# THE INFLUENCE OF DYNAMIC F0 ON THE PERCEPTION OF VOWEL DURATION: CROSS-LINGUISTIC EVIDENCE

*Heike Lehnert-LeHouillier*

University at Buffalo (SUNY) and Haskins Laboratories
hlehnert@buffalo.edu

## ABSTRACT

This paper investigates the influence of a dynamic fundamental frequency (F0) on the perception of vowel duration. The perception of vowel duration of the vowels [a], [e], and [i] with a falling versus a level F0 was investigated. Native speakers of Thai, Japanese, German, and Latin American Spanish were presented with monosyllabic CV non-sense words, and their perception of the duration of vowels with a level F0 was compared to that of vowels with a falling F0 from 160 Hz to 80Hz. The results show that only Japanese listeners perceived the vowels with a falling F0 as longer. Hence the cross-linguistic investigation shows that the influence of F0 on the perceived duration of vowels is language specific rather than universally present in speech perception.

**Keywords:** vowel length perception, Thai, Japanese, German, Spanish.

## 1. INTRODUCTION

The current study takes up an issue – the increase in perceived duration of a vowel due to a dynamic F0, which has been addressed in a variety of studies, with somewhat contradictory results. The cross-linguistic investigation presented here adds more data, and more insight into our understanding of this issue.

Early investigations, such as Lehiste [1], Pisoni [2], and Wang et al. [3] found that listeners perceive vowels with a dynamic F0 as longer than vowels with a level F0. All these studies used synthetic stimuli consisting of a single vowel ([1] and [2]) or isolated vowels and non-speech [3]. While Lehiste's [1] stimuli with a dynamic F0 had either a rising-falling or a falling-rising F0 contour, Pisoni [2] and Wang et al.'s [3] stimuli compared a falling and a rising F0 to a level F0. Wang et al. found that vowels with a rising F0 contour are perceived as longer than those with a falling F0, these, in turn, were perceived as longer than the vowels with a level fundamental frequency. These results were recently replicated by Yu [4] with the additional finding that vowels with a low level tone were perceived as shorter than those with a high level tone.

In a later study, Rosen [5] investigated the same issue and concluded that the falling and rising F0 contours were irrelevant in the perception of long and short vowels for his Swedish subjects. More evidence against the universal nature of the perceptual lengthening effect of a dynamic F0 was found in a study by van Dommelen [6]. Using monosyllabic and disyllabic words, which were presented either in isolation or embedded in a sentence, he found that German listeners only perceived vowels with a dynamic F0 as longer when they occurred in isolated monosyllabic words. In all other conditions the perceptual lengthening effect was reversed.

The reason for these somewhat contradictory results – in addition to the reasons discussed in [6] – may be due to two factors. One is the language background of the participants in these studies, and the other is the nature of the stimuli used. While the studies that found an increase in perceived duration due to a dynamic F0 were all conducted with native speakers of American English – a language without a phonemic vowel length contrast, the other studies were conducted with native speakers of Swedish and German – two languages with phonemic vowel length contrast. The stimuli used in the studies supporting the perceptual lengthening effect of a dynamic F0 consisted of isolated vowels, whereas the other studies used actual lexical items as stimuli. Furthermore, the vowels in these lexical items were followed by a consonant, which may impact vowel length as well as F0 (see [7]).

The current study addresses the role of language background with respect to the influence of a dynamic F0 on vowel length perception. The stimuli used for this experiment consisted of CV/CV: non-sense syllables and – unlike the studies in [5] and [6] – were not lexical items.

However, the stimuli - unlike the ones in [1], [2], and [3] - were not synthetically created but consisted of manipulated natural speech.
In order to investigate whether either a phonemic vowel length contrast, or the contrastive use of F0 (tone or pitch accent) in the native language of listeners influences perceived vowel duration due to a dynamic F0, native speakers of four different languages, Thai, Japanese, German, and Spanish were presented with the same stets of stimuli. Three of these languages, Thai, Japanese, and German exhibit a phonemic vowel length contrast, while Spanish does not. Thai and Japanese use F0 contrastively with Thai being a tone language and Japanese a pitch accent language.

## 2. METHOD

### 2.1. Participants

The participants in this perception experiment were 12 native speakers of Thai, 12 native speakers of Japanese, 12 native speakers of German, and 12 native speakers of Latin American Spanish. Out of the 12 native speakers of Thai, 9 were female and 3 were male. 6 of the Japanese participants were female and 6 were male. Out of the 12 German speakers, 6 were female and 6 male. The native speakers of Latin American Spanish were 6 female and 6 male participants. 6 of the Spanish speakers were native speakers of Mexican Spanish, 3 of Argentinean Spanish, 2 of Peruvian Spanish, and one speaker was a native speaker of Venezuelan Spanish.

All participants were students, staff or alumni at the University at Buffalo, and all reported to have normal hearing. The time the participants had been living in the US at the time of the experiment ranged from one year to 25 years. However, only 2 participants had been living in the US for more than 15 years, and they reported that they either spend several months a year in a native language environment or speak the native language at home on a daily basis. Each participant was paid for participating.

### 2.2. Stimuli

All stimuli for this experiment were based on the speech of a 22 year-old female Estonian talker. Estonian was chosen in order to avoid a native language bias for the listeners of any of the investigated languages. Unlike any of these languages, Estonian has a three-way length contrast in vowels. However, only the short and the long vowels, not the overlong vowels were used for the creation of the stimuli, since overlong vowels in Estonian seem to differ from long vowels mainly in the F0 pattern.

Overall, six vowel continua going from a short vowel to a long vowel were created. Three continua one for each of the vowels [i], [e], and [a] for each of the two experimental conditions: level F0 at 180 Hz, and falling F0 from 160Hz to 80 Hz. The vowels in the nonsense CV syllables [ti], [te], and [ta] were lengthened in seven steps to [ti:], [te:], and [ta:] respectively using PSOLA, see Moulines & Charpentier [8]. The length of the original vowels in the CV syllables as produced by the Estonian talker as well as the step size by which the duration of the vowels was increased is shown in Table 1.

**Table 1:** Durations and ratios of original long and short vowels in the /tV/ and /tV:/ syllables.

| | ti/ti: | te/te: | ta/ta: |
|---|---|---|---|
| short V in ms | 222 | 192 | 237 |
| long V in ms | 361 | 408 | 415 |
| step size in ms | 29 | 36 | 23 |

### 2.2. Procedure

All participants in this study completed two experimental blocks: one for each experimental condition. Before starting the experiment, participants completed a practice block containing 7 trials. An AXB identification procedure (cf. [9]) was used here because the labels required for a standard forced-choice identification task could not be given to the Spanish participants since Spanish lacks a contrast between short and long vowels. All participants were instructed to identify the X-stimulus as being more similar to either the first (A) or the last stimulus (B). Each test stimulus was presented six times such that the vowel in the utterance in the A position contained three times a short vowel and three times a long vowel. The vowel in the utterance in B always contained the opposite vowel type. Hence, each block consisted of 126 trials. The experimental routine, such as playing the stimuli, randomizing the trials, and recording the responses was run using the software E-Prime.

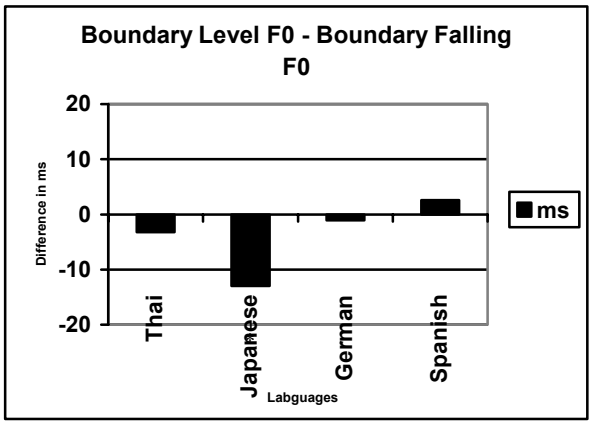
For the analysis of the results, the number of ‘short’ responses, i.e. the number of times out of

the 6 repetitions that participants identified the stimulus as a short vowel, was recorded. The data were then expressed in terms of percent 'short' responses as a function of stimulus, where stimulus number was converted into percentages. For each subject the crossover point from the 'short' category to the 'long' category on each continuum was determined by first transforming the sigmoid function yielded by the raw data into a probit function using SPSS. Then, the category boundary was determined for each of the two experimental conditions (level F0 and falling F0) for each subject. The difference between the values for category boundary in the two conditions was calculated by subtracting the value in the falling F0 condition from that in the level F0 condition. An analysis of variance (ANOVA) was performed on the results for category boundary as well as on the results for the difference between the category boundaries in the two conditions.

## 3. RESULTS

The repeated measurement ANOVA on the results for category boundary revealed a significant interaction between experimental condition and language ($F(3,44) = 4.25$; $p = .01$). Neither significant differences between the three vowels nor a significant interaction between language and vowels was found. A summary of the results for the difference in category boundary location between the two experimental conditions averaged across the three vowels is given in Figure 5:

**Figure 1:** Results for difference between category boundaries in the two conditions level F0 and falling F0, expressed in ms.




The probit functions for the two experimental conditions – again averaged across the three vowels – for each of the four languages are shown in Figures 2 – 5:

**Figure 2:** Percentage of 'short' responses as function of vowel duration steps on the continuum for level and falling F0 in Thai.

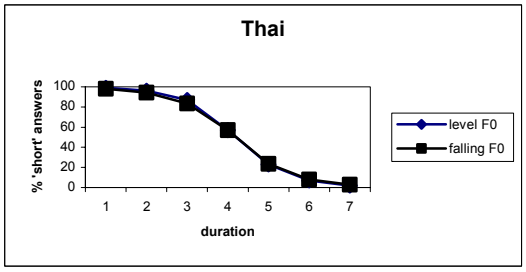



**Figure 3:** Percentage of 'short' responses as function of vowel duration steps on the continuum for level and falling F0 in Japanese.

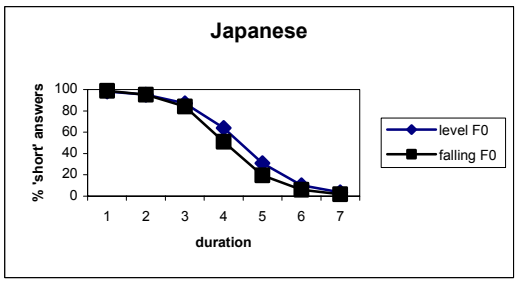



**Figure 4:** Percentage of 'short' responses as function of vowel duration steps on the continuum for level and falling F0 in German.

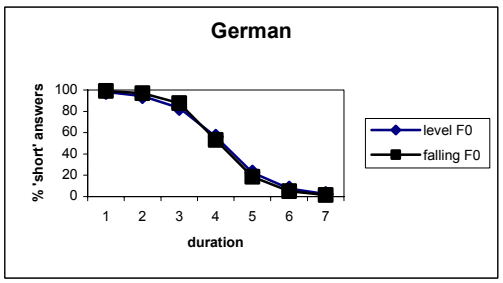
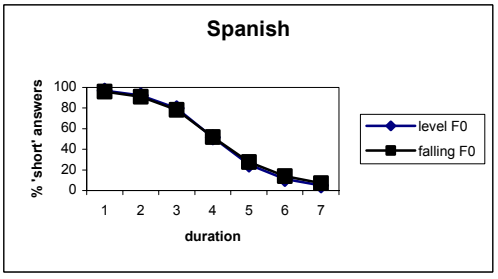



**Figure 5:** Percentage of 'short' responses as function of vowel duration steps on the continuum for level and falling F0 in Spanish.



As Figures 1, 2, 4 and 5 show, the Thai, German, and Spanish listeners did not judge the vowels with a falling F0 differently from those with a level F0. However, Japanese listeners judged the vowels with a falling F0 sooner as long compared to the vowels with a level F0, as shown in Figure 3. Averaged across the three vowels, Japanese listeners exhibited a category boundary shift of -13 ms. Although the difference between the vowels was not statistically significant, the largest shift in the boundary location occurred in the vowel [a] with -17 ms, next was [e] with -12.05 ms, and the smallest shift occurred in [i] with only -9.4 ms.

## 4. DISCUSSION AND CONCLUSION

The results show that whether or not a dynamic F0 leads to an increase in perceived vowel duration does not correlate with the presence of a phonemic vowel length contrast in the native language of the listener. The studies in [1], [2], and [3] found an increase in perceived vowel duration due to a dynamic F0 in native speakers of American English – a language without a phonemic length contrast, and the current study found the same effect in native speakers of Japanese – a language with a phonemic vowel length contrast.

The contrastive use of F0 as either tone or pitch accent does also not correlate with whether or not listeners show the perceptual lengthening effect due to dynamic F0. Japanese, but not Thai listeners were influenced in their judgment of vowel length by the falling F0. At this point the question arises: What leads Japanese listeners to judge vowels with a falling F0 as longer than those with a level F0?

The results here show – similar to the findings of Kinoshita et al. [10] – that Japanese listeners use a falling F0 to distinguish between short and long vowels when the duration is not adequately cued. The fact that the falling F0 leads to an increase in perceived duration of the vowel might be due to a distributional fact of Japanese phonology. While a level F0 may occur on both, short and long vowels in Japanese, a falling F0 may occur on long vowels only. If the duration is ambiguous, the falling F0 biases the Japanese listeners towards a long vowel.

Similar to this explanation for Japanese, is the proposal made by Lehiste [1] for why English listeners might have judged vowels with a dynamic F0 as longer. She suggests that since both vocalic length and changing F0 are correlates of stress in English, listeners might have responded in terms of English stress when they judged the vowels with a dynamic F0 as longer.

The results of this cross-linguistic investigation along with results from earlier studies suggest that an increase in perceived vowel duration due to a dynamic F0 occurs only in languages that associate a dynamic F0 with longer vowel duration.

**Acknowledgements:** This work was supported by NIH Grant DC-02717 to Haskins Laboratories. I would like to thank Doug H. Whalen for his helpful advice throughout the project, and Tsan Huang for use of her perception laboratory at the University at Buffalo.

## 5. REFERENCES

[1] Lehiste, I. 1976. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics* 8, 469-474.

[2] Pisoni, D. B. 1976. Fundamental frequency and perceived vowel duration. *J. Acoust. Soc. Am.* 59, S39.

[3] Wang, W. S. Y., Lehiste, I., Chuang, C. K., Darnovsky, N. 1976. Perception of vowel duration. *J. Acoust. Soc. Am.* 60, S92.

[4] Yu, A. C. L. (to appear). Tonal effects on perceived vowel duration. *Laboratory Phonology* 10. Berlin: Mouton de Gruyter.

[5] Rosen, S. M. 1977. Fundamental frequency patterns and the long - short vowel distinction in Swedish. *Speech Transmission Laboratory, Quarterly Progress and Status Report* 1, 31-37.

[6] Van Dommelen, W.A. 1993. Does dynamic F0 increase perceived duration? New light on an old issue. *Journal of Phonetics* 21, 367-386.

[7] Kohler, K. J. 1985. F0 in the perception of lenis and fortis plosives. *J. Acoust. Soc. Am.* 78, 21-32.

[8] Moulines E., Charpentier, F. 1990. Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9, 453-467.

[9] Best, C. T., Morrongliello, B., Robson, R. 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics* 29, 191-211.

[10] Kinoshita, K., Behne, D. M., Arai, T. 2002. Duration and F0 as perceptual cues to Japanese vowel quantity. *Proc. ICSLP* Denver, 757-760.