Tonal effects on perceived vowel duration

Alan C. L. Yu
University of Chicago
aclyu@uchicago.edu

Abstract
Cross-linguistically, there is a tendency for vowel duration to be inversely related to the approximate average $f_0$; all else being equal, vowels on low tones are longer than those on high tones, while vowels on rising tones are longer than those on falling tones. This type of interplay between tonal contrasts and duration are commonly reflected in the world’s languages. This paper reports a study investigating the role of perception in the effect of tone on duration. Both $f_0$ slope and height are found to exert an influence on perceived duration. Syllables with dynamic $f_0$ are generally perceived as longer than syllables with flat $f_0$. Syllables with higher $f_0$ are heard as longer than syllables with lower $f_0$. It is argued that the mismatch in the effects of $f_0$ height on perceived duration and acoustic duration is due to hypercorrection. Perceived duration equality is achieved by acoustic duration adjustment. While this type of duration compensation might be characteristic of level tone perception, no such effect is apparent in the case of dynamic tone realization. Regardless of perceived or acoustic duration, contour tone syllables are generally longer than flat tone syllables. The difference between the effects of dynamic and flat $f_0$ patterns on perceived and acoustic durations suggests that the perceptual mechanism for dynamic $f_0$ is different from that of level $f_0$. 
Tonal effects on perceived vowel duration

1. Introduction
The relationship between tone and its tone-bearing unit can be described as symbiotic. While tonal contrasts are realized primarily by differences in fundamental frequency ($f_0$) height or contour, they may also involve systematic differences in duration. The perception of duration, on the other hand, may be influenced by the $f_0$ pattern. This type of interplay between tonal contrasts and duration are commonly reflected in the world’s languages. It is repeatedly observed that dynamic tones tend to be restricted to phonetically long sonorous segments (Gordon, 1999, 2001; Yu, 2003, 2004; Zhang, 2001). Cross-linguistically, there also seems to be a tendency for vowel duration to be inversely related to the approximate average $f_0$. In particular, vowels on low tones are longer than those on high tones, while vowels on rising tones are longer than those on falling tones (Gandour, 1977). For example, various studies have shown that Mandarin Tones 1 (55) and 4 (51) tend to be shorter than Tone 2 (35), which in turn tends to be shorter than Tone 3 (214) (Brotzman, 1964; Ho, 1976; Howie, 1974, 1976; Nordenhake & Svantesson, 1983). Kong (1987) found similar results for Cantonese, although he observes that the high (55) and low (21) tones tend to be shorter than the mid tone (33). More curiously, however, is the fact that tone systems also often correlate with or, in some cases, evolve into non-tonal languages with length opposition. In Southern Thai, for example, short vowels became long under non-falling tones (i.e. low-rising, low-level, and mid-level), while long vowels became short under falling tones (i.e. high-falling, mid-falling, & low-falling). Kwon (2003) argues that many long vowels in Korean today correspond to Middle Korean rising tone syllables. Middle Korean has three lexical tones (i.e. low, high and rising) but no phonemic vowel length contrast. Modern Standard Korean, on the other hand, has no tonal contrast but developed a phonemic vowel length difference. As shown in (1), mal ‘word’ was realized with a short vowel and a rising tone in Middle Korean, but with a long vowel and no lexical tone in Modern Korean. No such long vowel is found in the Modern Korean reflexes of Middle Korean words with high or low tone.

(1) Middle Korean Modern Korean
 mal (L) ‘horse’ mal
 mal (H) ‘a unit of measure’ mal
 mal (R) ‘word’ ma:l

Similarly, a vowel length contrast in the Weert dialect of Dutch corresponds to an accentual difference in neighboring dialects. The examples in (2), reproduced from Heijmans (2003), show that in the Baexem dialect of Dutch, singular and plural are marked by a difference in accent. In a declarative context, Accent II is realized with a falling-rising pattern, while a falling pattern is observed in the accented syllable in Accent I. In Weert, on the other hand, the singular and plural, which show similar $f_0$ contours, are distinguished in terms of vowel length. The singular contains a vowel that is about twice as long as the vowel in the plural.
While many diachronic changes and correspondence between tone and vowel length often involve dynamic tones (as illustrated by the Korean and the Weert dialect of Dutch cases), level tones show similar symbiotic relationship with vowel length as well. As noted earlier, Gandour (1977) reports that vowels on low tones are generally longer than those on high tones. This correlation is found in several diachronic changes as well. In the history of Cantonese, vowel length correlates with a tonal split in the Cantonese upper register tone. Thus, in plosive-final syllables, a high upper register tone (55) occurs in syllables containing a short vowel while a mid upper register tone (33) occurs in syllables containing a long vowel (Lee, 1993). No such correlation is found in the other syllable types, however. Similarly, in Hu, a Mon-Khmer language, high tones occur in words containing historically short vowels while low tones occur in words with historical long vowels; Hu does not have a vowel length distinction synchronically (Svantesson, 1988).

Many researchers have characterized the observed differences in vowel duration with respect to the level, extent, and direction of fundamental frequency as physiologically conditioned. For example, studies on the maximum speed with which pitch can be changed found that subjects were able to perform pitch drops considerably faster than pitch elevation even though the pitch ranges are comparable (Ohala & Ewan, 1973; Sundberg, 1973, 1979; Xu & Sun, 2002). However, such physiological explanation offers no explanation for the durational differences among the syllables with level tones.

Studies on the perceptual influence of $f_0$ on duration offer potential explanatory alternatives, however. Lehiste (1976), for example, found that listeners judge a dynamic (falling-rising or rising-falling), as supposed to a flat $f_0$ pattern, to be longer even when the stimuli are of equal acoustic durations. This finding was replicated in other studies on perceived duration of isolated vowels (Pisoni, 1976; Wang, Lehiste, Chuang, & Darnovsky, 1976). Brigner (1988) found that the duration of a high frequency pure tone was perceived to be longer than the duration of a low frequency tone. These results have serious implications for the perception of other phonological contrasts. For example, cross-linguistically, vowels before voiceless, fortis stops tend to be shorter than vowels before voiced, lenis stops. Peterson & Lehiste (1960), for example, found that the ratio of the duration of vowels before voiceless consonants to the duration of the vowels before voiced consonants is about 2:3 in American English. Subsequent studies have shown that a dynamic $f_0$ contours in a preconsonantal vowel shifts phoneme perception towards more lenis (van Dommelen, 1989; Gruenenfelder & Pisoni, 1980; Kohler, 1985; Lehiste, 1985). The usual explanation for this effect refers to the findings of Lehiste and postulates a longer perceived vowel duration connected with $f_0$ movement. The longer perceived duration would therefore bias the phoneme-judgment towards more lenis percepts.

While $f_0$ effect on perceived vowel duration is invoked in explaining perception of consonantal voicing contrast, surprisingly, no studies have tried to explore the perceptual
foundation of the typological connection between lexical tone and vowel duration. This study represents a first step toward such an account. The specific goal of this study is two-fold. While previous studies have found that dynamic movement in fundamental frequency affects English listener’s perceived duration, no studies have considered the effect of $f_0$ height on perceived duration.\footnote{While Brignier’s study establishes a positive correlation between pitch and perceived duration, the linguistic relevance of his study is diminished by the fact that his stimuli is based on pure tone and that the frequency tested are above and beyond normal human pitch range (i.e. 4000Hz and 500Hz). It is not clear whether the effect is observed with more linguistically natural materials (i.e. in a low and much narrower pitch range and CV stimuli rather than just pure tones).} This experiment widens the scope of previous investigations by including stimuli with different $f_0$ heights, as well as slope. To the extent that such a perceptual influence exists, a related question to be addressed in the experiment is whether there is a ceiling effect with respect to such an influence. That is, is the strength of the perceived duration lengthening effect connected with $f_0$ movement and height stable regardless of the acoustic duration of the stimuli or will the effect diminish as acoustic duration changes. This paper is organized as follows: Section 2 presents the methods of the experiment while Section 3 reports the results. The interpretation of the results and its implication with respect to the main concerns of this study appear in Section 4. Conclusions are given in Section 5.

2. Methodology

Subjects: Fifteen native speakers of English, all students at the University of Chicago, were paid a nominal fee to participate in the experiment. Subjects are screened for past experience with tonal languages. None of them report any speech or hearing problems.

Stimuli: A 300 ms [pa] syllable was synthesized using SynthWorks. A 3-step duration continuum was created with 100 ms. decreasing increments: 300, 200, and 100 ms. The $f_0$ of the syllable was then manipulated to make five stimuli of varying $f_0$ contours. The values of the five $f_0$ contours (referred to as tones henceforth) are given in Table 1. Level tones are steady throughout and rising or falling tones are linearly interpolated across the syllable.

<table>
<thead>
<tr>
<th>Name</th>
<th>$f_0$ beginning (Hz.)</th>
<th>$f_0$ ending (Hz.)</th>
<th>average $f_0$ (Hz.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>85</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>33</td>
<td>115</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>55</td>
<td>145</td>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td>15</td>
<td>85</td>
<td>145</td>
<td>115</td>
</tr>
<tr>
<td>51</td>
<td>145</td>
<td>85</td>
<td>115</td>
</tr>
</tbody>
</table>

Table 1 Following the transcription conventions for Chinese tones, a low tone will be called 1, a mid 3, and a high 5.

Procedures: Subjects were asked to rate each stimulus for duration on a 7-point scale, with 1 being the shortest duration and 7 being the longest. Subjects first completed a practice session of nine trials with the 33 tone stimuli at three durations presented three
times in a random order. After practice, subjects completed an experimental session of 300 trials (5 tones x 3 durations x 20 blocks) with the order of trials randomized within each block. The presentation of the stimuli was controlled by the subject. Presentation of the next item starts after a response is recorded. Subjects were allowed to take a short break after every five blocks, with a mandatory 2-minute break after the first 10 blocks. The experiment was administered using E-Prime. Stimuli were presented binaurally with high quality headphones in a quiet room. The sound will be adjusted to a comfortable level during the practice session.

3. Results
In order to facilitate cross-subject comparison, the data was first normalized by transforming the rating value to a z-score scale. Figure 1 summarizes the z-score transformed duration rating results across tones, while Figure 2 summarizes the results across different stimulus durations. To determine the overall effect of tone and stimulus duration group upon the listener perceived duration responses, a two-way repeated-measures analysis of variance with TONE (11, 15, 33, 55, 51) and LENGTH (300, 200, 100 ms.) as within-subject factors was conducted. The results of the ANOVA analysis are given in Table 1. There are significant main effects of stimulus duration and tone but no significant interaction effect is observed.

![Figure 1 Mean z-score transformed duration rating across five tones by English speakers.](image-url)

<table>
<thead>
<tr>
<th>Tone</th>
<th>Mean z-score(Duration Rating)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>-0.3</td>
</tr>
<tr>
<td>33</td>
<td>-0.1</td>
</tr>
<tr>
<td>55</td>
<td>0.1</td>
</tr>
<tr>
<td>15</td>
<td>0.2</td>
</tr>
<tr>
<td>51</td>
<td>-0.2</td>
</tr>
</tbody>
</table>
Simulus duration (msec.)

![Graph showing mean z-score transformed duration rating by stimulus duration.]

Figure 2 Mean z-score transformed duration rating by stimulus duration.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>14</td>
<td>453.878</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>.002</td>
<td>.967</td>
</tr>
<tr>
<td>DUR</td>
<td>2</td>
<td>3088.503</td>
<td>.000</td>
</tr>
<tr>
<td>TONE</td>
<td>4</td>
<td>41.771</td>
<td>.000</td>
</tr>
<tr>
<td>DUR * TONE</td>
<td>8</td>
<td>1.441</td>
<td>.174</td>
</tr>
<tr>
<td>Error</td>
<td>4478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>4492</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 ANOVA of z-score transformed duration rating

Because tones had a statistically significant main effect for perceived duration ratings, a post-hoc analysis, using Bonferroni adjusted alpha levels of 0.01 per test (0.05/5), was conducted for this variable using Student-Newman-Keuls (SNK) tests. The SNK tests were used because they do not produce overly conservative results when many levels of a single independent variable are compared. The results of the SNK tests are shown in Table 3, which list the tone and their normalized z-score ratings in ascending order. In other words, the tones are ordered from the least amount of duration effect on perceived duration (therefore receiving negative z-score ratings) to the most effect (thus receiving positive z-score rating). Table 3 also shows the SNK result grouping. In SNK tests, means that are assigned to the same subset are not significantly different from each other. Three subgroups are observed. Tone 11 and Tone 33 each formed their own subgroup, while the 55, 51, and 15 tones form the other subgroup.
### Table 3 Mean transformed z-score duration rating of five tones.

<table>
<thead>
<tr>
<th>TONE</th>
<th>Subset 1</th>
<th>Subset 2</th>
<th>Subset 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.00</td>
<td>-.2045983</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.00</td>
<td></td>
<td>-.0747133</td>
<td></td>
</tr>
<tr>
<td>51.00</td>
<td></td>
<td></td>
<td>.0522410</td>
</tr>
<tr>
<td>55.00</td>
<td></td>
<td></td>
<td>.1065372</td>
</tr>
<tr>
<td>15.00</td>
<td></td>
<td></td>
<td>.1207909</td>
</tr>
<tr>
<td>Sig.</td>
<td>1.000</td>
<td>1.000</td>
<td>.062</td>
</tr>
</tbody>
</table>

4. Discussion

The experimental results suggest that the height and slope of fundamental frequency can significantly affect the listener’s perceived duration of speech sounds. The results on the perceived duration of syllables with dynamic $f_0$ are consistent with previous findings. That is, dynamic $f_0$ generally lengthens subjective syllable duration. No significant difference between the perceived durations of syllables with falling and rising $f_0$ is found, however, although syllables with a falling $f_0$ contour are generally shorter than syllables with a rising $f_0$ contour. The matching between results of previous production studies and results of the present perception test suggests that the positive correlation between dynamic $f_0$ and syllable duration found in synchronic typological patterns and diachronic changes may not be merely a matter of physiological constraints on tonal implementation, but might have arisen, at least in part, as the result of perceptual skewing that is inherent in the perception of dynamic $f_0$ movement. The lack of interaction between acoustic duration of the stimuli and tone strongly suggests that the perceptual skewing is robust even when the stimulus is inherently long. Applying this reasoning in the diachronic context, for example, the Middle Korean speaker might have heard *mal* with a rising tone as longer than the *mal* syllable with flat pitch. The continued parsing of *mal*(R) as long might have prompted some Korean speakers to misinterpret the extra length as intentional. The hyporrection of the tonal effect on perceived duration might have resulted in the creation of distinctive vowel length in Korean today.

The viability of this theory, however, is dependent on our interpretation of the results concerning the perception of syllables with a flat $f_0$ pattern. While the experimental results are consistent with previous findings that syllables with a dynamic $f_0$ pattern are perceived as longer than syllables with a flat $f_0$ pattern, this advantage of dynamic $f_0$ over flat $f_0$ pattern is only significant when compared to flat $f_0$ patterns with mid or low $f_0$ heights. No significant difference is found between syllables with dynamic $f_0$ and those with a flat high $f_0$ pattern. This finding suggests that the effect of flat $f_0$ patterns on perceived duration is not monolithic. Syllables with flat high $f_0$ are perceived as significantly longer than those with a flat mid $f_0$ pattern. Syllables with flat low $f_0$ pattern are perceived as significantly shorter than syllables with a mid or high flat $f_0$ profile even though the acoustic durations are identical. The discovery of the differential effects of $f_0$ height (of a flat $f_0$ pattern) on perceived duration is significant in several respects. First, the results point to a potential gradient effect of $f_0$ height on perceived duration. That is,
the higher the $f_0$, the longer the perceived syllable duration. This positive correlation between $f_0$ height and perceive duration is unexpected, however, since previous production studies have found that, all else being equal, syllables with high tone tend to be shorter than syllables with low tone. That is, an inverse correlation between $f_0$ height and acoustic duration is observed. Thus, the disassociation of $f_0$ height with perceived duration and acoustic duration demands explanations.

One interpretation is that the $f_0$ effect on perceived vowel duration may be confounded by the influence of stress perception in English. Stress in English is generally correlated with a rise in $f_0$, lengthened duration, and increase in intensity. Thus, the English listeners might associate a high $f_0$ pattern with stress, which might in turn prompt a longer duration judgment. Similarly, since an unstressed syllable in English is correlated with low $f_0$ and short vowel duration, the listeners might have perceived syllables with low flat $f_0$ as unstressed, thus responded with low duration rating judgments. Analogously, a mid $f_0$ pattern might be perceived as reflecting secondary stress, thus such syllables are given medium duration ratings by the listeners. On the strength of the English results alone, it is difficult to rule out this potential confound. However, there are reasons to believe that the results are more systematic than the stress-induced-lengthening hypothesis might suggest.

As noted earlier, Brigner (1988) found that the perceived duration disparity connected with tone height is observed at pitch levels far beyond the range of linguistic $f_0$ modulations (e.g., 500Hz. & 4000Hz.). His results are less amendable to the stress-induced perceptual lengthening explanation since his stimuli are machine-generated pure tones, rather than based on linguistic materials and the $f_0$ of stressed syllables do not normally go beyond 500Hz. in any case. Thus, the effect observed in Brigner’s study calls for a psychoacoustic or psychophysical explanation. To the extent that the results obtained in the current study are comparable to Brigner’s results, both sets of data should be explained by the same mechanism.

![Figure 3](image)

**Figure 3** Mean z-score transformed duration ratings across five tones by Cantonese speakers.
Furthermore, preliminary results from a pilot study looking at the effect of $f_0$ slope and height on perceived duration in speakers of tone languages found similar positive correlation between $f_0$ height and perceived duration. For example, Figure 3, which summarizes the duration rating results of five Cantonese speakers (the experimental conditions are identical to the experiment reported in this paper), shows a positive correlation between $f_0$ height and perceived duration even when the speakers’ native tongue is tonal and has no distinctive stress. The fact that a positive correlation is found between $f_0$ height and perceived duration across speakers of typologically diverse languages (i.e. English and Cantonese) and in pitch ranges beyond normal linguistic parameters as observed in Brigner’s study suggests that the perceived duration effect associated with different $f_0$ height might stem from deep-seeded motivations that go beyond ad hoc explanations that are language-specific.

One theory was advanced in Brigner (1988), which suggests that the effect of pitch level on perceived duration is motivated by perceived space and perceived time covariation. Two sets of assumptions are crucial to his proposal. First, previous studies have found that a change in perceived space results in a change in perceived time and vice versa (Abe, 1935; Brigner, 1984; Helson & King, 1931). In particular, a decrease in perceived time may occur with a increase in perceived size/space (Brigner, 1984) (although he also cited other studies (Abe, 1935; DeLong, 1981) that arrived at the opposite conclusion, i.e. a decrease in perceived time may occur with a decrease in perceived size/space). Second, low-frequency tones are perceived as large while high-frequency tones are perceived as small (Stevens, 1934). Thus, a low-frequency tone, which is perceived as large, is expected to have a shorter perceived duration than a high-frequency tone, which is perceived as small. If this \textit{perceived-pitch-as-perceived-size} theory is correct, then the positive correlation between $f_0$ height and perceived duration is explained: the lower the $f_0$, the larger the perceived size; the larger the perceived size, the shorter the perceived duration.

At this point, a paradox arises. If syllables with higher $f_0$ are perceived as longer than syllables with lower $f_0$, then why are high tones more often associated with short vowels, while low tones more often associated with long vowels in both synchronic and diachronic domains? In essence, how is it possible that the production observations directly contradict the perceptual results?

Here, we suggest that the surface perception/production disassociation between tone and duration is the result of hypercorrection (Lindblom, 1990; Ohala, 1993). The theory goes as follows: speakers are aware of the perceptual skewing induced by the presence of pitch. In order to normalize the effect, they compensate in the production to achieve perceptual equality. That is, a vowel with low tone may be realized with a longer duration to compensate for the fact that low tone triggers an automatic shorter duration percept. On the other hand, a vowel with high tone may be realized with a shorter duration to reduce the perceived duration lengthening effect. This hypercorrective effect is more forcefully demonstrated in Brigner’s original study on the effect of pitch on perceived duration. Using the method of adjustment, he tested the effect of pitch level on perceive duration by asking the subject to judge the duration equivalence of a high tone (4000 Hz) and a low tone (500 Hz) stimuli. After each presentation of tones, the experimenter asked the subject whether the duration of the high-frequency tone seems longer, shorter, or equal to the duration of the low-frequency tone. The experimenter then adjusts the
duration of the high-frequency tone in steps of approximately 27 msec, ascending from 110 msec. or descending from 659 msec. The tones were then repeated after each adjustment. This procedure continued until the subject report an equality of durations. He found that the duration of the low-frequency tone was perceived to be equal to that of a high-frequency tone when the latter had a shorter duration. On average, the 500-Hz tone (at a constant duration of 385 msec.) was reported as equal in duration to the 4000-Hz tone when the latter had duration of approximately 361 msec. (averaged across seven subjects). Thus when the subjects are given a range of duration options, they invariably prefer a shorter high tone stimulus when judged against a fixed duration low tone stimulus.

Such normalization effects are readily observed in other domains as well. For example, Birnbaum and Veit (1974) found that a small size box is judged heavier than a bigger box of equal physical weight. Hombert (1978) found that a low vowel [a] has a tendency to be judged higher in pitch than the high vowels [i] or [u] even when their $f_0$ were in fact equal. Hombert suggests this might be a normalization response to the intrinsic $f_0$ effect with respect to vowel height. That is, under normal condition, higher vowels tend to be associated with higher $f_0$ while lower vowels with lower $f_0$ (e.g., Whalen & Levitt, 1995). To compensate for this effect, speakers normalize for the influence of vowel height on $f_0$ in their perception. The hypercorrective mechanism observed in this study is most reminiscent of the compensatory listening proposal advanced by Gussenhoven (LabPhon 9). He found that higher vowels sound longer than lower vowels when acoustic durations are equal. He reasons that this is the result of listeners compensating for the intrinsic vowel duration effect whereby higher vowels are intrinsically shorter than lower vowels. To be sure, there is an important qualitative difference between the normalization effect proposed in this study and Gussenhoven’s compensatory listening proposal. In particular, while the compensatory effect in Gussenhoven’s study is evident in the experimental results itself, in the present context, we argue the compensatory mechanism is manifested in speakers’ production and in case of diachronic changes. As such, the results obtained in the current study represent the normative situation in speech perception. The adjustment is observed in the production and fossilized in diachronic changes instead. That is, the inverse correlation between tone and syllable duration found in previous production studies is understood to be the result of compensatory duration adjustment. Such compensation and adjustment may result in lengthening or shortening, or a mixture of both, depending on whether it is the high or the low tone syllables that are adjusted.

5. Conclusion
This study investigated the perceptual relation between tone and perceived duration. Both $f_0$ slope and height exert an influence on perceived duration. Syllables with dynamic $f_0$ are generally perceived as longer than syllables with flat $f_0$. Syllables with higher $f_0$ are heard as longer than syllables with lower $f_0$. It is argued that the mismatch in the effects of $f_0$ height on perceived duration and acoustic duration is due to hypercorrection. Perceived duration equality is achieved by acoustic duration adjustment. While this type of duration compensation might be characteristic of level tone perception, no such effect is apparent in the case of dynamic tone realization, however. That is, while syllables with dynamic tone are generally perceived as longer than syllables with flat tone, no
disassociation effect is observed in syllables with dynamic tones. Regardless of perceived or acoustic duration, contour tone syllables are generally longer than flat tone syllables. The difference between the effects of dynamic and flat $f_0$ patterns on perceived and acoustic duration suggests that the mechanism of tonal perception is not monolithic. The perceptual mechanism for dynamic $f_0$ appears to be different from that of level $f_0$. This conclusion echoes the results of previous investigations on the perceptual dimensions of tone (e.g., Gandour, 1981) which have repeatedly found $f_0$ slope and height to be distinct attributes. Thus, rather than speaking of tone-duration interaction generically where all tones are lumped together as a homogeneous group, the relation between duration and $f_0$ slope on the one hand, and $f_0$ height on the other may need to be examined differently.

An important corollary of this conclusion is that the effects of $f_0$ on the emergence of sound patterns and on linguistic universals might differ depending on the slope and height of the $f_0$ contour. For example, while patterns showing interaction between dynamic-tone related lengthening is observed in many languages (Gordon, 1999, 2001; Zhang, 2001), patterns showing interaction between level tone and length are scant. Moreton (2006) calls this phenomenon of differential phonological developments “underphonologization”. That is, given two sets of phonetic precursors that might $a$ priori give rise to sound changes, one phonetic precursor gives rise to a new phonological pattern more readily than the other. Why should this be? Moreton (2006) suggests that there are two possible factors, which he calls pattern selectivity and phonetic precursor robustness. Pattern selectivity refers to cognitive biases that make some patterns difficult or impossible to acquire, even from perfect training data. On the other hand, phonetic precursor robustness refers to the fact that some phonetic precursors to sound change are less subtle or more frequent than others. Based on a comparison between phonological processes involving tone-tone interactions and consonant-tone interactions, Moreton concludes that pattern selectivity is the likely cause of underphonologization. From the Optimality Theoretic point of view of a language learner, the constraints responsible for triggering alternations may interact with other constraints in different ways. The learning algorithm is biased toward phonologizing processes driven by markedness constraints that interact with few other constraints. Leaving aside the specifics of his method of accessing relative robustness of phonetic precursors, it is instructive to note that Moreton only takes into account the relative magnitude of the articulatory influence involved in tone-tone interactions and consonant-tone interactions. As highlighted in the present study, however, the robustness of a phonetic precursor must be measured in both articulatory and perceptual terms. That is, if the phonetic precursor is observed in the articulatory domain, then the effect of perceptual normalization must be taken into account. Similarly, if a phonetic precursor has its origin in speech perception, then potential articulatory compensatory measures must be considered as well. The nature of the normalization effect may ultimately determine how likely a phonetic precursor to a sound pattern will prevail. That is, if normalization, be it perceptual or articulatory, is robust, then the chances of a phonetic precursor becoming phonologized should be considerably undermined.
References


