing one random variable (e.g., participants) is embedded within a model containing two (e.g., participants and items), the two models can be compared with a likelihood ratio test to determine whether the second random variable helps improve the fit to the data. Currently the only tool for LME analysis with two random variables is the lmer function in the lme4 package (Bates, Maechler, & Dai, 2008) in R (R Development Core Team, 2008).

The results of our LME analysis are shown in Table 2. The lmer function currently does not report *p* values due to controversies over how the degrees of freedom (*df*) should be computed from the number of observations and parameters, so only *t* values are shown. However, with very large data sets like ours ($n > 2300$), the *df* controversies become irrelevant, and the *t* values can be treated as *z* values (i.e., the number of standard deviations from 0 in a standard normal distribution). Hence two-tailed significance at the alpha level of .05 is indicated by absolute *t* values greater than 1.96. Since a likelihood ratio test showed that the by-participants-and items model provided a significantly better fit with the data than the by-participants-only model, only results from the former model are shown.

[INSERT TABLE 2 ABOUT HERE]

The analysis showed four significant effects (which remained significant even after removing all nonsignificant factors), namely a positive effect of duration on trough depth (shorter utterances were associated with shallower troughs), positive effects of onset plosive and onset fricative (both associated with deeper troughs), and most importantly, a negative effect of orthogonalized log uncontracted frequency. The negative sign indicates that higher frequencies were associated with shallower intensity troughs, hence a greater degree of syllable contraction.

The effect of log contracted frequency trended in the same direction, though it didn't reach statistical significance. The effect of log corpus frequency was not only not significant, but didn't even trend the right direction. Assuming that our attempts to assure the relative independence of the three frequency estimates were reasonably successful, these results suggest that subjective uncontracted frequencies provide the estimates most relevant to syllable contraction. The fact that these subjective estimates performed better than the corpus frequencies presumably follows from the fact that the corpus was too small for the purposes we put it to. Similarly, the most conservative interpretation of the nonsignificance of the contracted frequency effect would simply be that like corpus frequency, this measure is not sufficiently reliable for our purposes.

Higher lexical frequencies were associated with a greater degree of contraction (i.e., shallower intensity troughs), as measured through subjective frequency estimates of uncontracted forms. Since these effects appeared even when all of the other variables were factored out, this result implies that frequency is directly correlated with the degree of contraction, rather than working via the other variables.

Another noteworthy observation in the results is the lack of a significant interaction between uncontracted frequency (apparently the currently best available estimate of frequency in Southern Min) and lexical category. These results thus fail to show a difference in frequency effects between items with function morphemes vs. those without. However, there might still be value in pursuing this issue in further research, since in the by-participants-only analysis the interaction does reach statistical significance ($\beta = 3.085$, $t = 6.264$); the positive sign suggests that frequency effects tend to be stronger for items with function morphemes.

Finally, recall that nine of our items were examined by Hsu (2003) in terms of a phonological analysis. If these items have been reanalyzed by speakers as phonologically rather than merely articulatorily contracted (in some sense), it is possible that their inclusion in our analysis skews our results. To check this, we reran the above LME by-participants-and-items analysis with these nine items removed. The significant effect on trough depth of uncontracted frequency was unaffected ($\beta = -3.981$, $t = -2.180$).

4.2.3 Testing the gradience of frequency effects on segment reduction

As a linear model, LME is designed to find straight lines that best describe the relationship between dependent and independent variables. Thus a significant effect in such a model does not show that the effect of frequency on trough depth is truly gradient, where a gradual increase in frequency is associated with a gradual increase in syllable contraction. Instead, it may be that frequency merely influences the probability of an item appearing in one of two categorical alternates: fully uncontracted or fully contracted. If the lenited alternate is selected only above a fixed threshold frequency, the frequency-lenition relationship should show significant nonlinearity (a step-like function). In addition, regardless of the shape of this relationship, variable selection of categorical alternates would also predict a bimodal distribution in trough depth itself, representing the fully contracted (low trough depth) and fully uncontracted (high trough depth) forms.

The small number of materials (120 items) and less than ideal frequency estimates preclude any decisive conclusions concerning the structure of the frequency-contraction relationship. Nevertheless, we explored the issue using two sources of evidence: the shape of the trough depth distribution, and the shape of frequency-contraction trend lines.

Figure 2 shows the density plots (smoothed histograms) for trough depth for each of the 20 participants (labeled s01 through s20). As can be seen, they do not appear to be normally distributed and indeed, several of them seem to be bimodal. This impression receives statistical support from the Engelman-Hartigan test (Engelman & Hartigan, 1969; Thode, 2002), which finds the partition of the distribution maximizing the ratio of between-partition sum of squares to within-partition sum of squares (a technique common in cluster analysis), and then gives the probability p that this maximum ratio is associated with a single normal distribution (the null hypothesis) as opposed to a mixture of two normal distributions with different means but the same variance. This test was significant ($p < .05$) for 18 of the 20 participants, confirming the impression of bimodality.

[INSERT FIGURE 2 ABOUT HERE]

The bimodality suggests a categorical distinction, but given the involvement of so many phonetic variables, there is no reason to suspect distinct frequency-dependent variants rather than distinct phonemic targets. In fact, the best predictor of whether an utterance fell in the lower or upper partition of the distribution, as defined by the Engelman-Hartigan test, turned out to be the presence/absence of an onset sonorant: if it was present, trough depth always fell into the lower partition (shallow trough), without exception. Moreover, the difference in mean trough depth for items with onset sonorants (3.38 dB) vs. those without (16.18 dB) was 12.80 dB, which, based on Figure 2, appears to be roughly the average distance between the two peaks in the participants' bimodal distributions. This general pattern was mirrored by two other phonetic factors, albeit to a lesser degree: A greater proportion of tokens in the higher partition had onset plosives (57% = 638/1128) than in the lower partition (29% = 356/1231), and the same asymmetry was found for onset fricatives (41% = 466/1128 vs. 30% = 370/1231).

These patterns were confirmed by mixed-effects logistic regression (e.g., Agresti, Booth, Hobert, & Caffo, 2000; Baayen, 2008). The by-participant Engelman-Hartigan partitions were used as categorical dependent variable (1 = upper partition, 0 = lower partition, as defined for each participant) and all of the independent variables in the LME analysis were used except for onset sonorant, which, as a perfect predictor, violates assumptions of the logistic model. In the by-participants-and-items analysis, uncontracted frequency predicted whether trough depth fell into the lower or upper partition ($\beta = -0.945$, $z = -5.686$, $p < .0001$), but just as in the LME analysis, it was not the only significant predictor; larger regression coefficients were associated with the phonological factors of onset plosive ($\beta = 3.025$, $z = 12.235$, $p < .0001$) and onset fricative ($\beta = 2.790$, $z = 11.471$, $p < .0001$). Put together, these observations suggest that the bimodality is merely a consequence of the acoustic measure of trough depth being strongly affected by the inherent acoustic properties of certain segment types, and does not indicate distinct frequency-dependent alternate forms.

We turn now to the second source of evidence, namely the shape of the trend line for the frequency-contraction relationship. One simple, unbiased way to plot such a line is to use a form of local regression called loess modeling (Cleveland & Devlin, 1988) in a scatterplot with trough depth as dependent and log frequency as the only independent variable. Loess lines track the local trends in a scatterplot by fitting the data to polynomial functions in a moving span of fixed size. Hakuta, Bialystok, & Wiley (2003) note that it is customary to test spans varying from 0.25 (less smoothing) to 0.75 (more smoothing) of the total range. Two such loess lines are shown in Figure 3, which relate trough depth with log uncontracted frequency. The figure combines the data from all 20 participants since separate analyses by participants would leave only 120 points per analysis, too sparse to conduct a meaningful loess analysis. Production tokens of the nine items listed in Hsu (2003) are highlighted as black triangles.

[INSERT FIGURE 3 ABOUT HERE]

The general downward slope of the loess lines is clear, and in the more highly smoothed curve (span = 0.75), the line is quite flat. Although it seems to be divided into two roughly horizontal segments with a drop between them, around where the log subjective estimates reach 0.4, this drop is unlikely to represent a threshold between uncontracted forms on the left (with deeper trough depths) and contracted forms on the right. The scatterplot as a whole doesn't support such an analysis, given the very small difference in trough depth across this putative threshold, a mere 3.89 dB. This is less than 9% of the total range of trough depths; clearly there is a lot more variation than would be expected if the two halves of the loess line truly reflected two distinct types of forms.

Notice that the items listed in Hsu (2003), marked by triangles in Figure 3, are not clustered at the upper end of the subjective frequency range. Interestingly, however, the most frequent of these items, /to^{55}ui^{33}/ "where" (item 97 in the Appendix) is not only at the very top of the uncontracted frequency range, but it also appears to be produced categorically (trough depths of all tokens close to zero). The least frequent of the items listed by Hsu (2003), according to these estimates, namely /hɔ^{21}laŋ^{33}/ "by someone" (item 94 in the Appendix), also seems to be produced categorically. Nevertheless the seven items from Hsu falling along the frequency range between these two show a great deal of cross-participant variation; the vertical distribution of triangles in Figure 3 makes it clear that the productions don't cluster into two distinct alternate forms, but rather vary gradiently.

4.3 Results and discussion: Tonal merger

4.3.1 Dependent and independent measures for tonal merger

Syllable contraction in Southern Min is accompanied by merger of the two syllables' tonal contours, a phenomenon that has received much attention in its own right (see review in 2). It is a nontrivial matter, however, to find a single continuously varying index of the merger of dynamic tonal contours. Our solution to this challenge is a measure we call tonal distance, based on the insight that as tone contours merge, they join together into a single monotonic curve, since naturally produced fully contracted syllables in Southern Min never contain complex tone contours (e.g., rising then falling). To quantify this insight, we defined tonal distance in terms of the height and slope of the tones of the first and second half of each utterance token according to the following algorithm. First we divided the f0 contour at the point defined as the syllable boundary by the Mermelstein (1975) algorithm, or if trough depth was zero, by the intensity maximum. We then computed a linear regression equation for each of the two f0 sub-contours (dependent: f0 in Hz, independent: time in ms). This gave us two regression parameters for each sub-contour, with the intercept a representing the overall height (in Hz) and the coefficient b representing the slope (in Hz/ms). Thus the two sub-contours could be coded as vectors (a_1, b_1) and (a_2, b_2), respectively. Tonal distance was then defined as the Euclidean distance between these two vectors, as in (1). As with trough depth, tonal distance is inversely related to lenition (here, tonal merger).

(1) Tonal distance $=\sqrt{(a_1-a_2)^2+(b_1-b_2)^2}$

Although tonal distance is expressed in the awkward units of $\sqrt{[\mathrm{Hz}^2 + (\mathrm{Hz/ms})^2]}$, this measure nevertheless seems superior to alternatives based on intercept difference or slope difference alone, since distinct tone contours may accidentally share the same intercept at the start of the disyllable (e.g., [53-55]) or the same slope (e.g., [55-33]).

Independent measures were the same sixteen (plus three interactions) used in the analysis of trough depth.

4.3.2 Testing the independence of frequency effects on tonal merger

Tonal distance proved to be positively skewed, perhaps because tonal coarticulation made adjacent tones a bit more similar regardless of the degree of syllable contraction. Hence before running the LME analyses, we applied a logarithmic transformation as we had done for the frequency measures.

The by-participants-and-items results of a LME analysis for log tonal distance are shown in Table 3. For the reasons explained earlier, parameters with $|t| > 1.96$ are considered statistically significant. Four significant effects were observed, namely of duration, maximum intensity, coda obstruents, and orthogonalized log uncontracted frequency (all but maximum intensity remained significant when nonsignificant factors were removed from the analysis).

[INSERT TABLE 3 ABOUT HERE]

Tonal distance was significantly greater with a coda obstruent, indicating less tonal merger. This makes sense, given that obstruents tend to block syllable contraction in general.

There was also a significant negative effect of duration, indicating smaller values (i.e., greater tonal merger) in longer utterances. This seems to be an artifact of our measures: longer syllables necessarily have flatter f0 contours than shorter syllables, even if both syllable types have the same phonological tone contour (e.g., both [53]), and this means they will be associated with more similar regression equations. There was also a trend towards a negative effect of maximum intensity, with louder items showing a greater degree of tonal merger. This result is somewhat counterintuitive, given that louder items are presumably more stressed, and thus presumably less likely to reduce, but it was also statistically the weakest effect.

Crucially, as with trough depth, tonal distance was also significantly negatively correlated with log uncontracted frequency. Moreover, it was again not significantly correlated with the other two frequency measures. These results confirm both that lexical frequency affects tonal merger, a phenomenon associated with syllable contraction, and that subjective frequency estimates are a more reliable measure than frequencies based on the small corpus. This time when we removed the nine items listed by Hsu (2003), however, even uncontracted frequency was no longer significant in the by-participants-and-items analysis ($\beta = -47.041$, $t = -1.718$), though it remained so in the by-participants-only analysis ($\beta = -46.932$, $t = -2.461$). Without overinterpretating this null result, we can merely speculate that tonal distance may be a somewhat weaker correlate of syllable contraction than trough depth, making frequency effects in it harder to detect if fewer items are tested.

4.3.3 Testing the gradience of frequency effects on tonal merger

The same two analyses used to examine the relationship between frequency and trough depth were then applied to tonal distance. After the log transform, tonal distance showed roughly normal distributions by participants, as demonstrated in Figure 4. Thus there was no need for further analyses of distribution shape; the variation in tonal merger could not involve an alternation between two categorical forms.

[INSERT FIGURE 4 ABOUT HERE]

Similarly, loess curves showing the relationship between log tonal distance and uncontracted frequency (the only frequency measure that showed a significant effect) showed no sign of a sharp drop. This is shown in Figure 5, with the scale of the y-axis adjusted in order to highlight the slope. The nine items listed in Hsu (2003), marked as small triangles, seem to be distributed randomly across the uncontracted frequency range.

[INSERT FIGURE 5 ABOUT HERE]

There is thus no reason to reject the hypothesis that the degree of tonal merger, as measured by tonal distance, is gradiently correlated with uncontracted frequency (our best measure of lexical frequency), without any hint of categorical alternate forms.

5. General discussion

This study has provided answers, of various degrees of decisiveness, to all three of our research questions. First, the degree of syllable contraction in Southern Min, as measured by both of our acoustic measures (trough depth as a measure of intersyllabic segment reduction, and tonal distance as a measure of tonal merger), does indeed correlate with lexical frequency, independently of reaction time, duration, and various other phonetic and phonological factors (including intensity, manner of consonant articulation, vowel height features, number of segments, tone features, and degree of contraction in the shadowing stimuli themselves). Moreover, since the utterances were produced in isolation and in random order, there was no influence of contextual predictability. Lenition, at least in this case, seems to be correlated with lexical frequency directly, not via some other factor.

Frequency effects have played an important role in debates over the motivating force behind fine-grained adjustments in articulation. A speaker-oriented explanation of frequency-sensitive lenition would focus on notions like automatization, whereby articulations become more energy-efficient with practice. Models of this sort are advocated by Bybee (2001), Pierrehumbert (2001, 2002, 2003) and Oudeyer (2005, 2006) and Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands (2000) provide experimental evidence that speakers do not take listener needs into account when articulating. A listener-oriented explanation would agree that speakers want to save energy, but would explain frequency effects in terms of how predictable the target items are to the listener; the more predictable, the more lenition the speaker can get away with. Production models that take the listener into account are discussed by Lindblom (1990), Boersma (1998), Steriade (1999), Jurafsky et al. (2001), Flemming (2002), and Aylett & Turk (2004).

Since frequency is itself correlated with predictability, frequency effects are in principle consistent with both speaker-oriented and listener-oriented approaches. Thus our finding that frequency has an independent effect on syllable contraction is not enough to argue for one approach over the other. Nevertheless, our findings do provide evidence against certain versions of both speaker-oriented and listener-oriented approaches. In particular, the fact that frequency effects on lenition remained after reaction times were factored out in regression potentially challenges models that ascribe such effects solely to ease of lexical access by the speaker (Geffen & Luszcz, 1983; Jescheniak & Levelt, 1994). We also reconfirmed the findings of earlier studies showing that frequency effects are distinct from the effects of contextual predictability (Jurafsky et al., 2001).

Regarding our second research question, we found no compelling evidence that the relationship between frequency and the degree of syllable contraction is anything other than gradient; namely, no evidence was found for a frequency-sensitive alternation between categorically distinct uncontracted vs. contracted forms. This was implied both by the lack of a sharp drop in the nonlinear regression lines in scatterplots relating frequency and acoustic measures of contraction, and by the apparently continuous variation in the dependent variable reflected in these scatterplots. However, further research is necessary to establish the point more decisively. Frequency-sensitive variation between apparently categorical allomorphs does seem to occur in natural language, as in the choice between competing prosodic structures in Finnish (Anttila, 2006) or English (Hammond, 1999, 2004). Based on our current data, however, frequency effects in Southern Min syllable contraction seem to be more like the gradient frequency effects that have been observed in duration (e.g., segment duration in Dutch: Pluymaekers et al., 2005).

Third, these gradient frequency effects were found in a pattern that had formerly attracted more interest from phonologists than phoneticians, and had been described primarily in terms of categorical representations. Our phonetic results are consistent with earlier observations that syllable contraction is affected by the manner, especially the sonority, of the intervening consonants, and that it is accompanied by merger of tone into a single tonal contour. We also showed, however, that syllable contraction may be realized gradiently. This holds even for some of the examples previously described by Hsu (2003) as involving optional variation between two competing alternate forms.

Nevertheless, Southern Min syllable contraction is not thereby reduced to a mere articulatory process. This cannot be the case, since as already noted in section 2, syllable contraction often (though not always) obeys phonotactic constraints that hold of the Southern Min syllabary more generally. Moreover, our study has demonstrated that some of the items we tested from Hsu (2003) do not vary at all, gradiently or otherwise, but are produced in a consistently contracted form. These items truly seem to have been reanalyzed as categorical, similar to the way English speakers are generally not free to contract *going to* partially to **g[o:]nna*. In any case, we do not attempt to provide an articulatory explanation for the characteristics of syllable contraction, and we leave open the possibility that some of these characteristics depend on abstract phonological representations or principles.

Beyond addressing our three major empirical questions, this study has also argued for the benefits of certain underused methodological techniques. The most important of these is the use of a regression-based experimental design, which not only permits detailed study of relationships between continuous variables, but also gives a handle on the eternal bane of lexical research, namely the proliferation of potentially relevant independent variables.

Another methodological novelty is even more relevant to future research on frequency effects in understudied languages, namely the reliance on subjective frequency estimates over corpus-based frequency estimates. Very few languages have sufficiently large and well-balanced corpora, a limitation that is even more severe for languages without a standardized orthography. It is natural to have misgivings about substituting "subjective" estimates for the "objective" estimates of a corpus, until one realizes that any corpus is inevitably a miniscule sample of the totality of speech processed by the typical experimental participant over a lifetime. Moreover, as Leech (2007) notes, and a moment's thought makes clear, the "representativeness" of a corpus is impossible to measure objectively (unless it is embedded in a still larger corpus). Previous research has shown that subjective frequency estimates work as well, if not better, as corpus-based frequencies at predicting certain behavioral measures (Gernsbacher, 1984; Balota et al., 2001), and our own study confirms that the same holds for aspects of articulation.

Our final methodological point is that as research on lenition probes phonetic detail more deeply, sample sizes will have to increase considerably. Proper sample size is inversely related to grain size; grain size in a sample of objects that are always either 1 or 2 cm long will be detectable in a much smaller sample than if the lengths range from 1 to 2 cm in 1 mm increments. Since we do not yet know what the grain size is for the representations used in the production of lenition, the safest bet is to design experiments with several times the number of lexical items used in this study.

References

Agresti, A., Booth, J. G., Hobert, J. P., & Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, *30*, 27-80.

Ang, U. (1995). On problems with the orthography in Southern Min teaching materials [Lun Taiwan Minnanyu jiaocai de shuxie wenti]. In F.-F. Tsao & M.-H. Tsai (Eds.), *Papers from the 1994 Conference on Language Teaching and Linguistics, Vol. I: Southern Min* (pp. 319-336). Taipei: Crane Publishing.

Anttila, A. (2006). Variation and opacity. *Natural Language & Linguistic Theory, 24*, 893-944.

Au, C.-P. (2005). Acquisition and evolution of phonological systems. PhD dissertation, City University of Hong Kong.

Aylett, M., & Turk, A. (2004) The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech, 47*(1), 31-56.

Baayen, R. H. (2001). *Word frequency distributions*. Dordrecht: Kluwer Academic Publishers.

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

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: lexical access and/or production? *Journal of Memory and Language*, *24*, 89-106.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639-647.

Bard, E. G., Anderson, A. H., Sotillo, G., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1-22.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes. R package.

Beckman, M. E., & Pierrehumbert, J. (2003). Interpreting ‘phonetic interpretation’ over the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology VI* (pp. 13-37). Cambridge: Cambridge University Press.

Berkenfield, C. (2001). The role of frequency in the realization of English *that*. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 281-307). Amsterdam: John Benjamins.

Boersma, P. (1998). *Functional phonology: Formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer. Software retrieved from World Wide Web: http://www.fon.hum.uva.nl/praat.

Bonin, P., & Fayol, M. (2002). Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology*, *14*(3), 289-313.

Bybee, J. L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*, 261-290.

Bybee, J. L., & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics, 37*, 575-596.

Cacoullos, R. T., & Ferreira, F. (2000). Lexical frequency and voiced labiodental-bilabial variation in New Mexican Spanish. *Southeast Journal of Linguistics*, *19*(2), 1-17.

Chen, K.-J., Huang, C.-R., Chang, L.-P., & Hsu, H.-L. (1996). SINICA CORPUS: design methodology for balanced corpora. *Language, Information and Computation, 11*, 167-176.

Cheng, C.-E. (2004). An acoustic phonetic analysis of tone contraction in Taiwan Mandarin. MA thesis. National Cheng Chi University.

Cheng, R. L. (1985). Sub-syllabic morphemes in Taiwanese. *Journal of Chinese Linguistics*, *13*(1), 12-43.

Chung, K. S. (2006). Contraction and backgrounding in Taiwan Mandarin. *Concentric: Studies in Linguistics*, *32*(1), 69-88.

Chung, R.-F. (1996). *The segmental phonology of Southern Min in Taiwan*. Taipei: The Crane Publishing.

Chung, R.-F. (1997). Syllable contraction in Chinese. In F.-F. Tsao, & H. S. Wang (Eds.), *Chinese languages and linguistics III: Morphology and lexicon* (pp. 199-235). Taipei: Academia Sinica.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*, 596-610.

Cohn, A., Brugman, J., Crawford, C., & Joseph, A. (2005). Phonetic duration of English homophones: An investigation of lexical frequency effects. LAS 79th meeting, Oakland, CA, January 7, 2005.

Connine, C. M, Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 81-94.

Engelman, L., & Hartigan, J. A. (1969). Percentage points of a test for clusters. *Journal of the American Statistical Association*, *64*(328), 1647-1648.

Ernestus, M., & Baayen, R. H. (2003). Predicting the unpredictable: Interpretating neutralized segments in Dutch. *Language, 79*(1), 5-38.

Ernestus, M., Lahey, M., & Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustic Society of America, 120*(2), 1040-1051.

Fidelholtz, J. L. (1975). Word frequency and vowel reduction in English. *Chicago Linguistics Society*, *11*, 200-213.

Flemming, E. (2002). *Auditory representations in phonology*. New York: Routledge.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35*, 116-124.

Gahl, S. (2008). *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*(3), 474-496.

Gaygen, D. E., & Luce, P. A. (1998). Effects of modality on subjective frequency estimates and processing of spoken and printed words. *Perception and Psychophysics, 60*(3), 465-483.

Geffen, G., & Luszcz, M. A. (1983). Are the spoken durations of rare words longer than those of common words? *Memory & Cognition*, *11*(1), 13-15.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256-281.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31-38.

Hammond, M. (1999). Lexical frequency and rhythm. In M. Darnell, E. Moravcsik, F. J. Newmeyer, M. Noonan, & K. M. Wheatley (Eds.), *Functionalism and formalism in linguistics I: General papers* (pp. 329-358). Amsterdam: John Benjamins.

Hammond, M. (2004). Frequency, cyclicity, and optimality. *Studies in Phonetics, Phonology, and Morphology, 10*, 349-364.

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. M. Christie, Jr. (Ed.), *Current progress in historical linguistics* (pp. 95-105). Amsterdam: North Holland.

Hsiao, Y. E. (1999). From Taiwanese syllable contraction to the relationship between phonology, morphology and syntax: A new direction for an old issue. In Y.-M. In, Y.-L. Yang, & H.-Z. Zhan (Eds.), *Chinese languages and linguistics V: Interactions in language* (pp. 251-288). Taipei: Academia Sinica.

Hsiao, Y. E. (2002). Tone contraction. In *Chinese languages and linguistics VIII* (pp. 1-16). Taipei: Academia Sinica.

Hsu, H.-C. (2003). A sonority model of syllable contraction in Taiwanese Southern Min. *Journal of East Asian Linguistics, 12*(4), 349-377.

Hundt, M., Nesselhauf, N., & Biewer, C. (Eds.) (2007). *Corpus linguistics and the Web*. Amsterdam: Rodopi.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824-843.

Jurafsky, D., Bell, A., & Griand, C. (2002). The role of lemma in form variation. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 3-34). Berlin: Mouton de Gruyter.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P.

Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Kawamoto, A. H., Kello, C. T., Higareda, I., & Vu, J. V. Q. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset an rime duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 362-381.

Krug, M. (1998). String frequency: A cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics, 26*, 286-320.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the Web* (pp. 133-149). Amsterdam: Rodopi.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.) *Speech production and speech modeling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

Lorch, Jr., R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 149-157.

Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. *Journal of the Acoustic Society of America, 58*(4), 880-883.

Munson, B. (2007). Lexical access, lexical representation, and vowel articulation. In J. Cole & J. Hualde (Eds.), *Laboratory phonology 9* (p. 201-228). New York: Mouton de Gruyter.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech*, *Language*, *and Hearing Research*, *47*(5), 1048-1058.

Myers, J., & Tsay, J. (2003). Phonological competence by analogy: computer modeling of experimentally elicited judgments of Chinese syllables (I). (Project report). National Cheng Chung University. Research Project funded by National Science Council, Taiwan (NSC 91-2411-H-194-022).

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, *22*(2), 93-136.

Oudeyer, P.-Y. (2005). The self-organization of combinatoriality and phonotactics in vocalization systems. *Connection Science, 17*(3-4), 325-341.

Oudeyer, P.-Y. (2006). *Self-organization in the evolution of speech: Studies in the evolution of language*. Oxford, UK: Oxford University Press.

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica*, *58*, 254-275.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101-139). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177-228). Cambridge: The MIT Press.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. Berlin: Springer.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustic Society of America*, *118*(4), 2561-2569.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Racine, I., & Grosjean, F. (2002). La production du E caduc facultatif est-elle prévisible? Un début de réponse. *Journal of French Language Studies, 12* (3), 307-326.

Ranbom, L. J. & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*(2), 273-298.

Scarborough, R. A. (2004). Coarticulation and the structure of the lexicon. PhD dissertation, University of California, Los Angeles.

Schneider, W., Eschman, A., & Zucolotto, A. (2005). E-Prime reference guide. Pittsburgh: Psychology Software Tools Inc.

Steriade, D. (1999). Phonetics in phonology: The case of laryngeal neutralization. In M. K. Gordon (Ed.), *Papers in phonology 3, UCLA working papers in linguistics 2* (pp. 25-146). Los Angeles: Department of Linguistics, University of California, Los Angeles.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153-181.

Thode, J. H. C. (2002). *Testing for normality*. New York: Marcel Dekker.

Tseng, C.-C. (1999). Contraction in Taiwanese: Synchronic analysis and its connection with diachronic change. In Y.-M. In, Y.-L. Yang, & H.-Z. Zhan (Eds.), *Chinese languages and linguistics V: Interactions in language* (pp. 205-232). Taipei: Academia Sinica.

Tseng, S.-C. (2005a). Contracted syllables in Mandarin: Evidence from spontaneous conversations. *Language and Linguistics*, *6*(1), 153-180.

Tseng, S.-C. (2005b). Syllable contraction in Mandarin conversational dialogue corpus. *International Journal of Corpus Linguistics, 10*(1), 63-83.

van Bergem, D. R. (1995). Perceptual and acoustic aspects of lexical vowel reduction: A sound change in process. *Speech Communication*, *16*, 329-358.

van Son, R. J. J. H., Bolotova, O., Lennes, M., & Pols, L. C. W. (2004). Frequency effects on vowel reduction in three typologically different languages (Dutch, Finnish, Russian). In *Proceedings of INTERSPEECH 2004* (pp. 1277-1280).

Warren, P., & Marslen-Wilson, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, *41*(3), 262-275.

Wedel, A. B. (2006). Feedback and regularity in the lexicon. *Phonology, 24*, 147-185.

Wong, W. Y. P. (2006). Syllable fusion in Hong Kong Cantonese connected speech. PhD dissertation. The Ohio State University.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition*, *7*(6), 411-419.

Yu, A. C. L. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology, 24*, 187-214.

Table 1.

Examples of Southern Min syllable contraction

	Uncontracted forms	Contracted forms	Gloss
(a)	tsai^{55} + k^{h}i^{53}	tsai^{53}	‘this morning’
(b)	sio^{33} + kaŋ^{13}	siaŋ^{33}	‘the same’
(c)	u tsai^{33} + tiau^{33} k^{h}i^{21}	u tsau^{33} k^{h}i^{21}	‘be able to go’
(d)	toʔ$^{\underline{55}}$ + ui^{33}	toi^{53}	‘where’
(e)	bo^{33} + e^{33}	bue^{33} / be^{33}	‘unable’
(f)	si^{33} + tsun^{33}	sin^{33} / sun^{33}	‘moment’
(g)	tsa^{33} + bɔ^{55} laŋ^{13}	tsau^{35} / tsɔ^{35} laŋ^{13}	‘woman’
(h)	bo^{33} + iau^{53} kin^{53}	bua^{33} / bau^{33} kin^{53}	‘it doesn’t matter’
(i)	k^{h}i ^{21} + lai^{21}	k^{h}iai^{21} / k^{h}ai^{21}	‘get up’
(j)	tsit$^{\underline{55}}$ + e^{13}	tse^{55}	‘this one’
(k)	li^{21} + tsap$^{\underline{53}}$	liap$^{\underline{21}}$	‘twenty’
(l)	tsiaʔ$^{\underline{55}}$ + ni^{21} ho^{53}	tsian^{53} ho^{53}	‘so good’
(m)	ka^{21} + guan^{55} kɔŋ^{53}	ka^{35} kɔŋ^{53}	‘talk to us’
(n)	hɔŋ^{21} + laŋ^{33} p^{h}aʔ$^{\underline{53}}$	hɔŋ^{24} / haŋ^{24} p^{h}aʔ$^{\underline{53}}$	‘beaten by someone’

Table 2

Linear mixed-effects modeling of trough depth (by participants and items).

Parameter	*β*	*SE*	*t*
(Intercept)	7.817	8.522	0.917
RT	0.001	0.001	1.040
Duration	0.013	0.002	7.189 *
Maximum intensity	-0.121	0.086	-1.405
Coda obstruent	3.153	1.961	1.608
Coda nasal	-1.755	1.624	-1.081
Onset plosive	10.148	2.260	4.491 *
Onset fricative	7.453	2.286	3.260 *
Onset sonorant	-4.355	3.573	-1.219
Cross-syllable sonority rise	-1.419	1.469	-0.966
Tonal register difference	-0.696	1.556	-0.447
Tonal slope difference	-1.801	1.416	-1.272
Number of target segments	0.858	0.688	1.248
Orthogonalized log corpus frequency	0.813	1.163	0.699
Orthogonalized log subjective frequency of contracted stimuli	-2.640	2.388	-1.106
Orthogonalized log subjective frequency of uncontracted stimuli	-4.245	1.682	-2.523 *
Lexical category (content vs. function)	0.020	0.729	0.027
Orthogonalized log corpus frequency × Lexical category	-0.631	1.154	-0.547
Orthogonalized log subjective frequency of contracted stimuli × Lexical category	-3.628	2.090	-1.735
Orthogonalized log subjective frequency of uncontracted stimuli × Lexical category	2.937	1.687	1.741

NOTES: * $p < .05$. See text for details on orthogonalization.

Table 3

Linear mixed-effects modeling of log tonal distance (by participants and items).

Parameter	*β*	*SE*	*t*
(Intercept)	1280.098	362.592	3.530 *
RT	-0.081	0.051	-1.579
Duration	-0.543	0.081	-6.679 *
Maximum intensity	-8.083	4.095	-1.974 *
Coda obstruent	105.677	31.260	3.381 *
Coda nasal	-16.929	24.957	-0.678
Onset plosive	-33.748	35.105	-0.961
Onset fricative	7.029	34.457	0.204
Onset sonorant	-42.659	52.909	-0.806
Cross-syllable sonority rise	-17.277	21.563	-0.801
Tonal register difference	13.180	22.866	0.576
Tonal slope difference	30.655	20.942	1.464
Number of target segments	11.059	10.221	1.082
Orthogonalized log corpus frequency	8.522	17.176	0.496
Orthogonalized log subjective frequency of contracted stimuli	-23.345	35.656	-0.655
Orthogonalized log subjective frequency of uncontracted stimuli	-63.046	24.913	-2.531 *
Lexical category (content vs. function)	-3.003	10.709	-0.280
Orthogonalized log corpus frequency × Lexical category	-14.700	16.952	-0.867
Orthogonalized log subjective frequency of contracted stimuli × Lexical category	-19.049	30.720	-0.620
Orthogonalized log subjective frequency of uncontracted stimuli × Lexical category	-18.445	24.877	-0.741

NOTES: * $p < .05$. See text for details on orthogonalization.

Appendix

Experimental items tagged for category of morphemes (C = content morpheme, F = function morpheme) and whole words (C = containing no function morpheme, F = containing at least one function morpheme). Tone contours are indicated with digits (1 = lowest level, 5 = highest); underlined digits indicate short tones (in syllables ending in an obstruent or glottal stop). The nine items also listed in the appendix in Hsu (2003) are marked with *.

	Uncontracted forms	Contracted forms	Word gloss and category	Morpheme category	Token frequencies
1	kam^{55} + kak$^{\underline{21}}$	kaːk$^{\underline{51}}$	‘feel’ (C)	(C) + (C)	967
2	piŋ^{33} + iu^{53}	piːu^{353}	‘friend’ (C)	(C) + (C)	947
3	tak$^{\underline{21}}$ + ke^{33}	tae^{213}	‘everyone’ (C)	(C) + (C)	755
4	tsai^{33} + ĩã^{53}	tsãĩːã^{353}	‘know’ (C)	(C) + (C)	650
5	ka^{33} + ti^{33}	kai^{33}	‘oneself’ (C)	(C) + (C)	552
6	lai^{21} + te^{53}	lae^{153}	‘the interior’ (C)	(C) + (C)	387
7	miʔ^{21} + kĩã^{33}	mĩːã^{213}	‘goods’ (C)	(C) + (C)	354
8	si^{33} + kan^{55}	sian^{35}	‘time’ (C)	(C) + (C)	332
9	bun^{21} + te^{13}	bue^{213}	‘question’ (C)	(C) + (C)	299
10	tset$^{\underline{53}}$ + bɔk$^{\underline{53}}$	tseɔk$^{\underline{535}}$	‘program’ (C)	(C) + (C)	298

11	tai^{21} + tsi^{21}	taiː^{21}	‘event’ (C)	(C) + (C)	292
12	tai^{33} + uan^{13}	taiuan^{313}	‘Taiwan’ (C)	(C)	257
13	ten^{21} + ue^{33}	teue^{213}	‘telephone’ (C)	(C) + (C)	257
14	ho^{55} + tsiaʔ$^{\underline{53}}$	hoiaʔ$^{\underline{53}}$	‘delicious’ (C)	(C) + (C)	156
15	kan^{55} + tan^{55}	kaːn^{55}	‘simple’ (C)	(C) + (C)	141
16	kɔŋ^{55} + ue^{33}	kɔue^{53}	‘speak’ (C)	(C) + (C)	130
17	p^{h}ɔ^{55} + t^{h}ɔŋ^{55}	p^{h}ɔːŋ^{55}	‘ordinary’ (C)	(C) + (C)	129
18	kuan^{33} + he^{33}	kuae^{33}	‘relation’ (C)	(C) + (C)	120
19	tiɔŋ^{21} + iau^{21}	tiɔiau^{21}	‘important’ (C)	(C) + (C)	118
20	hũã^{33} + hi^{53}	hũãĩ^{353}	‘joyful’ (C)	(C) + (C)	117
21	sin^{33} + t^{h}e^{53}	sie^{353}	‘body’ (C)	(C) + (C)	104
22	tsun^{55} + pi^{33}	tsui^{53}	‘prepare’ (C)	(C) + (C)	82
23	pe^{21} + bu^{53}	peu^{153}	‘parents’ (C)	(C) + (C)	80
24	iŋ^{33} + gi^{53}	iː^{353}	‘English’ (C)	(C) + (C)	68
25	to^{33} + sia^{33}	toia^{33}	‘many thanks’ (C)	(C) + (C)	67
26	t^{h}au^{33} + ke^{55}	t^{h}aue^{35}	‘boss’ (C)	(C) + (C)	54
27	un^{21} + tɔŋ^{33}	uɔŋ^{213}	‘sports’ (C)	(C) + (C)	49
28	ts^{h}iŋ^{33} + k^{h}i^{21}	ts^{h}iː^{31}	‘clear’ (C)	(C) + (C)	47
29	ge^{21} + sut$^{\underline{53}}$	geut$^{\underline{153}}$	‘art’ (C)	(C) + (C)	40

30	tai^{21} + hak$^{\underline{53}}$	taiak$^{\underline{153}}$	‘university’ (C)	(C) + (C)	38
31	ts^{h}an^{33} + t^{h}ĩã^{55}	ts^{h}ãĩã^{35}	‘restaurant’ (C)	(C) + (C)	37
32	ket$^{\underline{53}}$ + hun^{33}	keun^{53}	‘marry’ (C)	(C) + (C)	32
33	ten^{21} + nau^{53}	teau^{153}	‘computer’ (C)	(C) + (C)	29
34	k^{h}ui^{33} + ts^{h}ia^{55}	k^{h}uiːa^{35}	‘drive a car’ (C)	(C) + (C)	23
35*	loʔ$^{\underline{21}}$ + hɔ^{33}	loɔ^{213}	‘rain’ (C)	(C) + (C)	22
36	tsiu^{33} + ni^{13}	tsiui^{313}	‘anniversary’ (C)	(C) + (C)	21
37	tɔŋ^{21} + tsɔk$^{\underline{21}}$	tɔːk$^{\underline{21}}$	‘action’ (C)	(C) + (C)	21
38	tau^{21} + iu^{13}	tauiu^{13}	‘soy sauce’ (C)	(C) + (C)	20
39	ts^{h}in^{53} + ts^{h}ai^{53}	ts^{h}iai^{535}	‘causal’ (C)	(C) + (C)	20
40	kue^{53} + bin^{53}	kuein^{535}	‘allergy’ (C)	(C) + (C)	18
41	bin^{33} + kan^{55}	bian^{35}	‘folk’ (C)	(C) + (C)	18
42*	tsai^{55} + k^{h}i^{53}	tsaiː^{53}	‘this morning’ (C)	(C) + (C)	16
43	to^{21} + ien^{53}	toien^{153}	‘director’ (C)	(C) + (C)	16
44	bu^{55} + to^{33}	buo^{53}	‘dance’ (C)	(C) + (C)	16
45	ien^{55} + tsau^{21}	ieau^{51}	‘play instrument’ (C)	(C) + (C)	14
46	iu^{33} + iŋ^{53}	iuiŋ^{353}	‘swim’ (C)	(C) + (C)	12
47	se^{53} + zi^{33}	sei^{53}	‘careful’ (C)	(C) + (C)	12
48	bak$^{\underline{21}}$ + kĩã^{21}	bãĩã^{21}	‘glasses’ (C)	(C) + (C)	9

49	k^{h}au^{55} + tsai^{13}	k^{h}auai^{513}	‘eloquence’ (C)	(C) + (C)	9
50	ki^{33} + p^{h}io^{21}	kiːo^{31}	‘airplane ticket’ (C)	(C) + (C)	9
51	bi^{55} + iɔŋ^{13}	biːɔŋ^{513}	‘beautify’ (C)	(C) + (C)	9
52	p^{h}ũã^{53} + tuan^{33}	p^{h}ũãːn^{53}	‘judge’ (C)	(C) + (C)	8
53	tsui^{55} + tsun^{53}	tsuiun^{53}	‘norm’ (C)	(C) + (C)	8
54	iɔŋ^{55} + kam^{53}	iɔam^{53}	‘brave’ (C)	(C) + (C)	7
55	tsɔŋ^{33} + kau^{21}	tsɔau^{31}	‘religion’ (C)	(C) + (C)	7
56	hap^{21} + ts^{h}ĩũ^{21}	hãĩũ^{21}	‘chorus’ (C)	(C) + (C)	6
57	hueʔ$^{\underline{53}}$ + ap$^{\underline{21}}$	hueap$^{\underline{51}}$	‘blood pressure’ (C)	(C) + (C)	6
58	zin^{33} + tsai^{13}	ziai^{313}	‘talent’ (C)	(C) + (C)	6
59	k^{h}i^{53} + hau^{33}	k^{h}iau^{53}	‘climate’ (C)	(C) + (C)	5
60	tsik$^{\underline{53}}$ + zim^{33}	tsiːm^{53}	‘duty’ (C)	(C) + (C)	5
61	t^{h}e^{55} + keʔ$^{\underline{21}}$	t^{h}eːʔ$^{\underline{51}}$	‘physique’ (C)	(C) + (C)	5
62	ti^{21} + an^{55}	tian^{215}	‘public security’ (C)	(C) + (C)	5
63	gan^{55} + kɔŋ^{55}	gaɔŋ^{55}	‘eyesight’ (C)	(C) + (C)	5
64	tsiu^{21} + giap$^{\underline{53}}$	tsiuiap$^{\underline{153}}$	‘get a job’ (C)	(C) + (C)	4
65	liɔŋ33 + sim^{55}	liɔim35	‘conscience’ (C)	(C) + (C)	4
66	ts^{h}i^{21} + k^{h}u^{55}	ts^{h}iu^{215}	‘urban district’ (C)	(C) + (C)	4
67	sat$^{\underline{53}}$ + siŋ^{55}	saiŋ^{535}	‘kill’ (C)	(C) + (C)	3

68	tɔk$^{\underline{21}}$ + p^{h}in^{53}	tɔin^{153}	‘drug’ (C)	(C) + (C)	3
69	gan^{55} + k^{h}o^{55}	gao^{55}	‘ophthalmology’ (C)	(C) + (C)	2
70	iu^{33} + lam^{53}	iuam^{353}	‘sightseeing’ (C)	(C) + (C)	2
71	ki^{33} + kim^{55}	kiːm^{35}	‘fund’ (C)	(C) + (C)	2
72	te^{21} + kiu^{13}	teiu^{213}	‘earth’ (C)	(C) + (C)	2
73	tioʔ$^{\underline{21}}$ + kip$^{\underline{21}}$	tioip$^{\underline{21}}$	‘worry’ (C)	(C) + (C)	2
74	ti^{33} + iu^{13}	tiːu^{313}	‘lard’ (C)	(C) + (C)	2
75	kua^{53} + ho^{33}	kuao^{53}	‘register at hospital’ (C)	(C) + (C)	2
76	u^{33} + ziam^{53}	uiam^{353}	‘pollution’ (C)	(C) + (C)	2
77	bi^{21} + sɔ^{21}	biɔ^{21}	‘MSG’ (C)	(C) + (C)	1
78	k^{h}i^{53} + ts^{h}uan^{53}	k^{h}iuan^{535}	‘asthma’ (C)	(C) + (C)	1
79	tsu^{53} + ts^{h}eʔ$^{\underline{21}}$	tsueʔ$^{\underline{51}}$	‘register at school’ (C)	(C) + (C)	1
80	tai^{21} + k^{h}uan^{53}	taiuan^{153}	‘loan’ (C)	(C) + (C)	1
81	hit$^{\underline{53}}$ + le^{33}	hie^{53}	‘that one’ (F)	(F) + (F)	1207
82	k^{h}o^{55} + liŋ^{13}	koiŋ^{513}	‘maybe’ (F)	(F) + (F)	756
83	e^{21} + sai^{21}	eai^{21}	‘be able to use’ (F)	(F) + (F)	521
84	i^{55} + kiŋ^{33}	iːŋ^{53}	‘already’ (F)	(F) + (F)	431
85	iŋ^{53} + kai^{55}	iai^{535}	‘ought to’ (F)	(F) + (F)	345
86	tɔŋ^{33} + zen^{13}	tɔen^{313}	‘for sure’ (F)	(C) + (F)	272

87	tu^{55} + tsiaʔ$^{\underline{21}}$	tuiaʔ$^{\underline{51}}$	‘just now’ (F)	(F) + (F)	231
88	m̩^{21} + ko^{55}	mo^{215}	‘however’ (F)	(F) + (F)	177
89	kɔk$^{\underline{53}}$ + ui^{33}	kɔui^{53}	‘everyone’ (F)	(F) + (C)	126
90*	loʔ$^{\underline{21}}$ + kʰi^{21}	loi^{21}	‘fall down’ (F)	(C) + (F)	116
91	i^{55} + au^{33}	iau^{53}	‘afterwards’ (F)	(F) + (F)	115
92	u^{21} + kau^{21}	uau^{21}	‘enough’ (F)	(F) + (C)	106
93*	kʰi^{21} + lai^{21}	kʰiai^{21}	‘get up’ (F)	(C) + (F)	104
94*	hɔ^{21} + laŋ^{33}	hɔːŋ^{213}	‘by someone’ (F)	(F) + (C)	71
95*	ka^{21} + gua^{21}	kaː^{21}	‘to me’ (F)	(F) + (F)	57
96*	na^{53} + e^{33}	nae^{53}	‘how come’ (F)	(F) + (F)	48
97*	to^{55} + ui^{33}	toui^{53}	‘where’ (F)	(F) + (C)	45
98*	be^{21} + hiau^{53}	beiau^{153}	‘not understand’ (F)	(F) + (C)	35
99	tsɔŋ^{55} + kiɔŋ^{33}	tsɔiɔŋ^{53}	‘totally’ (F)	(F) + (F)	25
100	kiŋ^{53} + zien^{13}	kiːen^{513}	‘unexpectedly’ (F)	(C) + (F)	25
101	hɔ^{21} + siɔŋ^{33}	hɔiɔŋ^{213}	‘each other’ (F)	(F) + (F)	22
102	kɔŋ^{53} + kue^{21}	kɔue^{51}	‘have talked’ (F)	(C) + (F)	19
103	tsi^{33} + ha^{33}	tsia^{33}	‘below’ (F)	(F) + (F)	19
104	sui^{33} + si^{13}	suiː^{313}	‘anytime’ (F)	(F) + (C)	18
105	tai^{21} + siŋ^{55}	taiːŋ^{215}	‘beforehand’ (F)	(C) + (F)	18

106	ki^{33} + kan^{55}	kian^{35}	‘period of time’ (F)	(C) + (F)	17
107	k^{h}o^{55} + si^{33}	k^{h}oi^{53}	‘but’ (F)	(F) + (F)	14
108	huan^{55} + tsiŋ^{21}	huaiŋ^{51}	‘anyway’ (F)	(F) + (F)	13
109	be^{21} + hu^{21}	beu^{21}	‘not catch up’ (F)	(F) + (C)	11
110	pit$^{\underline{53}}$ + iau^{21}	piːau^{51}	‘be required to’ (F)	(F) + (F)	9
111	tsi^{53} + tsio^{53}	tsiːo^{535}	‘at least’ (F)	(F) + (C)	8
112	taŋ^{33} + tse^{13}	tae^{313}	‘together’ (F)	(F) + (C)	8
113	piŋ^{33} + siɔŋ^{13}	piːɔŋ^{313}	‘usually’ (F)	(F) + (F)	7
114	tsiɔŋ^{33} + kin^{33}	tsiɔin^{33}	‘nearly’ (F)	(F) + (C)	7
115	tsi^{55} + u^{33}	tsiu^{53}	‘only’ (F)	(C) + (F)	6
116	lɔŋ^{53} + tioʔ^{21}	lɔioʔ^{51}	‘collide with’ (F)	(C) + (F)	5
117	tsiaʔ$^{\underline{21}}$ + tiau^{33}	tsiaiau^{213}	‘finish eating’ (F)	(C) + (F)	3
118	tsiŋ^{33} + kiŋ^{55}	tsiːŋ^{35}	‘already’ (F)	(F) + (F)	3
119	pue^{21} + au^{33}	pueau^{213}	‘at the back’ (F)	(C) + (F)	1
120	ho^{33} + pit$^{\underline{21}}$	hoit$^{\underline{31}}$	‘no need’ (F)	(F) + (C)	1

Figure 1.

Spectrogram of a partially contracted disyllable, with the associated intensity contour (dB) superimposed.

Figure 2.

Density plots (smoothed histograms) for trough depth for each of the 20 participants (labeled s01 through s20).

Figure 3.

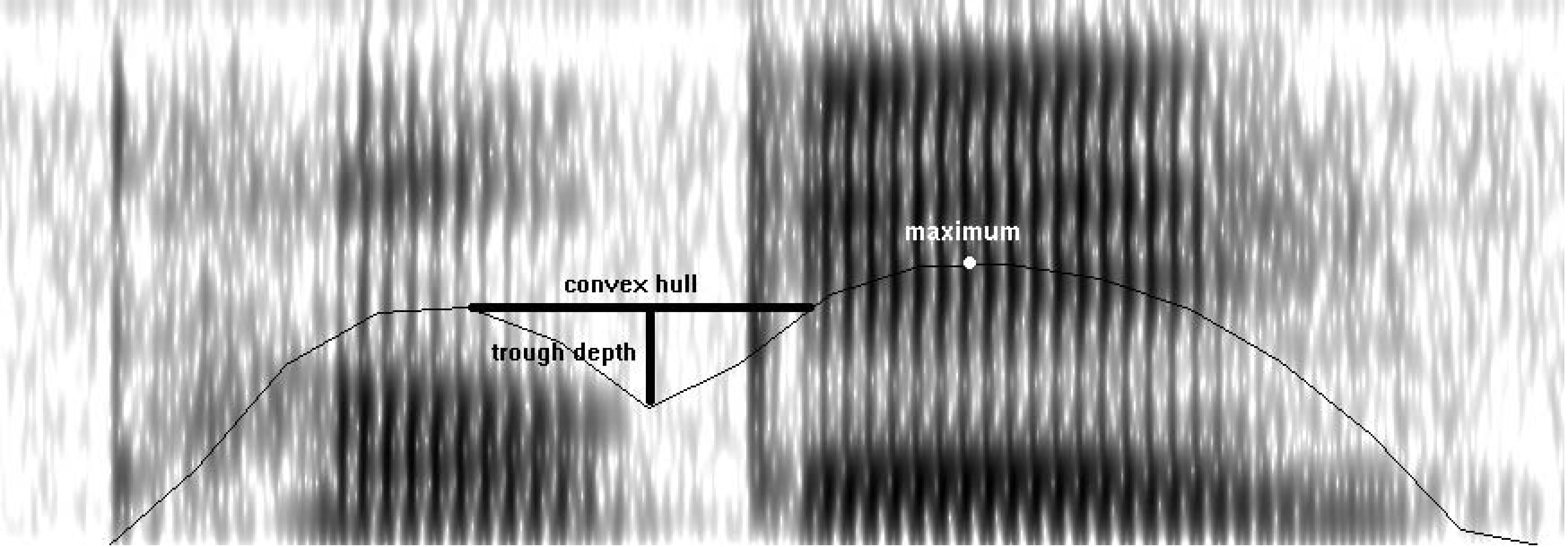
Loess plot of log frequency estimates of uncontracted forms to trough depth. Black triangles represent production tokens of items listed in Hsu (2003).

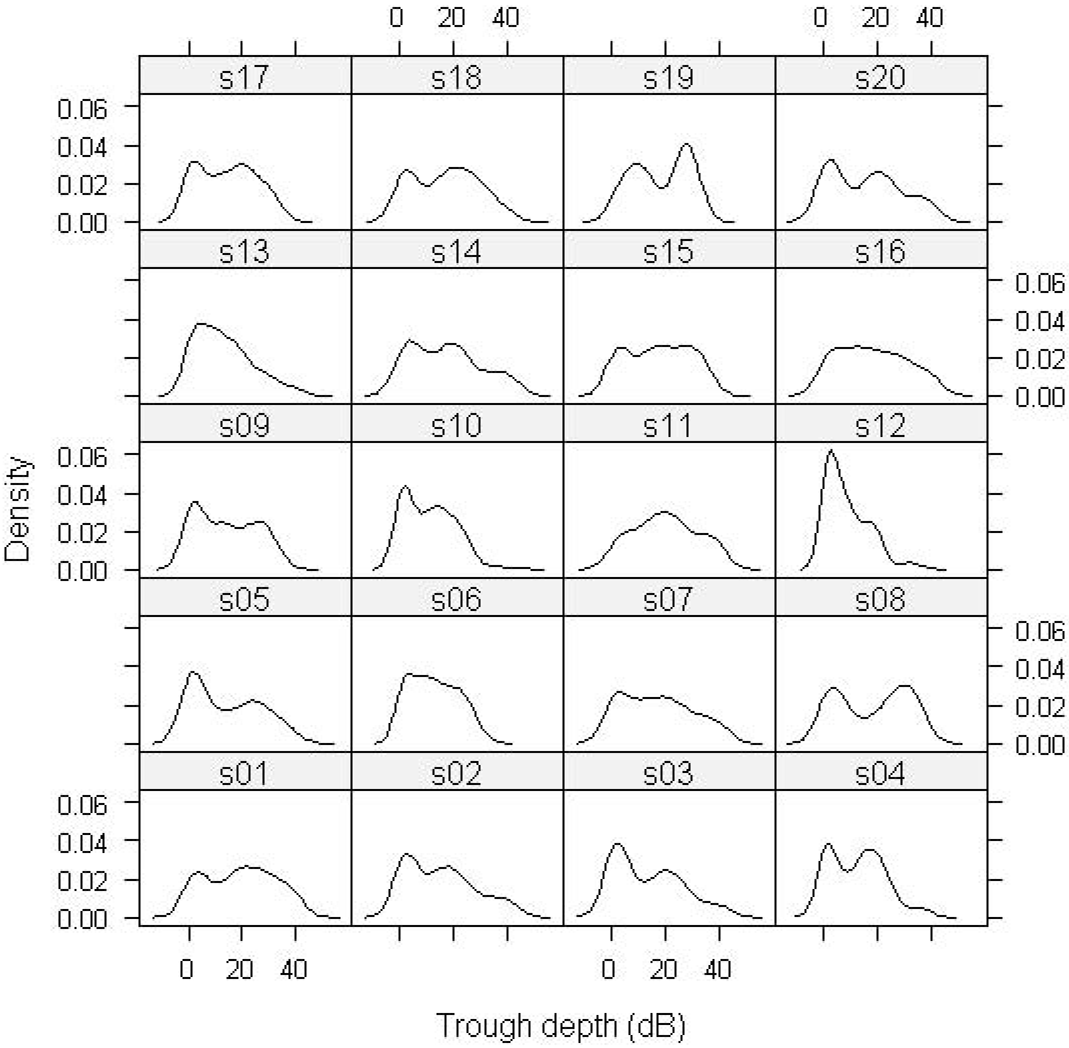
Figure 4.

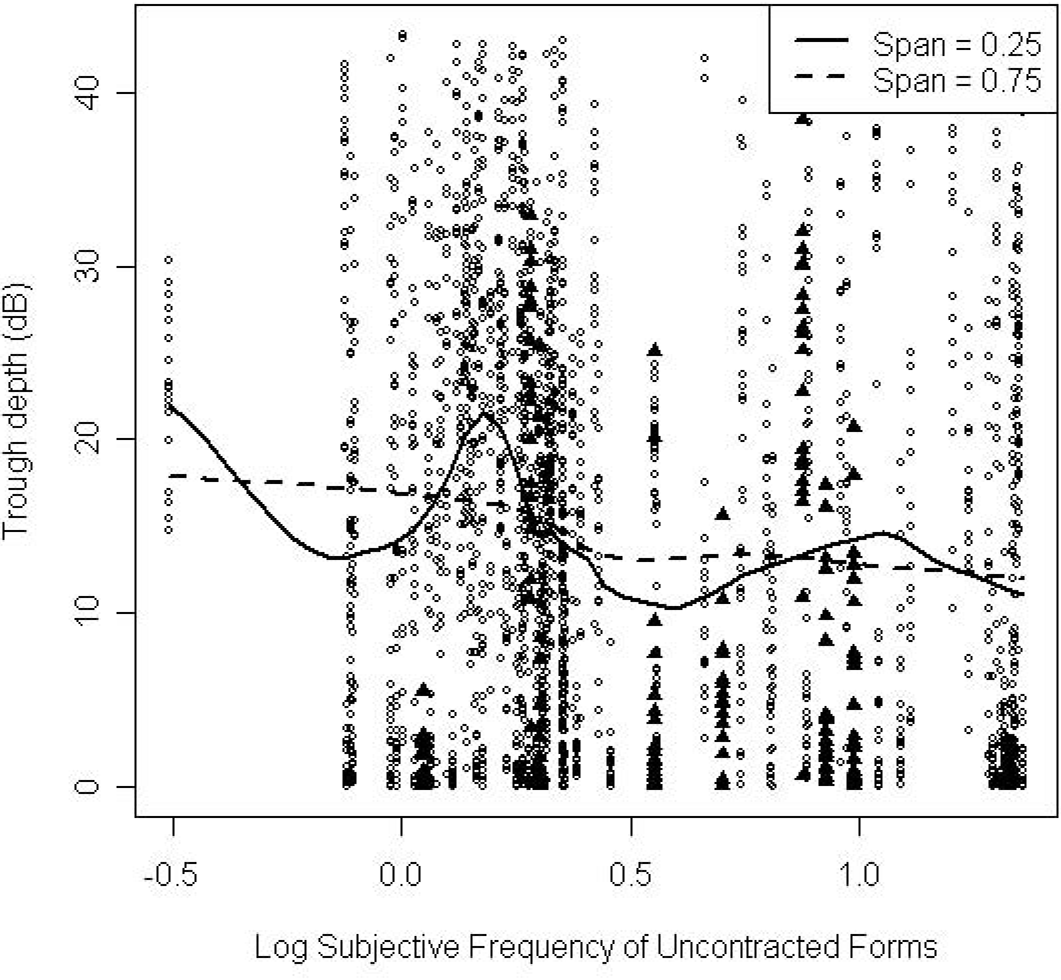
Density plots (smoothed histograms) for log tonal distance for each of the 20 participants (labeled s01 through s20).

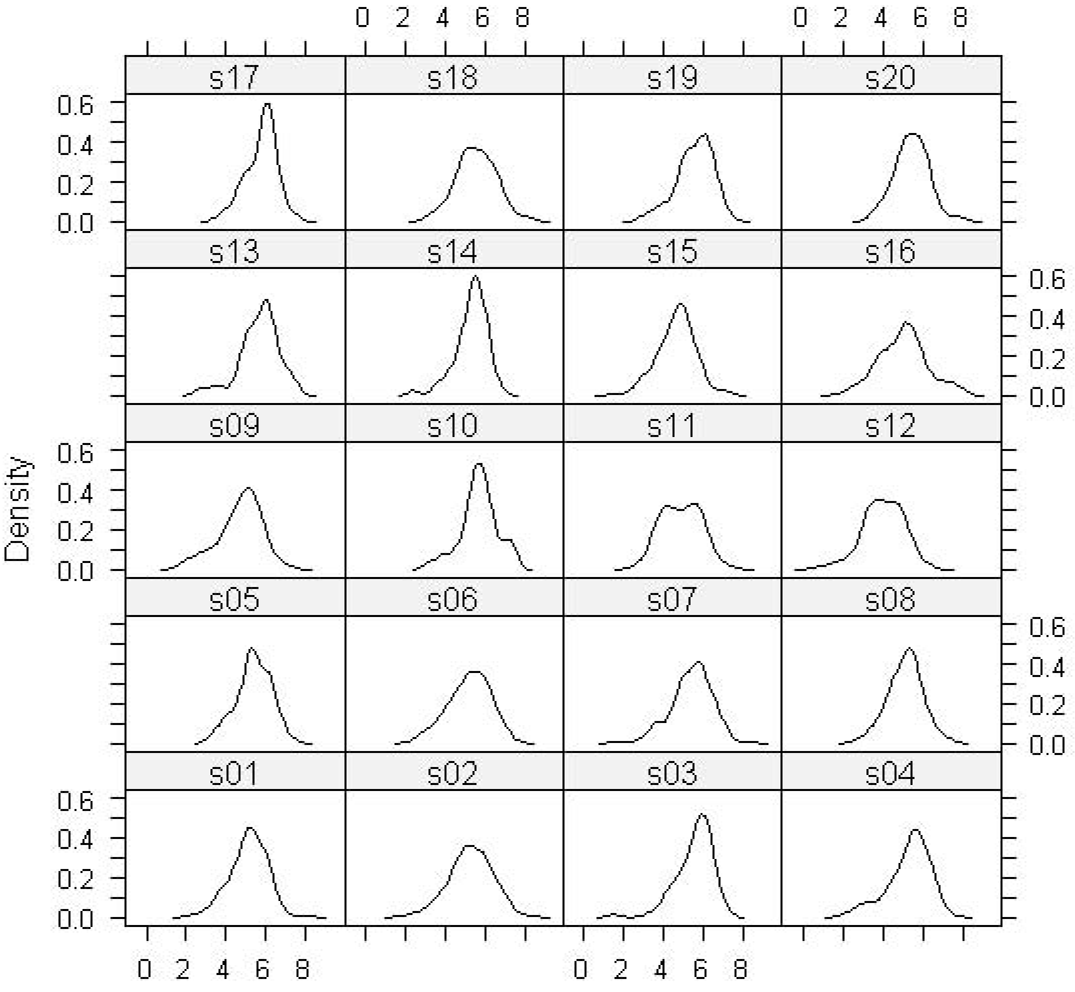
Figure 5.

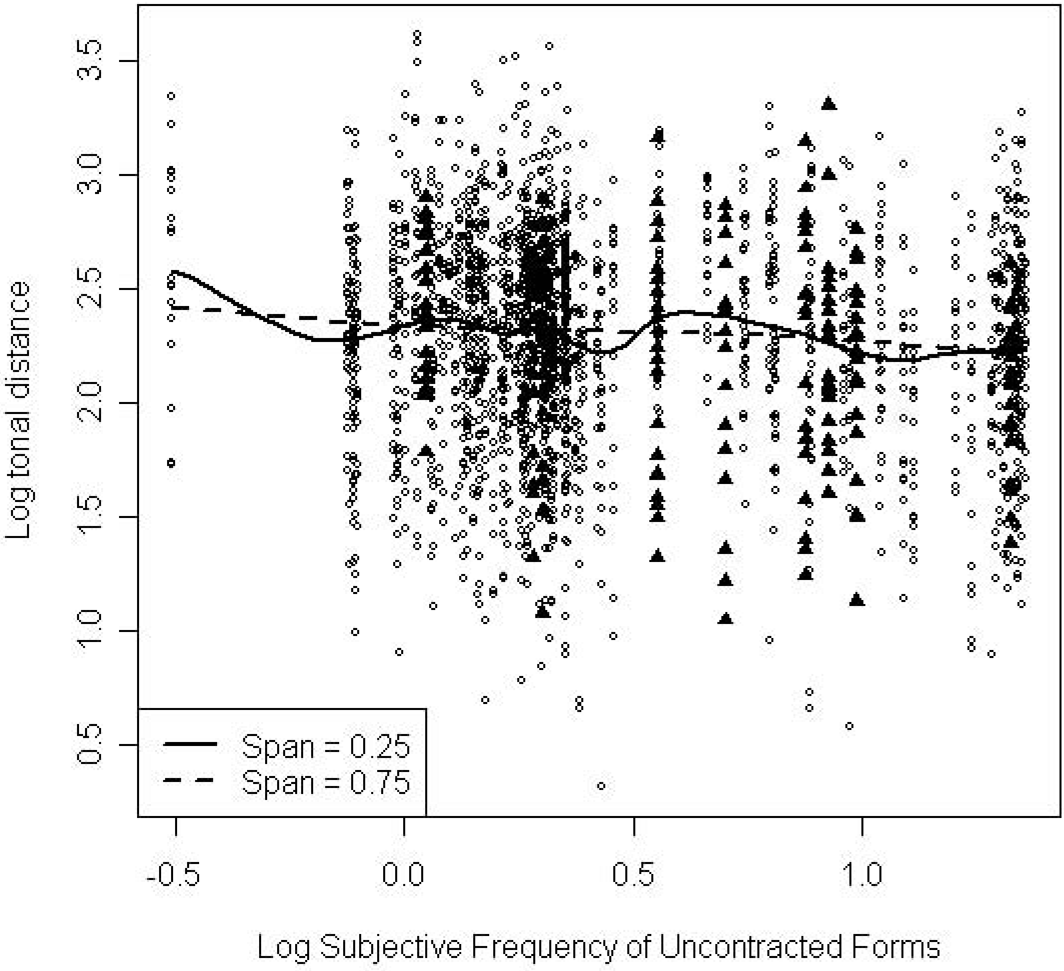
Loess plot of log frequency estimates of uncontracted forms to log tonal distance. Black triangles represent production tokens of items listed in Hsu (2003).


maximum
convex hull
trough depth


0
20
40
s17
s18
s19
s20
s13
s14
s15
s16
s09
s10
s11
s12
s05
s06
s07
s08
s01
s02
s03
s04
0.06
0.04
0.02
0.00
Density
Trough depth (dB)


Span = 0.25
Span = 0.75
Trough depth (dB)
0
10
20
30
40
-0.5
0.0
0.5
1.0
Log Subjective Frequency of Uncontracted Forms


s17
s18
s19
s20
s13
s14
s15
s16
s09
s10
s11
s12
s05
s06
s07
s08
s01
s02
s03
s04
Density
0.0
0.2
0.4
0.6
0 2 4 6 8
Log tonal distance



Log tonal distance
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-0.5
0.0
0.5
1.0
Log Subjective Frequency of Uncontracted Forms
Span = 0.25
Span = 0.75