

Tonal effects on perceived vowel duration

Alan C. L. Yu

Cross-linguistically, vowel duration is often found 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 is commonly reflected in the world's languages. This paper reports a study showing that 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. Duration compensation is characteristic of level tone perception, but not of dynamic tone realization. Regardless of perceived or acoustic duration, contour tone syllables are generally longer than flat tone syllables. This type of differential normalization effect might explain why certain phonetically motivated sound patterns are more prevalent than others typologically.

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 is 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; Gordon 2001; Zhang 2001; Yu 2003; Yu 2004). 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; Howie 1974; Ho 1976; Howie 1976; Nordenhake and 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 nonfalling 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) (Gandour 1977). Kwon (2003) argues that many long vowels in Korean today correspond to Middle Korean rising tone syllables. Middle Korean had 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, 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.

(2)	Weert	vowel dur. (ms.)	Baexem	vowel dur. (ms.)	gloss
	knɪ:n	220	knɪ:n ^{II}	186	‘rabbit’
	knɪn	143	knɪ:n ^I	177	‘rabbits’
	stɛɪn	273	ʃtɛɪn ^{II}	254	‘stone’
	stæjn	192	ʃtɛɪn ^I	219	‘stones’

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 and Ewan 1973; Sundberg 1973; Sundberg 1979; Xu and Sun 2002). However, such a physiological account offers no explanation for the durational differences among 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 opposed 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 et al. 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, crosslinguistically, 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 Gruenenfelder and Pisoni 1980; Kohler 1985; Lehiste 1985; Dommelen 1989). The usual explanation for this effect refers to the findings of Lehiste and postulates longer perceived vowel duration in connection with f_0 movement. That is, the longer perceived duration appears to bias the phoneme-judgment towards more lenis percepts. While the f_0 effect on perceived vowel duration is invoked in explaining per-

ception 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 study of the interplay between articulatory and perceptual effects in speech has gained a renewed sense of urgency in recent year in light of recent efforts by some researchers to identify intrinsic grammatical factors in the origins of sound patterns. That is, while the fact that articulatory and perceptual factors may influence the organization of a phonological system is undisputed, the nature and the extent of the influence has been a matter of intense debate in recent years. Some researchers, most notably Ohala (1993), argue that articulatory and perceptual factors shape phonological systems through listener misperception-induced sound changes and the synchronic typology of sound pattern is a consequence of the phonologization of such phonetic precursors (Blevins and Garrett 1998; Kavitskaya 2001; Barnes 2002; Blevins 2004; Yu 2004; Blevins and Garrett 2005; Barnes 2006). Others argue that phonological typology is shaped by cognitive and substantive biases that cause learners to acquire certain patterns more readily than others (Steriade 2001; Wilson 2003; Moreton 2006; Wilson 2006). The interaction between tone and duration in the world's languages is a fertile testing ground for this debate in view of the skewed typology of tone-duration interaction. That is, while patterns of dynamic-tone related lengthening are observed in many languages (Gordon 1999; Gordon 2001; Zhang 2001), patterns showing interaction between level tone and length are scant. To what extent then is this asymmetry accountable in articulatory and perceptual terms alone or must this asymmetry be attributed to some sort of cognitive bias intrinsic to the learning mechanism of language?

The following experiment was set up to explore the answer to the first question. In particular, we want to identify what potential perceptual motivations underlie the abovementioned asymmetry. While previous studies have found that *dynamic* movement in fundamental frequency affects English listeners' perceived duration, no studies have considered the effect of f_0 height on perceived duration.¹ 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 in connection 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 implica-

tion with respect to the main concerns of this study appear in Section 4. The conclusion appears in Section 5.

2. Method

Subjects: Sixteen native speakers of English, all students at the University of Chicago, were paid a nominal fee to participate in the experiment. Subjects were screened for past experience with tonal languages in order to eliminate those who had been exposed to tonal contrast in a linguistic context. 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. Following the usual conventions for transcribing tone in East Asian languages, the different f_0 contours are represented henceforth by two numbers that indicate the pitch of the beginning and the end of the tone; the numbers range from 1–5, with 1 being the lowest level and 5 the highest. Thus, “T33” represents a mid-level tone.

Table 1. Values of the five f_0 contours employed in the present study.

Name	f_0 beginning (Hz)	f_0 ending (Hz)	average f_0 (Hz)
T11	85	85	85
T33	115	115	115
T55	145	145	145
T15	85	145	115
T51	145	85	115

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 T33 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 the EPrime testing program. Stimuli were presented binaurally with high quality headphones in a quiet room. The sound was 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 zscore scale. Figure 1 summarizes the z-score transformed duration rating results across tones and stimulus durations. To determine the overall effect of tone and stimulus duration 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. There are significant main effects of stimulus duration ($F(2, 30) = 424.6, p < 0.01$) and tone ($F(4, 60) = 17.5, p < 0.01$). No significant interaction effect is observed, however.

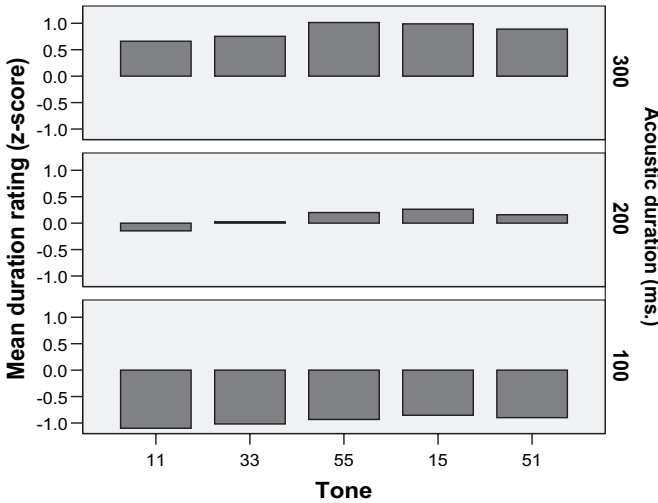


Figure 1. Mean z-score transformed duration rating across five tones by English speakers.

The fact that stimulus duration has a significant effect on perceived duration is to be expected. A series of pairwise comparisons, using Bonferroni adjusted alpha levels of 0.005 per test (0.05/10), was conducted to further examine the effect of tone on perceived duration ratings. A summary of results of the pairwise comparison is shown in Table 2. Between level tone syllables, T55 syllables were

Table 2. Results of pairwise comparisons between syllables with different tone contours

Comparison		<i>t</i> (15)	<i>p</i>
Between level tones	T11 ~ T33	-2.628	0.019
	T11 ~ T55	-5.665	0.000
	T33 ~ T55	-8.082	0.000
Level vs. dynamic tones	T11 ~ T51	-6.174	0.000
	T11 ~ T15	-6.848	0.000
	T33 ~ T51	-2.444	0.027
	T33 ~ T15	4.314	0.001
	T55 ~ T15	0.760	0.459
	T55 ~ T51	-0.880	0.393
	T15 ~ T51	1.960	0.069

perceived as significantly longer than syllables with T33 or T11. The difference between T11 and T33 did not reach the adjusted level of significance. Not all dynamic tone syllables were perceived as significantly longer than level tone syllables. Dynamic tone syllables, which were generally perceived as longer than T11 syllables, were comparable in perceived duration as T55 syllables. T33 syllables were found to be significantly longer than T15 syllables, but not so compared to T51 syllables.

4. Discussion

The experimental results suggest that the height and slope of fundamental frequency can significantly affect listeners' 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 effect 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 duration as intentional. The hypocorrection of the tonal effect on perceived duration might have resulted in the creation of distinctive vowel length in Korean today. That is, the Middle Korean listener might have failed to factor out the perceptual lengthening effect rising tone has on duration perception and mistakenly parsed the lengthened syllable at face value. As such, rising tone syllables, which were underlyingly short previously, are now taken to be underlyingly long.

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 high flat f_0 are perceived as significantly longer than those with a mid flat f_0 pattern. Syllables with low flat 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 perceived 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 dissociation between the effect of f_0 height on perceived duration and on acoustic duration demands explanation.

4.1. Differential effects of f_0 height on perceived and acoustic duration

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 an 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 in connection with pitch height is observed at pitch levels far beyond the range of linguistic f_0 modulations (e.g., 500Hz. & 4000Hz.). His results are less amenable 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.

Furthermore, preliminary results from a pilot study looking at the effect of f_0 slope and height on perceived duration in speakers of a tone language found similar positive correlation between f_0 height and perceived duration. For example, Figure 2, which summarizes the duration rating results of seven Cantonese speakers (the experimental conditions are identical to the experiment reported in this paper), shows a positive correlation between f_0 height and perceived

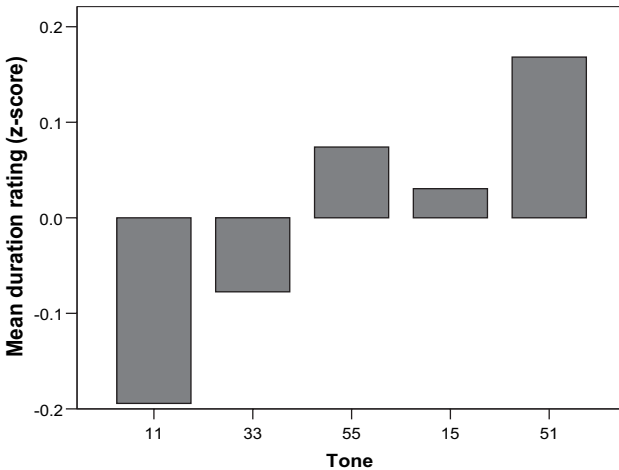


Figure 2. Mean z-score transformed duration ratings across five tones by Cantonese speakers.

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-seated motivations that go beyond *ad hoc* explanations that are language-specific.

Why should f_0 height correlate inversely with perceived duration? One theory was advanced in Brigner (1988). He 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 (Helson and King 1931; Abe 1935; Brigner 1984). In particular, a decrease in perceived time may occur with an 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 *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. Admittedly, this explanation of the effect of pitch on perceived duration is speculative. More investigations must be conducted to examine the viability of such an approach.

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 dissociation 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 of different heights. 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 perceived duration by asking the subject to judge the duration equivalence of a high tone (4000 Hz) and a low tone (500Hz) 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 adjusted the duration of the high-frequency tone in steps of approximately 27 ms, ascending from 110 ms or descending from 659 ms. 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 were given a range of duration options, they invariably preferred 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 was judged heavier than a bigger box of equal physical weight. Hombert (1978) found that a low vowel [a] had 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 suggested 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 and 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 (2007). He found that higher vowels sounded longer than lower vowels when acoustic durations were equal. He reasoned 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 themselves, 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.

4.2. Differential phonologization as differential normalization

An important corollary of the above 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. Recall that patterns of dynamic-tone related lengthening are observed in many languages, while patterns showing interaction between level tone and length are few. Given that both dynamic and level tones exert an influence on duration at both the articulatory and perceptual levels, why is one type of tone-duration interaction more prevalent than the other? This type of *differential phonologization* effect (i.e. when one phonetic precursor gives rise to a new phonological pattern more readily than another, even though both sets of phonetic precursors may give rise to sound changes) is not at all uncommon. At face value, differential phonologization seems incompatible with a phonologization approach to phonological typology since, all else being equal, any systematic variation in speech production and perception may become phonologized. Additional qualification is obviously needed.

Moreton (2006) identifies two potential factors that could give rise to what he calls underphonologization (i.e., the under-representation of sound patterns induced by a particular phonetic precursor that might *a priori* give rise to phonologizable sound changes): *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. Moreton concludes that pattern selectivity is the likely cause of underphonologization, since the magnitudes of phonetic precursor robustness between the pairs of phonetic precursors he considered do not seem to differ significantly.

This conclusion, however, is premature, in light of the inadequacies of his method for assessing the relative robustness of phonetic precursors. Moreton defines phonetic precursor robustness in terms of a ratio between the target measurements in what he refers to as a “Raising” context divided by that in the “Lowering” context. Thus, for example, in order to determine whether the phonetic influence of tone on neighboring f_0 is substantially greater than that of voicing, Moreton searches through the literature of previous published production studies where vowel f_0 was measured (and reported in physical units) in

the context of neighboring tones and in the context of voicing or aspiration of the preceding consonant. Thus, in the case of tone-tone interaction, a “Raising” context is the context which is likeliest to raise the target f_0 (e.g., [35β_] is the Raising context in a study measuring a [33] target in the contexts of [44_], [35_] and [21_]). The “Lowering” context is determined through a similar procedure. Relative phonetic precursor robustness is, therefore, determined solely in terms of speech production, according to Moreton. 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. If normalization, be it perceptual or articulatory, is robust, then the chances of a phonetic precursor becoming phonologized should be considerably undermined.

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 dissociation effect is observed in syllables with dynamic tones (i.e. 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. The experiment reported here focused on the distinction between dynamic and flat (i.e. static) f_0 . Yet, it must be noted that, in natural languages, static tones are understood to be level only in contrast with tones with shaped pitch contours. That is, dynamicism is a relative matter. In practice, level tones often display shallow upward or downward movement. It remains a real question as to what extent this type more subtle f_0 movement affects duration perception and the implementation of the compensatory lengthening effect. Further research is obviously needed in this respect.

The present study also provides an illustrative example demonstrating the importance of considering both perceptual and articulatory factors in differential phonologization research. In particular, underphonologization may be rooted in differential normalization in speech. That is, a phonetic precursor is less likely to become phonologized into sound patterns when a corresponding articulatory or perceptual normalization effect obscures the robustness of the phonetic precursor. As such, appeals to cognitive bias-induced pattern selectivity should be employed only when phonetic precursor robustness genuinely offers no explanatory insights.

Notes

1. While Brigner'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 are based on pure tones and that the frequencies 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).

References

- Abe, Saburo
1935 Experimental study on the co-relation between time and space. *Tohoku Psychologia Folia*, 3, 53–68.
- Barnes, Jonathan Allen
2002 Positional neutralization: A phonologization approach to typological patterns. PhD Dissertation. Department of Linguistics. Berkeley, University of California, Berkeley.

- Barnes, Jonathan Allen
2006 *Strength and weakness at the interface*. Berlin and New York: Mouton de Gruyter.
- Birnbaum, Michael H. and Clairice T. Veit
1974 Scale-free tests of an averaging model for the size-weight illusion. *Perception & Psychophysics*, 16, 276–282.
- Blevins, Juliette
2004 *Evolutionary Phonology: The emergence of sound patterns*. Cambridge, UK: Cambridge University Press.
- Blevins, Juliette and Andrew Garrett
1998 The origins of consonant-vowel metathesis. *Language*, 74(3), 508–556.
- Blevins, Juliette and Andrew Garrett
2005 The evolution of metathesis. Edited by Bruce Hayes, Robert Kirchner and Donca Steriade. *The phonetic basis of phonology*. Cambridge, Cambridge University Press.
- Brigner, Willard
1984 Rotation of space-time plane predicts a new illusion of spatial displacement. *Perceptual and Motor Skills*, 59, 359–369.
- Brigner, Willard
1988 Perceived duration as a function of pitch. *Perceptual and Motor Skills*, 67, 301–302.
- Brotzman, Robert L.
1964 Progress report on Mandarin tone study. *Project on linguistic analysis report*, 8, 1–35.
- DeLong, Alton J.
1981 Phenomenological space-time: toward an experiential relativity. *Science*, 213, 681–683.
- Dommelen, Wim A. van
1989 The influence of pre- and postplosive fundamental frequency on /t-/d/ perception in German. *Phonetica*, 46, 217–230.
- Gandour, Jack
1977 On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica*, 34, 54–65.
- Gandour, Jack
1981 Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics*, 9, 20–36.
- Gordon, Matthews
1999 Syllable weight: phonetics, phonology, and typology. *Linguistics*. Los Angeles, University of California, Los Angeles.
- Gordon, Matthews
2001 A typology of contour tone restriction. *Studies in Language*, 25(3), 423–462.

- Gruenenfelder, Thomas M. and David B. Pisoni
1980 Fundamental frequency as a cue to postvocalic consonantal voicing: Some data from speech perception and production. *Perception & Psychophysics*, 28, 514–520.
- Gussenhoven, Carlos
2007 A vowel height split explained: Compensatory listening and speaker control. Edited by Jennifer Cole and José Ignacio Hualde. *Laboratory Phonology 9*. Berlin/New York, Mouton de Gruyter. 145–172.
- Heijmans, Linda
2003 The Relationship between Tone and Vowel Length in Two Neighboring Dutch Limburgian Dialects. Edited by Paula Fikkert and Jacobs Haike. *Development in prosodic systems*. Berlin: Mouton de Gruyter. 7–45.
- Helson, Harry and Samuel M. King
1931 The *tau*-effect, an example of psychological relativity. *Journal of Experimental Psychology*, 14, 202–218.
- Ho, Aichen T.
1976 The acoustic variation of Mandarin tones. *Phonetica*, 33, 353–367.
- Hombert, Jean-Marie
1978 Consonant types, vowel quality, and tone. Edited by Victoria A. Fromkin. *Tone: a linguistic survey*. New York: Academic Press. 77–112.
- Howie, John M.
1974 On the domain of tone in Mandarin: Some acoustical evidence. *Phonetica*, 30, 129–148.
- Howie, John M.
1976 *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Kavitskaya, Darya
2001 Compensatory lengthening: Phonetics, phonology, diachrony. *Linguistics*. Berkeley, University of California, Berkeley.
- Kohler, Klaus J.
1985 F0 in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America*, 78, 21–32.
- Kong, Qingming
1987 Influence of tones upon vowel duration in Cantonese. *Language and Speech*, 30(4), 387–399.
- Kwon, Kyung-Keun
2003 Prosodic Change from Tone to Vowel Length in Korean. Edited by Paula Fikkert and Jacobs Haike. *Development in prosodic systems*. Berlin: Mouton de Gruyter. 67–89.
- Lee, Gina Maureen
1993 Comparative, diachronic and experimental perspectives on the interaction between tone and vowel in Standard Chinese. *Department of Linguistics*. Columbus, The Ohio State University.

- Lehiste, Ilse
1976 Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 4, 113–117.
- Lehiste, Ilse
1985 An Estonian word game and the phonematic status of long vowels. *Linguistic Inquiry*, 16(3), 490–492.
- Lindblom, Bjorn
1990 Explaining phonetic variation: a sketch of the H and H theory. Edited by William Hardcastle and A. Marchal. *Speech production and speech modeling*. Dordrecht: Kluwer Academic Publishers. 403–439.
- Moreton, Elliott
2006 Underphonologization and modularity bias. *University of North Carolina, Chapel Hill*.
- Nordenhake, Magnus and Jan-Olof Svantesson
1983 Duration of standard Chinese word tones in different sentence environments. *Lund University Working papers in Linguistics*, 25, 105–111.
- Ohala, John J.
1993 The phonetics of sound change. Edited by Charles Jones. *Historical Linguistics: problems and perspectives*. London: Longman. 237–278.
- Ohala, John J. and William G. Ewan
1973 Speech of pitch change. *Journal of Acoustical Society of America*, 53, 345.
- Peterson, Gordon E. and Ilse Lehiste
1960 Duration of syllable nuclei in English. *Journal of Acoustical Society of America*, 32, 693–703.
- Pisoni, David B.
1976 Fundamental frequency and perceived vowel duration. *Journal of Acoustical Society of America*, 59, S39.
- Steriade, Donca
2001 Directional asymmetries in place assimilation. Edited by Elizabeth Hume and Keith Johnson. *The role of speech perception in phonology*. San Diego: Academic Press. 219–250.
- Stevens, Stanley S.
1934 The volume and intensity of tones. *American Journal of Psychology*, 46, 397–408.
- Sundberg, Johan
1973 Data on maximum speech of pitch changes. *Royal Institute of Technology, Stockholm. Speech Transmission Laboratory. Quarterly Progress and Status Report*, 4, 39–47.
- Sundberg, Johan
1979 Maximum speech of pitch changes in singers and untrained subjects. *Journal of Phonetics*, 7, 71–79.

- Svantesson, Jan-Olaf
 1988 Hu – a language with unorthodox tonogenesis. Edited by Jeremy Davidson. *Contributions to Mon-Khmer Studies: Essays in honour of Professor H. L. Shorto*. London: SOAS.
- Wang, William, Ilse Lehiste, Chin-Kueng Chuang and Nancy Darnovsky
 1976 Perception of vowel duration. *Journal of the Acoustical Society of America*, 60, S92.
- Whalen, Doug H. and Andrea G. Levitt
 1995 The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23, 349–366.
- Wilson, Colin
 2003 Experimental investigation of phonological naturalness. Edited by G. Garding and M. Tsujimura. *Proceedings of the 22nd West Coast Conference on Formal Linguistics*. Somerville, MA.: Cascadilla. 533–546.
- Wilson, Colin
 2006 Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, 30(5), 945–982.
- Xu, Yi and Xuejing Sun
 2002 Maximum speed of pitch change and how it may relate to speech. *Journal of Acoustical Society of America*, 111, 149–174.
- Yu, Alan C. L.
 2003 Contour tone induced lengthening in Cantonese. Edited by M. J. Sole, D. Recasens and J. Romero. *The 15th International Congress of Phonetic Sciences*. Barcelona, Spain. 2381–2384.
- Yu, Alan C. L.
 2004 Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language*, 80(1), 73–97.
- Yu, Alan C. L.
 2004 Some methodological issues in phonetic typology research: Cantonese contour tone revisited. Edited by Pawel Nowak and Corey Yoquelet. *Proceedings of the 29th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, Berkeley Linguistics Society.
- Zhang, Jie
 2001 The effects of duration and sonority on contour tone distribution-typological survey and formal analysis. PhD dissertation. Department of Linguistics. University of California, Los Angeles.