Perception of South Kyungsang Korean tones

Seung-Eun Chang

Department of Linguistics
University of Texas, Austin
Sechang71@mail.utexas.edu

Abstract

The research provides a perception experiment which tests the claims made in the previous production experiment (Chang 2008). It tests the perception of isolated synthetic stimuli, in which the crucial acoustic parameters (i.e., F0 peak delay, initial F0, and syllable duration) have been systematically manipulated. The results are generally consistent with the acoustic data, by showing that those acoustic cues contribute to the perception of two tone contrasts. The stimuli tend to be identified as the R item if the F0 peak is later in the syllable and the initial F0 is lower while the F0 peak is earlier in the syllable and the initial F0 is higher. However, although all three factors were certainly different between the two tones in the acoustic data, the three factors did not equally contribute to the perception of tone contrast.

1. Introduction

South Kyungsang Korean is spoken in southeastern part of Korea, and it has lexical tones and tone minimal pairs, such as nuâ 'eye' and nuù 'snow'. The previous acoustic study showed that the minimal pairs are different in initial F0 values, F0 peak delay, and syllable duration in unsuffixed words (the minimal pairs are referred to as H and R, respectively, in this work): R has a lower initial F0, longer peak delay, and longer syllable duration than H does [2]. Taking these results into consideration, we can tentatively hypothesize that these three factors play a part in contributing to the perception of two tones, and no single variable absolutely serves to distinguish H and R. However, the perception study of tonal contrast in this language has not been examined in previous studies.

The purpose of this perception study therefore is to determine whether listeners use those acoustic differences in making word identification decisions. The experiment tests the perception of isolated synthetic stimuli in which the F0 peak delay, initial F0, and syllable duration have been systematically manipulated.

The production study [2] shows that F0 peak is later in the syllable, initial F0 is lower, and syllable duration is longer for R than for H. It was seen that the average relative peak delay for 6 speakers was 0.58 for H, suggesting that the peak came at 58% of the syllable duration for H. It was 0.84 for R, indicating that the peak came at 84% of the syllable duration. The average syllable duration for six speakers was 215 (ms) and 295 (ms) for H and R, respectively. The average initial F0 value for female was 243 (Hz) and 165 (Hz) for H and R, and the value for male was 121(Hz) and 102 (Hz) for H and R. In principle, all three factors could be strong cues for the distinction between H and R.

Therefore, I hypothesize the late peak delay, low initial F0, and long syllable duration would trigger R response. In sum, the proposed hypotheses can be summarized, as in (1).

(1) Hypothesis
a. The stimuli tend to be identified as the R item if the F0 peak is later in the syllable, initial F0 is lower, and duration is longer.
b. The stimuli tend to be identified as the H item if the F0 peak is earlier, initial F0 is higher, and duration is shorter.

2. Methodology

2.1. Stimuli

Two words that differ minimally in lexical tone, /kan/ (H 'taste' - R 'liver'), were chosen for synthesis. One female native speaker of South Kyungsang Korean produced the isolated H /R tone minimal pairs. Recordings were made in a sound-treated booth in the Phonetics Lab of the Linguistics Department, University of Texas at Austin, using Praat.

With the original sound object, I created a /kan/ continuum using Praat on a PC, by removing the pitch points except for the crucial three points, i.e., initial, peak, and final pitch point. The duration and F0 values at the initial, peak, and final points for each tone type in the original recording produced by a female speaker were as follows.

	Initial point	Peak point	Final point
H	0 (ms) - 216(Hz)	201.5 (ms) - 240 (Hz)	450 (ms) - 136 (Hz)
R	0 (ms) - 170(Hz)	483.6 (ms) - 224 (Hz)	620 (ms) - 126 (Hz)

Table 1: The F0 values (Hz) and duration (ms) at three points in original recordings

Based on these values in original recording, the stimuli for the perception tests were designed to vary the initial F0 in two steps, i.e., low (170 Hz) and high (216 Hz), and syllable duration in two steps, i.e., short (450 ms) and long (620 ms). The F0 peak delay was varied from 10% to 80% of syllable duration in 10% steps, for a total of 8 variants. The actual values for the F0 peak delay were 45ms, 90ms, 135ms, 180ms, 225ms, 270ms, 315ms, and 360ms for the short syllable duration (450 ms).

I manipulated each pitch point according to the proposed values above, by dragging the pitch point, and generated 16 syllables for the short syllable duration. The schematized picture for the manipulated parameters is given in Fig. 1 (a). The schematized picture for the manipulated parameters is given in Fig. 2 (b).

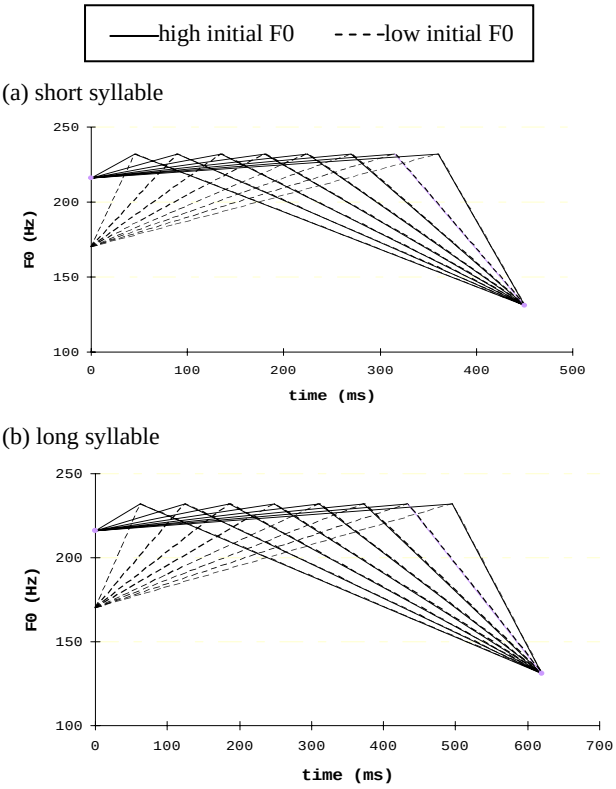

Figure 2: A schematic depiction for 36 /kan/ continua

The actual F0 peak delay for long syllable (620ms) were 62ms, 124ms, 186ms, 248ms, 310ms, 372ms, 434ms, and 496ms, and its /kan/ continuum also consists of 16 syllables. The total /kan/ continuum thus consists of 32 syllables (2 initial F0 * 2 duration * 8 turning point). The peak F0 and final F0 in each stimulus were fixed at constant values, i.e., 232 (Hz) and 131 (Hz) for each syllable, from the average values of H and R in the original recording.

2.2. Subjects

Ten adult subjects (aged 20 ~ 58 years old, 5 females and 5 males) from the South Kyungsang area participated. All subjects are linguistically naïve, and all reported that they had no history of hearing impediment. They were born and raised in Pusan, and five of them had lived only in Pusan before coming to the U.S., and five of them had lived in Pusan over 20 years and lived in Seoul for 7 to 10 years.

2.3. Procedure

The stimuli were played in randomized order on an audio system using an ALVIN software program on a PC. Subjects heard 10 repetitions of each stimulus, a total of 320 tokens. There was a two-second interval between stimuli.

The subjects were instructed to respond to each item as quickly as possible by pressing the button corresponding to pictures and English words for “taste” (H) or “liver” (R) because Korean orthography is exactly same for the two /kan/s. Since encoding picture was not available in ALVIN program to my knowledge, I made the appropriate pictures and each picture was attached over the corresponding English word on PC screen. I used the Korean soy sauce image for “taste” (H) because we have a common expression “taste with soy sauce”, and the liver image for “liver” (R).

To avoid misidentification, subjects were asked to pronounce the button labels prior to the experiment. A practice session consisting of 10 test items in each block preceded the test. These practice items provided listeners with end points for each parameter manipulated, as well as stimuli in between. Subject responses were collected by computer using an ALVIN program, and responses for each stimulus were added across speakers.

3. Results

The overall results indicate that R response was triggered when the timing of F0 peak was late, initial F0 was low, and syllable duration was long. The effect of three factors for all ten subjects is illustrated in one graph as in Fig 3.

It gives the average number of “R responses” out of 10 for all stimuli, arranged according to the “F0 peak delay” and the combination of “initial F0 and syllable duration”, showing all interactions of three factors. The F0 peak delay is represented along the horizontal axis as the relative peak delay (percent), and the average number of R responses is on the vertical axis. The maximum of average number R responses is “10”, because one stimulus was heard 10 times to the listeners.

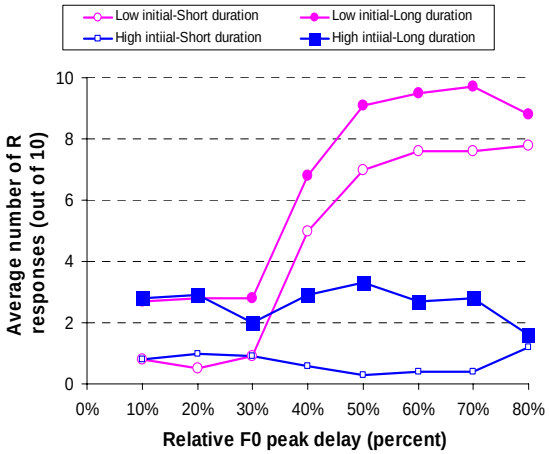

Figure 3: The average of R responses according to F0 peak delay, initial F0, and syllable duration (subjects pooled)

To take one example, when the relative peak delay is 70%, initial F0 is low, and syllable duration is long, which is marked with “dark circled line” at 70% timing, the average R responses for that stimulus is almost 10. This indicates that the stimulus was judged as an R item by the subjects for every 10 trial. On the other hand, when the relative peak delay is 10%, initial F0 is high, and syllable duration is short, which is marked with empty squared line at 10% timing, the average R responses is below 2, indicating that this stimulus was judged as an R item less than 2 times for all 10 trials.

Specifically, one notable pattern is that in the dark circled line, the rise begins at 30%, reaching peak slope at 40% with sharp rising curve, and this line has the highest number of R responses. The empty circled line also shows similar pattern

with the dark circled line, except that the average number of R responses are slightly lower than that.

This suggests that F0 peak delay have crucial effect on the combination of low initial F0, regardless of duration, triggering more R responses at the later peak delay. However, the F0 peak delay seems to have little effect on the high initial F0. It appears that if initial F0 is high, listeners don't pay attention to timing of peak in identifying H vs R.

It might be because the stimuli with 'high initial F0 and late peak delay' were heard unnatural. Namely, the peak F0 was fixed at 232(Hz) in all stimuli, and the high initial F0 value was 216(Hz). So, the initial F0 value was not that different from the peak F0 value in "high initial F0 stimuli", and this might be heard as a sound with a long peak plateau with a high initial F0, which is not a real word in South Kyungsang Korean. For these stimuli, most subjects seem to respond "H" toned word.

In general, the combination of late peak delay, low initial F0, and long duration triggered highest R responses, and the combination of high initial F0 and short duration has the lowest R responses.

To determine whether these three factors -the timing of the F0 peak, the initial F0, and the duration- significantly affect identification for H and R contrast, logistic regression analysis was conducted. This is appropriate because the dependent variable is binary, H or R. As such, the dependent variable was the dichotomous choice, H or R. H was coded as 0 and R as 1. The independent variables were the peak delay (in milliseconds), initial F0, and duration. The *peak delay* (a) was used instead of percentage of eight variants because continuous variables can provide more informative interpretation in logistic regression. *Initial F0* (b) was coded as 0 for low and as 1 for high. *Syllable duration* (c) was coded as 0 for short and as 1 for long.

The equation for such model is presented in Table 2, together with the coefficients (and odds ratios). The positive coefficient indicates a direct relationship to the dependent variable. For example, the fact that the coefficients for (a) peak delay are positive indicates that as the peak delay is longer, there is a *more* likelihood for the R response. On the other hand, negative coefficient indicates an inverse relationship with the dependent variable, indicating that as the (b) initial F0 is higher, there is a *less* likelihood for the R response.

(a) F0 peak delay	(b) initial F0	(c) duration	(d) constant
.003(1.003)	**-1.075 (.341)**	0.011(1.011)	-0.415(.661)

Equation: Log odds of R response
= (a×peak delay) + (b×initial F0) + (c×duration) + d

Table 2: Logistic regression analysis of R response in terms of peak delay, initial F0, and syllable duration across subjects

Since this coefficient is in log units, we cannot directly interpret the magnitude of the change. Therefore, the odds ratio converted in decimal units are provided with parenthesis, and it represents the change in the odds of having dependent variable for a one unit of change in the independent variable.

The result, for example, can be interpreted that every one-millisecond-increase in *peak delay*, the odds of R increase by a factor of 1.003 times, i.e., 1 (ms) longer peak delay is 1.003 times *more* likely to be judged as an R. Although 1.003 for peak delay is an extremely small increase, there's a lot of range in milliseconds. For example, if the peak delay is 100 (ms) longer for R than H, then it is 100.3 times more likely to be judged as an R. The stimuli with *high initial F0* is 0.341 times *less* likely to be responded as an R. Likewise, the stimuli with *longer duration* is 1.011 times *more* likely to be responded as an R.

The predictor variables make a significant contribution to the model ($p < .001$) is highlighted in boldface. Only "peak delay" and "initial F0" cues significantly contribute to the perception of H and R contrast.

4. Conclusions

The goal of this experiment was to determine whether the acoustic dimensions of peak delay, initial F0, and syllable duration contribute to the perception of South Kyungsang Korean H and R contrast. The results of the logistic regression analysis seen in Table 2 can be interpreted as perceptual models of H and R. The models show that South Kyungsang Korean listeners use these differences when making word identification decisions.

The critical predictor variables were the peak delay and initial F0, i.e., for R to be identified, the perceived target for the peak must be later, and the initial F0 must be lower, than H. The syllable duration also contributes to the model, i.e., for R to be identified, the perceived target for the duration must be longer than H. However, this was a significant factor in the model only for three listeners.

Although all three factors were certainly different between the two tones in the acoustic data, the three factors did not equally contribute to the perception of tone contrast for all subjects. It showed that if one cue is more extreme for a category, other cues can be less so, as observed in perception studies of other languages [1,3,4].

Further, the boundary between the two categories in perception data occurred rather early, i.e., 40% of relative peak delay. The average relative peak delay was 50% for H and 89% for R in the acoustic data [2], namely, a peak F0 located *after* 40% of the syllable duration tends to be judged as an "R" tone. Peaks located *before* the 40% are judged as an "H" tone. The boundary for the response time was consistent with the data of the R response, that is, the response time had a peak at 40% of relative peak delay, suggesting that the stimuli were heard as most ambiguous in this location of the peak.

In addition, H tone perception seems to tolerate more variability overall, while R requires a late peak delay and a low initial F0. This trend also appears to be relevant, as seen in the higher number of H responses than of R responses overall. As noted before, unnatural stimuli were present, including a combination of 'high initial F0 and late peak delay' as well as 'extremely early peak delay'. Most of those stimuli were finally judged as H, although the subjects appeared to have trouble responding. Another possible account is that the H toned words are much more frequent than R in this language, and it might induce listeners toward "H" response when they heard the unnatural stimulus.

An interesting aspect of the results was the between-subject variation. Although every female subject showed relatively consistent responses, male subjects had large within-subject variations in their responses. One possible interpretation is that the gender of the speaker might influence

the perception of the listener. If the speaker and listener are the same gender, the listener may be more familiar with the speaker's voice and thus perceive the stimuli more precisely than would a listener of the other gender.

Overall, the data confirm that South Kyungsang Korean listeners have two tone contrasts, H-class and R-class, which differ in F0 peak delay, initial F0, and syllable duration.

5. References

[1] Abramson, A.S. 1975. The tones of Central Thai: some perceptual experiments. In J.G. Harris and J. Chamberlain (eds.), *Studies in Tai Linguistics*. Bangkok: Central Institute of English Language, p.1-16.

[2] Chang, S-E. 2008. Tone alternations in South Kyungsang Korean. Linguistic Society of America (LSA), Chicago, IL, 2008

[3] Gandour, J.S. 1978. The perception of tone, In V.A. Fromkin (ed). *Tone: A Linguistic Survey*, p.41-76. New York: Academic press.

[4] Lin, H.B., and B.Repp. 1989. Cues to the perception of Taiwanese tones. *Language and Speech* 32.1: 25-44.