

## Variability in perceived duration: pitch dynamics and vowel quality

Alan C. L. Yu, Hyunjung Lee\*, Jackson Lee

Phonology Laboratory, University of Chicago, U.S.A.

aclyu@uchicago.edu, hyunjunglee123@gmail.com, jslllee@uchicago.edu

### Abstract

Why are some sound patterns more commonly attested than others? This paper explores the hypothesis that DIFFERENTIAL PHONOLOGIZATION—when one phonetic precursor gives rise to a new phonological pattern more readily than another even when both sets of phonetic precursors may give rise to sound patterns—might arise as a result of variability in how the perceptual system copes with variation in the speech signal; variations that are more robustly normalized for are less likely to lead to the phonologization. To investigate this hypothesis, we examine listeners' perceptual responses to covariation between duration, on the one hand, and conditioning factors, such as the dynamics of the fundamental frequency and vowel height on the other. We found that, while perceived duration of syllables is modulated by the abovementioned conditioning factors, the direction and magnitude of perceptual adjustments differ across conditioning factors, thus supporting the differential compensation approach to differential phonologization.

**Index Terms:** Pitch, vowel quality, perceived duration, perceptual compensation

### 1. Introduction

It has long been recognized that some sound changes, and by extension sound patterns, are more common than others. As common cross-language sound patterns are found to have physical and perceptual phonetic origins, the rarity of some sound patterns has been regarded as a result of the low probabilities of the corresponding phonetic effects being phonologized through sound change [1, 2]. For example, while patterns of tone affecting vowel height have been reported, the reverse is rarely, if ever, found. This infrequency of vowel-height effect on tone was thought to be a consequence of perceptual compensation; a low vowel [a] had a tendency to be judged higher in pitch than the high vowels [i] or [u] even when their fundamental frequency (f<sub>0</sub>) were in fact equal [3]. DIFFERENTIAL PHONOLOGIZATION—when one phonetic precursor gives rise to a new phonological pattern more readily than another even when both sets of phonetic precursors may give rise to sound changes—might be explained, if only partially, by differential compensation. That is, if not all variations in the speech signal are modulated by the perceptual and articulatory system to the same degree cross-linguistically and language-internally, then it is not surprising that variations that are more robustly normalized for are less likely to lead to the phonologization of listener-oriented sound changes. To investigate this hypothesis, we examine listeners' perceptual responses to co-variation between duration, on the one hand, and the dynamics of the f<sub>0</sub> and vowel height, on the other.

Segmental duration in speech is heavily context-dependent.

\*Corresponding author

Sonorous segments tend to be longer when hosting dynamic tones than level ones [4, 5, 6]. Diachronically, tone systems evolve into non-tonal languages with length opposition [7, 8]. Likewise, vowels with low tones tend to be longer than those with high tones [9]. Yet, while patterns of dynamic tone-related lengthening are observed in many languages [6], patterns showing interaction between level tone and duration are few [10]. Earlier research shows that perceived duration responses vary depending on the dynamics of f<sub>0</sub> [11, 12], recent studies, however, found that f<sub>0</sub> slope and height exert different influence on perceived duration [13, 14]. Across different level pitches, listeners experienced high f<sub>0</sub> syllables as longer than low f<sub>0</sub> syllables with equal acoustic duration, the reverse of production results; across level pitches and dynamic contours, dynamic contours elicit longer duration percept than level ones, in congruence with production results. The variable patterns in the effects of f<sub>0</sub> dynamics on duration production and perception have been explained as a difference in production compensation for psycho-acoustic perceptual effects [13] or as a matter of compensatory listening [14].

Beyond the effects of f<sub>0</sub> on the production and perception of segmental duration, vowel quality also plays a significant role. Higher vowels are shorter than lower vowels, presumably because the greater distance between the roof of the mouth and the articulatory excursion of the tongue and the jaw the longer the vowel. This tendency is sometimes phonologized as sound patterns. Original long /i, y, u/ in Dutch, for example, have phonologized as short /i, y, u/ and have merged in their quantity with /ɪ, ʏ/ [15]. However, phonological patterns of this sort are relatively infrequent.

How f<sub>0</sub> and vocalic cues affect perceived duration appears to differ cross-linguistically, and such differences are dependent and contrast-specific. A recent study [16] found strong effects of spectral cues on perceived vowel duration across Thai, Japanese, German, and Spanish, but, with the exception of Japanese, which has a restricted distribution falling f<sub>0</sub> on long vowels, only a weak or opposite effect of a falling f<sub>0</sub> contour is found in Thai, German, and Spanish. Unclear from this study is whether the spectral and f<sub>0</sub> effects on perceived vowel duration are related. That is, do spectral and f<sub>0</sub> effects on perceived duration covary in tandem? If so, it suggests that a unified perceptual mechanism is at work in regulating listener's perceptual compensation strategies.

Also unclear is to what extent spectral and f<sub>0</sub> interact in their influence on perceived duration. It has long been observed that higher vowels tend to be associated with higher f<sub>0</sub> while lower vowels with lower f<sub>0</sub> [17, 18]. If f<sub>0</sub> height exerted an influence on perceived duration, we might expect the vowel height effect to mitigate the f<sub>0</sub> influence. For example, if listeners perceptually compensate for the higher f<sub>0</sub> during the production of high vowels relative to low vowels, for high and low vowels that are of equal duration, the high vowels might be perceived as

longer than low vowels. However, given that syllables with high f0 tend to be shorter than lower f0 ones, if listeners compensate such a covariation, we might expect vowels with higher f0 to be perceived as longer than vowels with lower f0 as well. To the extent that the two types of compensatory responses interact, we should expect a high vowel to be perceived as the longest when f0 is high, compared to an acoustically equally long low vowel with a low f0. Crucially, when the compensatory effects conflict, we might see a mitigation effect between the two phonetic effects. The experiment below is designed to investigate the potential interaction between the compensatory responses for the dynamic and height of f0 and on vowel quality using a visual analog scaling paradigm. Are perceptual adjustments additive or are there ceiling and floor effects associated with these compensatory strategies?

## 2. Methods

### 2.1. Participants

200 participants completed a visual analog scale task. 26 participants (18 females, mean age of 26 (SD = 12.1)) completed the lab-based task while and 174 participants (99 females, mean age of 33 (SD = 9.53)) completed a similar task on Amazon Mechanical Turk, either for course credits (lab-based participants only) or a nominal fee. The participants were all native speakers of American English. None reported any speech or hearing problems.

### 2.2. Stimuli

Two 300 ms syllables, [pa] and [pi], synthesized using SynthWorks, were used to create two 5-step duration continua with 25ms decreasing increments: 250ms, 225ms, 200ms, 175ms, 150ms. The f0 of the syllables was then manipulated to create four f0 contours (tone henceforth) using the parameters given in Table 1.

Table 1: *F0 (Hz) values of the four tone types*

Tone	F0 onset	F0 offset	F0 mean
Low (L)	100	100	100
High (H)	200	200	200
Rising (R)	130	200	165
Falling (F)	200	130	165

### 2.3. Procedure

Participants completed a Visual Analog Scaling (VAS) task in which they judged the duration of a stimulus using a visual display that corresponds to contrastive sounds [19]. The lab-tested participants were given the visual scale shown in Figure 1 (top panel); the left and right sides of a horizontal line were labeled ‘100ms’ and ‘300ms’, respectively. The visual scale was displayed on the computer screen immediately after each stimulus was presented; the cursor on the scale was always reset to the midpoint of the horizontal line (i.e., 200ms). Participants were instructed to click anywhere on the visual scale, and were given 3 seconds to respond before the presentation of the next stimuli. A total of 320 trials (5 durations x 4 tones x 2 vowels x 8 repetitions) were randomly presented. The participants completed the task over E-Prime in a sound-proof booth. Since the visual scale provides more options to respond along the continuous scale, the responses might reflect listeners’ percept better in details than a simple identification task with fixed response options.

The online-based task was conducted on Amazon’s Mechanical Turk. To mimic the continuous scale, we created a

row of 31 radio buttons of which left and right endpoints were labeled ‘100ms’ and ‘300ms’, respectively (see lower panel of Fig. 1). Consistent with the lab-based test, participants were asked to estimate duration of each stimulus by clicking any buttons along the visual duration scale. After responding to each stimulus, the next stimulus was automatically presented to the listeners; there was no restriction on response time. A total of 160 trials (5 durations x 4 tones x 2 vowels x 4 repetitions) were randomly presented.

To familiarize the participants with duration estimation, before the main test block, participants were presented with mid-level tone [pe] syllables (i.e., onset/offset f0 of 150 Hz) of 100 ms and 300 ms without a response required. In the ensuing practice session, participants were asked to identify whether the duration of [pe] is 100 ms or 300 ms by pressing buttons, and feedback was provided to each response in the practice.

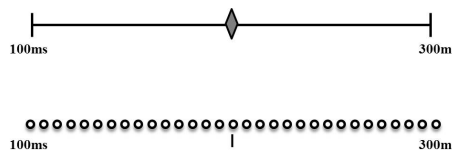


Figure 1: *Image of the visual duration scale used in the lab-based (top) and online (bottom) tasks.*

### 2.4. Analysis

Subjects’ duration estimation was recorded as VAS scores ranging from 0 (short) to 100 (long). The VAS scores were modeled using linear mixed-effects regression fitted in R, using the `lmer()` function from the `lme4` package. The model contains five factors: TRIAL indexed the stimulus’ presentation order, VOWEL quality (/a/ vs. /i/), TONE (High, Low, Rising, Falling), acoustic DURATION (150, 175, 200, 225, and 250ms) and COHORT (Lab- vs. Internet-based). To minimize collinearity, continuous variables were centered and binary variables were sum-coded. The TONE variable was coded as three contrasts: rising vs. falling tones (CONTRAST 1), high vs. low tones (CONTRAST 2), and level vs. contour tones (CONTRAST 3). According to forward selection, we first entered each of the fixed variables, and sequentially added the interaction terms. Model comparison was conducted with log-likelihood tests. The final model, given in a `lme4`-styled formula, was:  $VAS-SCORE \sim TRIAL + (TONE * VOWEL * DURATION)^2 + (1+TRIAL+TONE+ VOWEL+DURATION—SUBJECT)$ . The COHORT factor was dropped as likelihood ratio tests comparing between two-way interaction models with and without an interaction with COHORT did not reach significance ( $\chi^2(49) = 4.53, p = 0.604$ ).

## 3. Results

Table 2 summarizes the parameter estimate  $\beta$  for each of the fixed effects and interactions in the final model. All four main factors were significant. As expected, a main effect of DURATION indicates that longer stimuli have higher VAS scores. A main effect of TRIAL suggests that VAS scores became incrementally higher as the experiment progressed. An effect of VOWEL indicates that /i/ was rated as longer than /a/. In terms of the effects of tonal contour on perceived duration, as illustrated in the left panel of Fig. 2, stimuli with a rising contour are rated as longer than those with a falling contour. Syllables with a high tone are also rated longer than those with low tone. Con-

tour toned syllables are rated as longer than level toned syllables in general. A significant interaction between TONE and VOWEL suggests that the perceived duration difference in vowel height is attenuated in (i) the rising tone relative to the falling tone syllables and (ii) the low tone relative to the high tone syllables (see right panel of Fig. 2). The acoustic duration of the stimuli also interacts significantly with the tonal contour of the syllable. In particular, the difference in perceived duration between contour and level toned syllables is reduced in syllables with longer acoustic duration (bottom left panel of Fig. 3). Such an attenuation, however, occurs in the shorter duration stimuli comparing between low and high toned syllables (top right panel) and between rising and falling syllables (bottom right panel). Finally, there is a small, but significant, interaction between VOWEL and DURATION, suggesting that the vowel difference in perceived duration is attenuated when the acoustic duration of the stimulus is long (top left panel of Fig. 3).

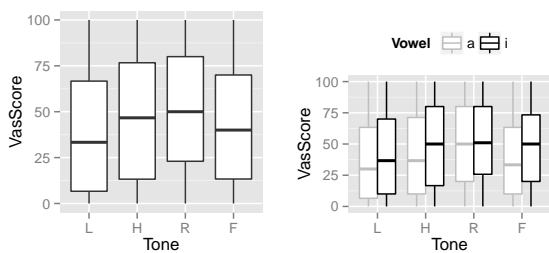


Figure 2: Perceived duration as mediated by TONE and VOWEL.

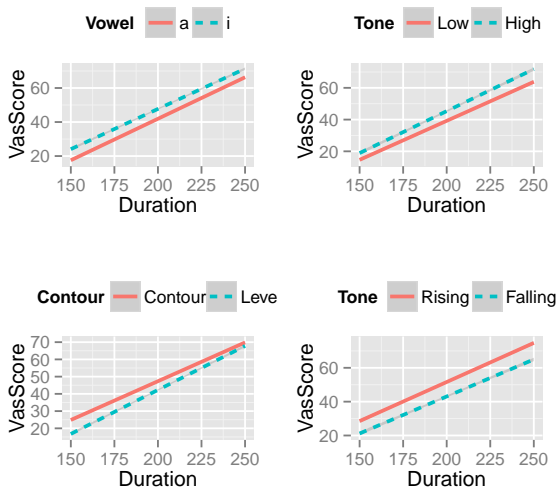


Figure 3: Perceived duration as mediated by DURATION, VOWEL, and TONE.

### 3.1. Discussion

The results of our study are generally consistent with earlier reports [13], revealing that i) contour-toned syllables are perceived as longer than level-toned ones, ii) rising-toned syllables are perceived as longer than falling-toned syllables, iii) high-toned syllables are perceived as longer than low-toned ones, and iv) syllables with the high vowel /i/ are generally perceived as longer than those with /a/. A novel finding of this study concerns potential ceiling and floor effects in the tonal and vocalic

Table 2: Estimates for predictors in a mixed-effects model in the duration perception task.  $t$ -value larger than 2 was considered significant.

Predictor	Coef. $\beta$	SE ( $\beta$ )	t value
Intercept	42.26	0.80	52.83
TRIAL	2.64	0.57	4.61
TONE-CONTRAST 1 (rising-falling)	11.13	0.70	16.00
TONE-CONTRAST 2 (high-low)	4.22	1.10	3.84
TONE-CONTRAST 3 (level-contour)	-4.58	0.75	-6.13
VOWEL-i	6.0	0.62	9.68
DURATION	17.20	0.50	34.15
TONE 1:VOWEL-i	-5.43	0.71	-7.64
TONE 2:VOWEL-i	3.18	0.71	4.48
TONE 3:VOWEL-i	-0.77	0.50	-1.52
TONE 1:DURATION	0.84	0.36	2.36
TONE 2:DURATION	1.35	0.36	3.80
TONE 3:DURATION	2.16	0.25	8.60
VOWEL-i:DURATION:	-0.53	0.25	-2.12

influence on perceived duration. Across tone height and across dynamic tones, their effects are attenuated in shorter acoustic duration. However, across vowel height and across level and contour tones, their effects are attenuated in longer duration. Concerning the interaction between tone and vowel quality, we found that vocalic compensation waned when the syllable carried a rising tone. Similarly, vocalic compensation attenuated when the tone was low level. These findings suggest that listeners exhibit response behaviors that are consistent with expectation adjustments in perception. However, the nature of the perceptual adjustments differ between feature types. With respect to the effects of level tone height and vowel height on perceived duration, listeners exhibit perceptual compensation. That is, for example, high vowels are perceived as longer than low vowels, even though the acoustic durations are in fact identical, presumably due to listeners' experience with high vowel being generally shorter in production than low vowels. Unlike tone and vowel height, however, the effects of rising and falling tones on perceived duration appear to mimic that of the production. That is, rising toned syllables are perceived as longer than falling-toned syllables and contour-toned syllables, as a class, are perceived as longer than level-toned syllables. This suggests that listeners do not compensate for covariation in the speech signal to the same extent. While the effects of tone height and vowel height are perceptually compensated for, the effects of contour tones on duration are not. This asymmetry in perceptual responses might help explain why phonological patterns and sound change involving duration/length on the one hand and tone height and vowel height on the other are few compared to associations between contour tone and length. Covariation in production that is compensated for (i.e. the tone and vowel height effects) is less likely to be phonologized into sound patterns.

What is unclear at this point is whether the perceptual adjustments observed here involves similar mechanisms. That is, are there different mechanisms involved in processing contour-toned syllables as opposed to syllables with other kinds of pitch contours and vowel levels? Also, do individuals vary by the extent of their perceptual adjustments? That is, is someone who perceptually adjusts for the vowel height effect on duration also likely to adjust for the contour tone effects? In the next section, we address this question, if only partially, by examining the correlation between the sizes of the perceptual adjustments across stimulus types.

## 4. Correlation across features

We examined the relationship between each individual's magnitude of compensation across three types of perceptual adjustments: tone height (high vs. low), tone dynamics (rising vs. falling), and vowel height (/i/ vs. /a/). The magnitude of compensation was measured for each subject by subtracting the average VAS score (i) for the falling toned syllables from those for the rising toned ones, (ii) for the low toned syllables from those for the high toned ones, and (iii) for the /a/ syllables from those for /i/ syllables. Figure 4 presents the correlations of individuals' perceived duration adjustments between level and contour tones (top:  $r(198) = 0.27$ ,  $p < 0.001$ ) and between the effects of vowels and contour tones (bottom:  $r(198)=0.14$ ,  $p = 0.04$ ). These results show that individuals who compensated greatly for one tonal property (e.g., high vs. low) also exhibited strong duration adjustments in the other tonal context (rising vs. falling). Likewise, individuals who show strong adjustments for the effects of contour tone on perceived duration also compensate for the effects of vowel. Two aspects of our findings are noteworthy. The strength of the correlations between perceptual adjustment types are rather weak, suggesting other factors might also be at work in modulating how the perceptual system processes different feature types. Also curious is the lack of a significant correlation between vowels and level tones ( $r(198) = 0.0667$ ,  $p = 0.35$ ). This suggests that the perceptual mechanisms handling the perceptual adjustments for vowel height might be different from those handling differences in tone height.

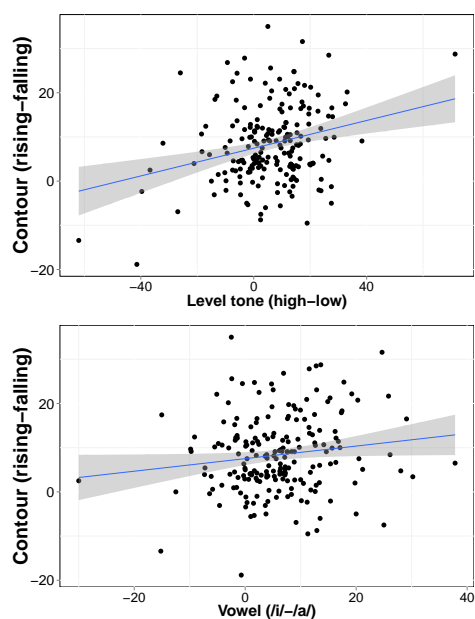


Figure 4: Correlations between Level and Contour tones (top) and between Vowel and Contour tone (bottom)

## 5. Conclusions

In this study, we found significant effects of tone contour and vowel quality on perceived duration. Correlation examination revealed significant individual variability in certain types of perceptual adjustments and that such variability appears to be consistent, albeit weakly, across feature types. Further research is needed to examine potential factors that explain such individual variability.

## 6. Acknowledgements

This work is partially supported by the National Science Foundation under grant #0949754.

## 7. References

- [1] J. Ohala, "The phonetics of sound change," in *Historical Linguistics: Problems and Perspectives*, C. Jones, Ed. London: Longman Academic, 1993, pp. 237–278.
- [2] J. Blevins, *Evolutionary Phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press, 2004.
- [3] J.-M. Hombert, "Consonant types, vowel quality, and tone," in *Tone: a linguistic survey*, V. Fromkin, Ed. New York: Academic Press, 1978, pp. 77–111.
- [4] M. Gordon, "A typology of contour tone restrictions," *Studies in Language*, vol. 25, pp. 405–444, 2001.
- [5] A. C. L. Yu, "Contour tone induced lengthening in cantonese," in *The 15th International Congress of Phonetic Sciences*, M. J. Sole, D. Recasens, and J. Romero, Eds., Barcelona, Spain, 2003, pp. 2381–2384.
- [6] J. Zhang, "The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis," Ph.D. dissertation, University of California, Los Angeles, Los Angeles, 2001.
- [7] L. Heijmans, "Relationship between tone and vowel length in two neighboring Dutch Limburgian dialects," in *Development in prosodic systems*, P. Fikkert and J. Haike, Eds. Berlin: Mouton de Gruyter, 2003, pp. 7–45.
- [8] K.-K. Kwon, "Prosodic change from tone to vowel length in Korean," in *Development in prosodic systems*, P. Fikkert and J. Haike, Eds. Berlin: Mouton de Gruyter, 2003, pp. 67–89.
- [9] J. Gandour, "On the interaction between tone and vowel length: Evidence from Thai dialects," *Phonetica*, vol. 34, pp. 54–65, 1977.
- [10] M. Faytak and A. C. L. Yu, "A typological study of the interaction between level tones and duration," in *Proceedings of the International Congress of the Phonetic Sciences XVII*, E. Zee, Ed. International Congress of the Phonetic Sciences, 2011.
- [11] I. Lehiste, "Influence of fundamental frequency pattern on the perception of duration," *Journal of Phonetics*, vol. 4, pp. 113–117, 1976.
- [12] W. Wang, I. Lehiste, C.-K. Chuang, and N. Darnovsky, "Perception of vowel duration," *Journal of the Acoustical Society of America*, vol. 60, p. S92, 1976.
- [13] A. C. L. Yu, "Tonal effects on perceived vowel duration," in *Laboratory Phonology 10*, C. Fougeron, B. Kuhnert, M. D'Imperio, and N. Vall'ee, Eds. Mouton de Gruyter, 2010, pp. 151–168.
- [14] C. Gussenhoven and W. Zhou, "Revisiting pitch slope and height effects on perceived duration," in *Interspeech 2013*. ISCA, 2013, pp. 1365–1369.
- [15] C. Gussenhoven, "Perceived vowel duration," in *On speech and language: Studies for Sieb G. Nootboom*, H. Quené and V. van Heuven, Eds. Utrecht: LOT, 2004, pp. 65–71.
- [16] H. Lehnert-Lehouillier, "From long to short and from short to long: Perceptual motivations for changes in vocalic length," in *Origins of Sound Change: Approaches to Phonologization*, A. C. L. YU, Ed. Oxford, UK: Oxford University Press, 2013, pp. 98–111.
- [17] B. Connell, "Tone languages and the universality of intrinsic f0: evidence from Africa," *Journal of Phonetics*, vol. 30, pp. 101–129, 2002.
- [18] D. H. Whalen and A. G. Levitt, "The universality of intrinsic f0 of vowels," *Journal of Phonetics*, vol. 23, pp. 349–366, 1995.
- [19] B. Munson, S. K. Schellinger, and K. U. Carlson, "Measuring speech-sound learning using visual analog scaling," *Perspectives on Language Learning and Education*, vol. 19, pp. 19–30, 2012.