



Effects of vowel duration on the perceived naturalness of English monosyllabic words ending in a stop: Some preliminary findings*

Eon-Suk Ko**

Department of English Language and Literature, Chosun University, Gwangju, Korea

Abstract

Preliminary findings are reported from five experiments testing the perceived naturalness of word tokens whose vowel durations are altered. The stimuli were minimal pairs of English words ending in a voiced/voiceless plosive. Results show an asymmetric effect of shortening and lengthening of the vowel on the perceived naturalness of the word. Incremental shortening of vowel duration initially shows a stable degree of perceived naturalness but rapidly deteriorates beyond a certain point. On the contrary, only a small degree of lengthening of the vowel made the perceived naturalness of the word quickly decay, but there was a floor effect such that the perceived degree of naturalness does not lower beyond a certain level. Further, the tokens with the original vowel duration were not always scored higher than the stimuli with a small degree of shortening. Future studies should address the issue of speaking rate and the ratio between the vowel and the stop closure duration to better understand the phenomenon. The issue investigated here has implications on the role of prototypical exemplars in the perception of phonotactic naturalness.

Keywords: perception, naturalness, prototype, vowel duration, voicing

1. Introduction

Vowels before voiced consonants are pronounced longer than the ones before a homorganic voiceless consonant if pronounced in the same prosodic conditions (Denes, 1955; Hogan & Rozsypal., 1980; House & Fairbanks, 1953; Raphael, 1972, among others). This durational difference in a pre-consonantal vowel before voiced and voiceless consonants has been known to affect the perception of the voice feature of the following consonant (Luce & Charles-Luce,

1985). A related question that has received little attention in previous research is how the perceived naturalness of the tokens, provided they are judged as having the same voice feature, differs as a function of the magnitude of manipulation. This paper explores the effects of vowel duration on the perceived naturalness of phonotactic patterns.

If speech perception involves evaluation of various features for prototype matching followed by pattern classification, as in the fuzzy-logical model (Massaro, 1989; Oden & Massaro, 1978), how

* The experiments reported in this paper were carried out while the author was affiliated to previous institutions. I would like to thank the speakers who recorded the stimuli, the research assistants who helped with the administration of the tests, and the three anonymous reviewers for their helpful comments.

** eonsukko@chosun.ac.kr, Corresponding author

Received 8 June 2021; Revised 18 June 2021; Accepted 18 June 2021

© Copyright 2021 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

do listeners process the segmental duration in perceiving the naturalness of the phonotactics in the words they hear? For example, if the vowel in the word *back* is lengthened twice its original duration, and the word *bag* is shortened to half its original duration, will listeners perceive two altered tokens natural or unnatural to a comparable degree?

In preparing for the stimuli used for the experiment in Ko et al. (2009), it was found that shortening of the vowel in a word such as *bag* to 50% of its base duration is perceived as more natural than doubling up the vowel duration in *back*. Inspired by this observation, a series of follow-up experiments have been conducted to explore the relationship between the degree of modified vowel duration and the perceived naturalness of the stimuli. Although a full picture explaining the various factors affecting the perceived naturalness of phonotactic patterns has yet to be completed, this paper reports some preliminary findings so that a more informed and better designed research could be performed building on these preliminary results.

The question of *naturalness* is related to the question of *goodness* of categories often investigated in the perception literature (e.g. Kuhl, 1991). The notion of naturalness investigated in this paper, however, is slightly different from goodness. If the notion of goodness is related to the prototypicality of a single phoneme, naturalness as used in this paper is a *syntagmatic* notion in that it is presented in the context of a word, not in an isolated phoneme. The listeners will, therefore, process and evaluate the (un)naturalness of the heard sequence of phonemes based on their knowledge about the phonotactics of their language.

In a study investigating prototype for speech categories (e.g. Kuhl, 1991; Samuel, 1982), it is standard to design the study with a set of stimuli from a single speech category with a range of acoustic variation so that we can inspect what factors affect the listeners' perception of the stimuli. The series of experiments reported in this paper, however, are somewhat deviant from such a practice. It is partly due to how this research got started, but also because the main issue addressed in this study is the question of naturalness with regard to phonotactics instead of phonological categories. It is an important distinction since phonotactic naturalness necessarily involves the context and the target locus of acoustic variation will always be gauged against the phonotactic contexts.

The series of experiments reported in this paper was initiated by an ulterior motive to find comparable degree of (un)naturalness, or (non-)prototypicality, to be used for another perception experiment with infants. Ko et al. (2009) investigated infants' development of perceptual sensitivity for the vowel duration conditioned by the voicing feature of the following consonant. In the experiment, half the infants were presented with stimuli containing short vowels and the other half with stimuli containing long vowels. In both groups, infants were presented with the Match type CVC stimuli with the natural phonotactic pattern of vowel and consonant (e.g. long vowel followed by a voiced stop), and the Mismatch type stimuli involving the opposite relationship between the vowel and consonant (e.g. long vowel followed by a voiceless stop). Infants might respond differently to the Match and Mismatch type if they have developed

sensitivity to the phonotactic pattern of English regarding vowel length conditioned by consonantal voicing. In such a design, it was important to balance the degree of naturalness between the shortened and lengthened stimuli so that the degree of naturalness does not become a confounding factor in comparing infants' response. The experiments 1–3 were conducted with the goal of selecting Mismatch tokens with a comparable degree of naturalness scores to be used for infants' perception tests, so the experimental design might not seem optimized for the task of goodness rating of phoneme categories. The experiments 4–5 were conducted without consideration for infant experiments, so were conducted with base tokens produced in a regular adult-directed speech register.

Before describing the experiments, it is worth noting at the outset that all based tokens used for deriving altered vowel durations of varying degrees were produced with an explicit cue for voicing during closure and a strong release burst of the stop consonant at the end of the word so that there can be a clear cue for the voice feature from the release burst as well as voicing during closure. All responses in the experiments are in the Likert scale on a 5-point step (5=natural, 1=unnatural). Some raise concern about using linear-mixed effects models for analyzing Likert-scale data, but there are recommendations to use parametric tests such as mixed effects linear regression for analyzing Likert-scale data (Gibson et al. 2011; Norman, 2010). All experimental sessions were preceded by a brief practice session where participants were presented with a randomly chosen set of three to six stimuli set to familiarize themselves with the test. In analyzing the responses of participants, statistical models based both on the scores z-transformed at an individual level and the raw scores on the Likert-scale were constructed, but the results did not show substantial differences. Therefore, only the analyses based on response scores in the Likert-scale are reported.

2. Experiment 1

The goal of this experiment was to select specific tokens of three minimal pairs which are similar in their perceived naturalness when their vowel duration is altered to be used for testing infants' response to unnatural phonotactic patterns (Ko et al., 2019). Given the goal, the range of vowel modification was limited to one instance of lengthening or shortening for each token. Instead, the intonation was systematically varied in generating the base tokens.

2.1. Methods

2.1.1. Stimuli

A female native speaker of American English in her mid-twenties produced three minimal pairs (*pig/pick*, *bag/back*, *cup/cup*) in varying intonational contours. In order to explore the effects of intonational contour, in addition to vowel duration, on the perceived naturalness of the word, three representative contours of final, question, and falling were selected for each word type.¹ The vowel durations of these tokens were lengthened to 160% and 200% of the original vowel duration if ending in a voiceless stop, and shortened to 50% and 60% of the original vowel duration if ending in a voiced stop

¹ Both the final and falling contour sound like a statement, but the final carries a signal for continuation whereas falling sounds as if the speaker is trying to garner attention with rapidly falling contour.

based on the PSOLA (Pitch Synchronous Overlap and Add) method using Praat (Boersma & Weenink, 2001). For each word token, the base token with the unmodified vowel was also presented. Thus, a total of 36 stimuli (6 word types×3 contours×2 levels of modified duration=36) were generated. The stimuli were equalized on their amplitude.

2.1.2. Participants

Thirteen undergraduate students (7 males) at Brown University participated in this study. They were in their early twenties and had no known hearing problems. All were paid \$5 for their participation.

2.1.3. Procedure

The participants were tested one at a time in a quiet office. The participants were told to rate the stimuli on their perceived naturalness. The stimuli were presented once in a random order. They listened to the stimuli over a high-quality headphone in a self-controlled pace, and responded by a mouse click on a PC. The task was carried out using Praat's Experiment MFC (Multiple Forced Choice) listening experiment object and was self-paced.

2.2. Results

The perceived naturalness for the unmodified ($M=3.62$, $SD=1.15$) and altered stimuli ($M=2.99$, $SD=1.20$) was significant ($t=-5.81$, $df=466$, $p<0.001$). An interesting observation in Figure 1 is the difference between the responses in lengthening and shortening. A mixed effect linear regression model with naturalness (5 levels) as the dependent variable, modification (original, lengthened, shortened) and prosody (falling, question, final) as fixed effects, and the participant and the stimulus as random effects showed significantly less naturalness responses for the lengthened stimuli than natural ($\beta=-.82$, $SE=0.16$, $t=-4.97$, $p<0.001$), but there was no significant difference in the perceived naturalness between the shortened and the natural vowel duration ($\beta=0.07$, $SE=0.16$, $t=0.48$, $p=0.63$). Final intonation was perceived to be more natural than falling ($\beta=0.32$, $SE=0.12$, $t=2.49$, $p<0.05$). In addition, a significant interaction was found between lengthening and question ($\beta=-0.69$, $SE=0.22$, $t=-3.11$, $p<0.01$). An inspection of the data revealed substantially lower naturalness when target words produced in a question intonation was lengthened than in other conditions. This is likely because word tokens produced in a question intonation were produced with relatively longer vowel duration than the others.

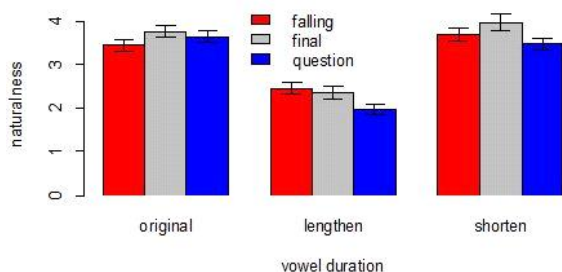


Figure 1. Perceived naturalness of words with the original, 200% lengthening, and 50% shortening of the base vowel duration in three intonational contours.

2.3. Discussion

Overall, shortening was perceived to be more natural than lengthening, and the final intonation was perceived to be more natural than falling. While the asymmetric perceived naturalness in shortening and lengthening was intriguing, there were limitations in this pilot test in that only one instance of each word was used as the base token for generating the natural and altered stimuli for the perception test. I address this limitation in the next experiment.

3. Experiment 2

The goal of this experiment was the same as the first one, i.e. selecting instances of lengthening and shortening that are perceived with similar degree of naturalness. In this experiment, the effect of lengthening and shortening the vowel duration on the perceived naturalness of words is investigated based on multiple instances of the same word type.

3.1. Methods

The methods adopted in Experiment 2 were largely the same as Experiment 1 except for some modifications in the generation of stimuli. Specifically, the intonational contour is now focused only on the final type since it was perceived to be more natural than the other types. Further details are described below.

3.1.1. Stimuli

Six instances of each word were selected from a multiple production of the same three minimal pairs used in Experiment 1. Based on the finding in Experiment 1 that shortened stimuli are perceived to have a greater degree of naturalness than lengthened stimuli, I shortened the stimuli to 50% in this experiment, but not to 60%, so that we can narrow down the candidates for tokens with comparable naturalness in the shortening and lengthening categories. For lengthening, I maintained the two-step manipulation of 160% and 200% elongation of the original vowel duration in the base tokens. A total of 90 tokens {[3 voiced word types×6 word tokens×2 levels of duration (100%, 50%)]+[3 voiceless word types×6 word tokens×3 levels of duration (100%, 160%, 200%)]} were generated to be tested.

3.1.2. Participants and procedure

Ten undergraduate students (5 males) at Brown University participated in the study. They were in their early twenties and reported no known issues in hearing. Each participant sat at a table in a sound attenuated booth and listened to the stimuli presented over a loud speaker. They were instructed in the same manner as in Experiment 1, and were paid \$5 for their participation. Presentation of the stimuli and collection of the responses were conducted using Praat's Experiment MFC (Multiple Forced Choice) listening experiment object.

3.1.3. Results

The mean perceived naturalness was 3.53 ($SD=1.25$) for the shortened stimuli, 3.80 ($SD=1.25$) for the stimuli with original duration, 3.07 ($SD=1.22$) for stimuli lengthening to 160%, and 2.60 ($SD=1.23$) for the stimuli lengthened to 200% of the original duration. We constructed a mixed effects linear regression model to test the effect of vowel duration on the perceived naturalness of the stimuli. The design of the experiment involves a nested structure in the

random effect because the stimulus set contains the 6 variations of a single word type (e.g. bag1, bag2, ...bag6) that serves as the basis of lengthened or shortened stimuli. We thus constructed a mixed effects model with the following formula to reflect this nested structure, where the dependent variable was the perceived degree of naturalness and the fixed effect was the degree of modification or vowel duration in percent of the base token, and the random effect was the participants and the base tokens of the stimuli nested under the word type: (lmer(naturalness~vowelDuration+(1|participant)+(1|wordType:baseToken)).

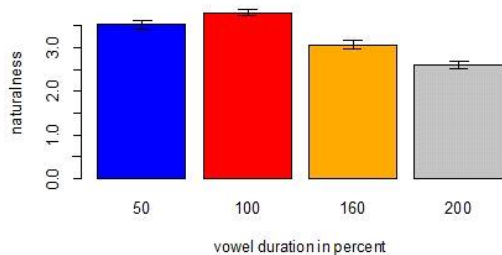


Figure 2. Perceived naturalness of words with 50% shortening, original, 160%, and 200% lengthening of the original vowel duration.

The results show that the stimuli with unmodified vowel duration are perceived to be significantly more natural than lengthened or shortened tokens (Figure 2; all p 's <0.001). The stimuli with 200% lengthening were perceived to be significantly less natural than the tokens with 50% shortening ($\beta=-0.65$, $SE=0.16$, $t=-4.0$, $p<0.001$), but there was no significant difference between the stimuli shortened by 50% and lengthened to 160%.

Since the original goal of this experiment was to select specific tokens of Mismatch tokens with comparable degree of perceived naturalness, an inspection was made at the individual token level. There was a great variation in the perceived naturalness among word types. For example, shortening of the word type *pig* ($M=4.09$, $SD=0.9$) was perceived as substantially more natural than shortening of *bag* ($M=3.03$, $SD=1.31$) or *cub* ($M=3.45$, $SD=1.24$) even though they are all words of CVC type ending in a voiced stop. I was able to select three tokens of each Mismatch word type from the 50% and the 160% tokens, whose mean naturalness score ranged from 3.2 to 3.6 to be used for another experiment in fulfillment of the goal of the experiment. The original vowel duration for these samples is in Table 1. The experiment based on these selected Mismatch stimuli are described in Ko et al. (2009).

Table 1. Mean vowel duration of the three selected Mismatch tokens

Voiced	Vowel duration (ms.)	Voiceless	Vowel duration
bag	297.4	back	119.5
cub	166.8	cup	95.6
pig	215.2	pick	110.4

3.1.4. Discussion

Experiment 2 confirmed the tendency found in Experiment 1 that shortening the vowel into half is perceived to be more natural than lengthening it twice as its original duration. Though not systematically analyzed, there was also a great variation in the perceived degree of naturalness among individual tokens of the same word type. The results call for a manipulation of the vowel duration in a

more systematic and incremental way with one instance of a word token for each word for a better understanding of the effect of vowel duration on the perceived naturalness of the word token.

4. Experiment 3

In this experiment, the perceived naturalness of nonce words with varying degrees of vowel duration was tested. The purpose of the test was to investigate if the results found in Experiments 1-2 are based on the stored memory traces of base word tokens, i.e. prototypical exemplars, or a higher-level grammatical knowledge of phonotactics in their language. If the latter, we would expect the results in the previous two experiments to be replicated in words that they have never heard.

4.1. Methods

4.1.1. Stimuli

A female native speaker of American English aged around sixty read three minimal pairs of nonce words (*zag/zack*, *gub/gup*, *mig/mick*) in infant-directed voice. Six tokens of each word type were selected to be manipulated. The vowel in the words ending in a voiced stop was shortened to 90% and proceeded stepwise by 10% increments to 40% of the original duration, and the words with a voiceless stop was lengthened to 140% and incrementally to 190% of the original vowel duration via the PSOLA method of Praat. A total of 216 stimuli were generated (6 word types×6 word tokens×6 steps of vowel duration).

4.1.2. Participants and procedure

Twelve native speakers of American English participated in the experiment. They were college undergraduate students in their early twenties attending University at Buffalo. None of them reported any known hearing problems. Three additional students participated but their data were not included in the analysis due to their being a non-native speaker of English. The participants received \$10 for their participation.

The test took place in a sound-attenuated room on the university campus via a high-quality headphone. Stimuli were presented in Praat's Experiment MFC in a self-controlled pace. Participants were asked to judge the voicing category of the final consonant, and then choose the level of naturalness for the heard stimuli. They were exposed to the stimuli during the practice session so were aware that they would be listening to nonce words.

The 216 stimuli were presented three times in a random order. Thus, each participant generated 648 responses, totaling in 7,776 judgments.

4.2. Results

The correct responses for the voicing category were 99.3% for the stimuli ending in a voiced consonant, and 95.6% for the voiceless. A visual inspection of the results in Figure 3 shows that the perceived degree of naturalness is greater in the case of shortening than in lengthening, consistent with the findings in Experiments 1-2. This has been confirmed by a mixed effects linear regression model (lmer(naturalness~voicing+(1|wordType:baseToken)+(1|participant)), which showed a significantly lower naturalness scores for the voiceless category that underwent lengthening than the voiced that

underwent shortening ($\beta=-0.92$, $SE=0.07$, $df=213$, $t=-11.97$, $p<0.001$).

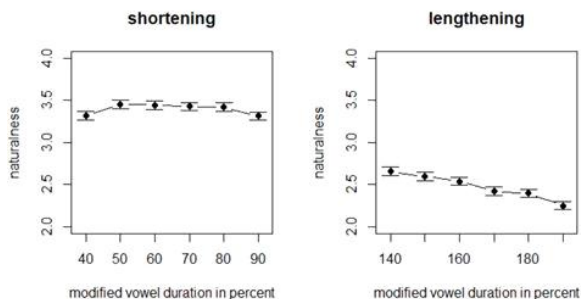


Figure 3. Perceived naturalness of words in nonce words. Words ending in a voiced stop was shortened (left panel) whereas words ending in a voiceless stop was lengthened (right panel).

4.3. Discussion

The results of Experiment 3 show that the asymmetric perception of the naturalness for lengthening and shortening is replicated in nonce words. This finding suggests that the judgment for naturalness is likely to be based on speakers' knowledge of native language phonotactic patterns rather than in reference to stored prototypical exemplars.

There could be two possible reasons for the relatively low naturalness scores in this data. First, unlike in the previous experiments, participants were asked to judge the voicing feature of the coda consonant before judging the naturalness of the stimuli. This could have added to task-related fatigue. Second, the stimuli were all nonce words which the participants might have never heard before, so the novelty of the stimuli might have added to the unnaturalness of the manipulated vowel duration. If this is the case, it could be that the judgment of the naturalness is a combination of the reference to prototypical exemplars as well as speakers' knowledge of native language phonotactics.

5. Experiment 4

The previous three experiments revealed interesting asymmetries in the perceived naturalness between lengthening and shortening of vowel duration. They were, however, targeted on finding specific tokens to be used for experimental stimuli for another perception test for infants, thus the range of manipulated degrees of vowel duration was somewhat ad hoc. Thus, a follow-up experiment was conducted to inspect the range of responses with more systematic modification of the vowel duration.

5.1. Methods

5.1.1. Stimuli

The three minimal pairs of English CVC words, *bag/back*, *cup/cup*, and *pig/pick*, contrasting in the voicing of the coda consonant were recorded by a male native speaker of American English. One of the multiple productions of each word was chosen to be used as a base token for the modification of the vowel duration. The vowel duration of the voiced series was shortened in 9 steps, starting with 90% and proceeding stepwise in 10% increments of the original duration all the way to 10%. The vowel duration of the voiceless series was lengthened in 10 steps, starting with 110% and proceeding

stepwise in 10% increments of the original vowel duration all the way to 200%.

5.1.2. Participants and procedure

Participants, 4 males and 5 females, were undergraduate students attending University at Buffalo. An additional participants' data were discarded due to a technical failure ($n=1$). They were recruited from an introductory linguistics class and received an extra credit for participation in the experiment.

The tests were conducted with one participant at a time in a sound attenuated booth using Praat's Experiment MFC listening experiment object. The stimuli were presented three times in a random order. A total of 5,670 judgments were generated $\{[(3 \text{ voiced words} \times 9 \text{ steps} \times 10 \text{ repetitions}) + (3 \text{ voiceless words} \times 10 \text{ steps} \times 10 \text{ repetitions}) + (6 \text{ unmodified words} \times 10 \text{ repetitions})\} \times 9 \text{ participants} = 5,670$.

5.2. Results

Compared with the first three experiments, the stimuli in this experiment is different in that the speaker producing the base tokens is a male, and the tokens were produced in an adult-directed manner as opposed to the child-directed register in the previous experiments. Nevertheless, we were able to see the tendency being replicated. If we compare the naturalness ratings for 50% ($M=3.43$, $SD=1.12$) and 160% ($M=3.14$, $SD=1.20$) as in the first two experiments, we can observe that a higher rating for shortening is maintained.

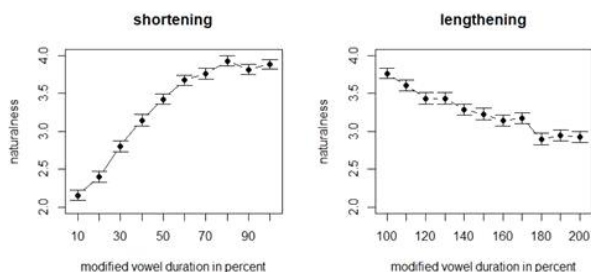


Figure 4. Perceived naturalness for CVC words with modified vowel duration. Words ending in a voiced consonant were shortened whereas those ending in a voiceless consonant were lengthened.

An observation of the change in the degree of naturalness in shortening and lengthening in Figure 4 suggests that the perceived naturalness of the word decreases more sharply in lengthening than in shortening earlier in the modification process moving away from the original duration. Thus, even a small degree of lengthening from the original duration makes the stimuli judged as substantially less natural whereas the perceived naturalness does not rapidly deteriorate in shortening up to a certain point in the incremental reduction process. But once a certain point is reached in shortening, the perceived naturalness rapidly deteriorates. In contrast, there is some sort of a floor effect in lengthening such that the perceived naturalness does not deteriorate any longer beyond the 180% point, for example.

To inspect the overall pattern of response in shortening and lengthening, I put together the data on shortening and lengthening, and compared the responses in the continuum (Figure 5). Often, manipulations of a linguistic unit on the same linear scale may have different degrees of effects on perception. To explore the question of whether the relationship between the perceived naturalness of the

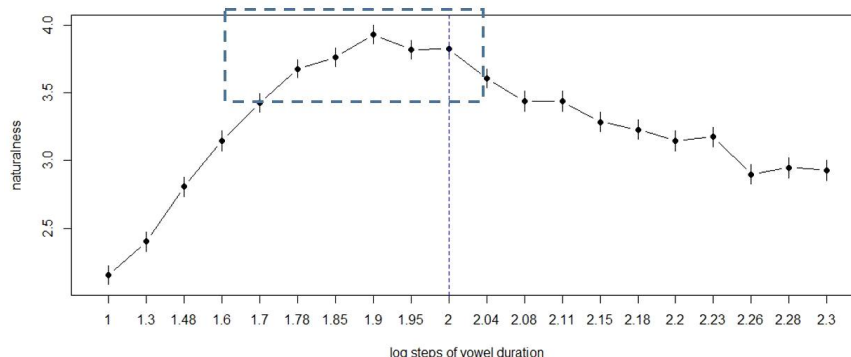


Figure 5. Figure 3 re-drawn with the two sets of data merged on a continuum, and the steps of vowel duration on the x-axis converted to a log scale.

vowel duration and steps of vowel duration is logarithmic in our sample, I converted the steps of vowel duration to a logarithmic scale. The original vowel duration, i.e. 100% is now represented as 2 ($\log_{10} 100=2$), and its location is marked with the blue line. Note that the scale of the log vowel duration on the x-axis does not reflect the uneven distances between the log steps of vowel duration.

If the perception of vowel duration was logarithmic, listeners would treat a greater elongation of the vowel to be comparable to a less degree of shortening. For example, a vowel shortened to 50% ($\log_{10} 50=1.7$) would be equated to a vowel lengthened to 200% ($\log_{10} 200=2.3$) since both are 0.3 away from the log of 100%, i.e. 2. However, the naturalness score at 50% ($M=3.43$, $SD=1.12$) is substantially higher than that at 200% ($M=2.93$, $SD=1.22$). Instead, it seems that there is a symmetric degradation of naturalness as the modification progresses away from the peak naturalness point at 1.9, i.e. 80%, to a certain extent, as indicated with a dotted box for ease of reference.

5.3. Discussion

Previous observation that shortening and lengthening show an asymmetric degree of perceived naturalness was again replicated. Interestingly, the highest degree of perceived naturalness fell on a slightly shortened stimuli instead of the tokens with the original vowel duration. It could be that the base word tokens were produced in isolation and are relatively long thus listeners are more used to a slightly shorter vowel duration than the production we used as base tokens.

6. Experiment 5

In Experiment 4, the modification of the vowel duration was asymmetric. That is, vowels followed by a voiced consonant were shortened whereas vowels followed by voiceless consonants were lengthened. In this experiment, we apply the same spectrum of modification for both the voiceless and voiced series, without going too extreme in the degree of modification. In the previous experiment, we observed that the perceived naturalness is below the neutral score of 3.0 when shortened to the 40% or shorter of the original base token or lengthened to the 180% or greater of the original duration. In this experiment, therefore, we focus our attention on the range of 40% to 160% of the vowel duration.

6.1. Methods

6.1.1. Stimuli

Base tokens of two minimal pairs, *bag/back* and *pig/pick*, were modified so that the vowel duration was altered in 13 steps from 40% to 160% of the base token in an increment of 10%. The stimuli were presented three times in a random order.

6.1.2. Participants and procedure

A total of 16 college students (7 males) participated in the experiment. They were recruited from students taking a linguistics class at University at Buffalo, and received an extra credit for their participation. All were native speakers of North American English in their early twenties without a hearing problem. The experiment was conducted in a quiet room. The test was composed of 8 blocks, each containing a total of 39 trials containing a three repetition of 13 word tokens of varying vowel duration of the same based token. One participant's data were eliminated due to a technical error, and the remaining data from 15 participants were analyzed.

6.2. Results

The results of this experiment replicated the skewness towards shortening for the perceived naturalness, which can be visually observed in Figure 6. In Experiment 4, the highest naturalness score fell on the stimuli shortened to 80% of the original vowel duration (Figures 3-4). In this experiment, a similar tendency was found for the minimal pair containing a low vowel /æ/, i.e. *back/bag*. The stimuli with the greatest score for naturalness for *bag* had a vowel duration adapted to 70% of the base token, at 267 ms. Similarly, the stimulus with the greatest score for naturalness for *back* had a vowel duration shortened to 80% of the vowel duration in the original token, i.e. 197 ms. The perceived naturalness for the high vowel /i/ in the words *pig* and *pick*, however, had a somewhat different pattern. Listeners perceived the durations adapted to 40% to 60% of the base token to have the highest degree of naturalness, with gradual decrease in the perceived naturalness.

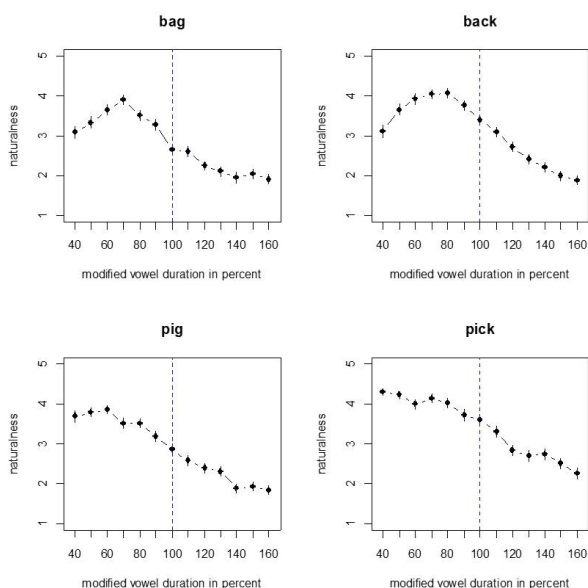


Figure 6. Perceived naturalness in two minimal pairs *bag/back* and *pig/pick* with modified vowel duration. Blue dotted line indicates the location of the original vowel duration in the continuum of vowel duration.

6.3. Discussion

In Experiment 4, vowels with the highest naturalness score were found with tokens shortened to 80% of the original duration. In that experiment, it was only the words ending in a voiced consonant that were shortened thus we did not have data showing how CVC words ending in a voiceless consonant would be perceived under shortening. Somewhat unexpectedly, the results of Experiment 5 for the *pig/pick* series show that the tokens with the greatest degree of perceived naturalness are all from the shortened stimuli regardless of whether the vowel is followed by a voiced or a voiceless consonant. On the contrary, the most natural tokens for *bag/back* involved a moderate degree of shortening but not as extreme as in the minimal pairs with a high vowel *pig/pick*.

7. General discussion and conclusion

Throughout the five experiments reported in this paper, we found an asymmetric degree of perceived naturalness in the shortened and the lengthened duration of a vowel. The shortened stimuli were consistently perceived to be more natural in all cases. As mentioned earlier, this could be due to the relatively long duration of the original vowel due to being produced in isolation. One way of testing this explanation in the future would be generating a base token with a shorter vowel duration under a faster speaking rate.

The effects of speaking rate on the perception of durational feature of segments is an area with substantial amount of research, yet this topic does not seem to have been covered. In previous research, it was found that speaking rate has no effect on the perception of VOT (Voice Onset Time; Utman, 1998). That is, listeners perceive voiceless consonants produced with longer VOTs as being better exemplars of the voiceless category regardless of the speaking rate in which they are presented. It would be interesting to see if the findings in this study might carry over when the stimuli were constructed based on a word with a shorter vowel duration produced in a faster speaking rate.

The natural, unaltered stimuli were not perceived as being very natural. If they were perceived as being prototypical, we would have expected for them a naturalness score approaching 5 in the 5-point scale, but it was only 3.62 in Experiment 1, and was consistently lower than the slightly shortened stimuli in subsequent experiments, i.e. Experiment 4 and 5. Again, this could be due to the deliberate and isolated production of the base token, which results in a relatively long vowel duration. However, it could be an effect of the other tokens that are presented to the listeners whose naturalness is inevitably gauged against each other. Several studies report the importance of the way stimuli are presented to the listeners including the presentation order. Kuhl (1991), for example, reports that a stimulus presented with a prototypical reference stimulus tends to generalize to that vowel, which is well-known as a *magnet effect*. The procedure of the current experiments does not involve detection of change in the stimuli as in Kuhl (1991) so a direct comparison is not immediately available, but it is an issue to consider for future studies. For example, a set of stimuli with vowel duration differing in increments of 30% as opposed to 10% of the current study might garner more favorable scores for the original duration than it has earned in the current paper.

The exploration into the effects of vowel duration reported in this study leaves more questions than have been answered, but might serve to inspire follow-up studies to better understand the nature of perception regarding duration. Future studies systematically controlling for the ratio between the vowel and the consonant duration as well as speaking rates and the order of presentation will be required to resolve the questions and provide a full explanation. In addition, as pointed out by an anonymous reviewer, it would be worth manipulating the stimuli based on an absolute (e.g. 10 ms.) incremental step as opposed to the percent of the base vowel duration as done here. Considering that there was a larger duration manipulation in vowels before voiced consonants, the results that shortening before voiced stops was perceived more natural is even more interesting, and it would be great to validate this effect with vowel manipulation using absolute values.

References

- Boersma, P., & Weenink, D. J. M. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.
- Denes, P. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America*, 27(4), 761-764.
- Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using mechanical turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, 5(8), 509-524.
- Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America*, 67(5), 1764-1771.
- House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1), 105-113.
- Ko, E.-S., Soderstrom, M. & Morgan, J. (2009) Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English. *Journal of the Acoustical Society of America*, 126.5, EL134-139.
- Kuhl, P. K. (1991). Human adults and human infants show a

- “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93-107.
- Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America*, 78(6), 1949-1957.
- Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, 21(3), 398-421.
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625-632.
- Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, 85(3), 172-191.
- Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America*, 51(4B), 1296-1303.
- R Core Team (2019). R: R foundation for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing, Retrieved from <https://www.R-project.org/>
- Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, 31(4), 307-314.
- Utman, J. A. (1998). Effects of local speaking rate context on the perception of voice-onset time in initial stop consonants. *The Journal of the Acoustical Society of America*, 103(3), 1640-1653.

- **Eon-Suk Ko**, Corresponding author
Assistant professor, Dept. of English Language and Literature,
Chosun University
309 Philmundaero, Donggu, Gwangju 61452, Korea
Tel: +82-62-230-6551
Email: eonsukko@chosun.ac.kr
Areas of interest: Phonetics, Child language acquisition