A Phonetic Study on Phrasing in Seoul Korean

Hae-Sung Jeon

University of Cambridge

1   Introduction

The aim of the experiments reported in this paper is to examine how duration and segmental properties among many other acoustic cues are manifested in Seoul Korean (henceforth Korean) speech and used by listeners. In particular, a pilot perception experiment was conducted to devise experimental methods for further investigation of the acoustic cues to junctures in speech.

The perception experiment was motivated by the experimenter's observation that listeners can disambiguate a pair of utterances with potential segmentation ambiguity recorded in a laboratory setting, and that they can still be disambiguated even when the intonation is eliminated. The resolution of such ambiguity is clearly related to how the utterance is divided up to make phrases.

A 'phrase' in this paper refers to any unit preceded and followed by perceivable junctures in speech, which is not necessarily defined by the F0 contour and not restricted to be larger than phonological words. Although this definition seems unclear and impressionistic, this approach was preferred in order to begin with without any presuppositions, particularly on the perceptual side.

2   Production Experiment
2.1   Experimental Materials

The two pairs of speech materials are listed below in the Yale system of romanisation, together with broad transcription, gloss, and English translation.

(1) Set A: *Hyen.sik.i.ka.sa.nun.cip.e.tu.ko.wass.tay.yo.*
a. 4+2 syllables phrasing

[hjənsi<u>giga</u>]	[<u>san</u>ɨn]	[cibe]	[tugo]	[wad'ɛjo]
[Hyensik-NOM]	[live-RL]	[house-LOC]	[leave-CONJ]	[come-PST-RT-POL]

(Somebody) left (something) in the house where Hyensik lives.

b. 3+3 syllables phrasing

[hjənsi<u>gi</u>]	[<u>kas</u>anɨn]	[cibe]	[tugo]	[wad'ɛjo]
[Hyensik]	[lyrics-TOP]	[house-LOC]	[leave-CONJ]	[come-PST-RT-POL]

Hyensik left the lyrics at home.

(2) Set B: *A.ce.ssi.to.kki.ka.iss.ess.u.myen.coh.keyss.tay.*
a. 4+2 syllables phrasing

[acəs'ido]	[k'iga]	[is'əs'ɨmjən]	[cokʰed'ɛ]
[uncle-INP]	[gut-NOM]	[be-PST-CONJ]	[like-PST-RT]

The uncle wishes he also had a gut.

b. 3+3 syllables phrasing

[acəs'i]	[tok'iga]	[is'əs'ɨmjən]	[cokʰed'ɛ]
[uncle]	[axe-NOM]	[be-PST-CONJ]	[like-PST-RT]

The uncle wishes he had an axe.

The meaning of the utterances within a pair differs depending on whether the first 6 syllables are parsed into two phrases with 4+2 syllables or with 3+3 syllables. In Set A, in (1a), *ka* in the first phrase functions as a nominative case particle, whereas in (1b), *ka* is the first syllable of a noun, *kasa* (*lyrics*). (1b) is an instance where the nominative case particle *ka* or *nun* is absent after the subject, *Hyensiki*. That is, the unambiguous equivalent of (1b) would be *Hyensiki-ka(/nun) kasanun*.

The same principle applies to the utterances in Set B. In (2a), *to* is a particle meaning *also*, whereas *to* in (2b) is the first syllable of a noun, *tokki* (*axe)*. The unambiguous version of this utterance is *Acessi-to(/nun) tokkika.*

The lenis stops are transcribed as voiced segments (e.g., [g], [d]) within a word and as voiceless segments word-initially.[1] In the experimental materials, lenis stops were placed on the fourth syllable of the sentences for the analysis of their lenition process depending on their location.

2.2 Procedures

Six native speakers of Korean (4 females, 2 males, aged between 19 and 34) were recorded. They were asked to read 5 pairs of randomly arranged speaking materials at three self-selected rates (slow, normal, fast), and then to repeat the whole procedure. They also read a few nonsense utterances for a different study, and the analyses of only two pairs of the recorded utterances are discussed in this paper.

Speakers were given enough time to become familiar with the experimental materials, and they were allowed to read a whole utterance again when they made mistakes. The experimenter did not give further instructions, and did not attempt to elicit any specific intonational contours.

Participants' speech was recorded in a sound attenuated booth in the Phonetics Laboratory, University of Cambridge. The sampling rate was 41kHz.

2.3 Acoustic Analysis

The focus of the analysis was on the three target syllables around the potential

phrase junctures only (i.e., 3rd-5th syllables in the speech materials, underlined in Section 2.1). Although the duration of each segment and F0 contours were also measured with Praat ver. 5.0.06, the results are not discussed here.

The experimenter listened to each utterance to ascertain that the recorded utterances did not sound ambiguous. All speakers clearly disambiguated the two utterances in a pair in all speaking rates.

The realisations of the underlyingly voiceless lenis stops were classified into 8 categories as 1) voiceless stop, 2) fricated stop, 3) fully voiced stop, 4) aspirated fricative, 5) fricative, 6) fully voiced fricative, 7) approximant, or 8) deleted segment. The lenition hierarchy is based on the presupposition of the strength hierarchy: stop > fricative > approximant. The subcategories of stop, fricative, approximant were established by the degree of voicing: the more voicing, the weaker the consonant.

If any voiceless interval during the stop closure was observed in the inspection of waveforms and spectrogram, the allophones were classified as voiceless stops. This implies that some of the allophones classified into voiceless stops are actually partially voiced. There were 72 utterances in each pair (2 sentences × 3 rates × 6 subjects × 2 repetitions).[2]

2.4 Results

Speaking rate Speakers clearly produced speech in three distinctive speaking rates. There was a significant main effect of SPEAKING RATE in a one-way repeated measures ANOVA [$F(2, 10) = 75.16$, $p < 0.001$ for Set A; $F(2, 10) = 93.05$, $p < 0.001$ for Set B]. *Post hoc* tests (dependent *t*-tests, significance level = 0.05) revealed the significant differences between Slow vs. Normal [$t(5) = -5.75$, $p < 0.005$ for Set A; $t(5) = -7.24$, $p < 0.005$ for Set B] and also between Normal vs. Fast [$t(5) = -9.03$, $p < 0.001$ for Set A; $t(5) = -6.15$, $p < 0.005$ for Set B]. The means and standard deviations of speaking rates are summarised in Table 1.

Rate	Set A	Set B
Slow	4.53 (0.28)	3.99 (0.26)
Normal	6.64 (0.81)	6.29 (0.64)
Fast	9.10 (1.08)	7.96 (0.49)

Table 1. Means (standard deviations) of local speaking rates by each set (*syls/sec*). Local speaking rate was calculated from the duration of the three target syllables around the potential phrase juncture.

Allophonic variations of lenis stops The allophonic distribution of two lenis consonants, *k* and *t*, placed as the onset of the fourth syllables in the speech materials, is presented in this section. As shown in Table 2, over all speaking rates, higher percentage of the *k*s in Set A are realised as stops (e.g., voiceless

stop, fricated stop, or fully voiced stop) in word-initial position, (i.e., [hjənsigi] [*k*asanɨn]) compared to word-medial position (i.e., [hjənsigi*g*a] [sanɨn]). On the other hand, word-medial *k*s show higher percentage of extremely lenited forms (i.e., approximant and deleted segment).

	Strong		Intermediate		Weak		Total	
	[g]	[k]	[g]	[k]	[g]	[k]	[g]	[k]
Slow	10 (83.3)	12 (100)	2 (16.7)	0	0	0	12 (100)	12 (100)
Normal	4 (33.3)	8 (66.7)	4 (33.3)	4 (33.3)	4 (33.3)	0	12 (100)	12 (100)
Fast	1 (8.3)	3 (25.0)	4 (33.3)	7 (58.3)	7 (58.3)	2 (16.7)	12 (100)	12 (100)

Table 2. Count of *k* ([k], [g]) realisations and frequency in percentages (%) within a rate and utterance type. [g] is from [hjənsigi*g*a] [sanɨn] and [k] is from [hjənsigi] [*k*asanin]. Lenition categories were grouped into strong (i.e., voiceless stop, fricated stop, and fully voiced stop), intermediate (fricative, aspirated fricative, and voiced fricative), or weak (approximant and deleted segment) allophones.

	Voiceless Stop		Fully Voiced Stop		Total	
	[d]	[t]	[d]	[t]	[d]	[t]
Slow	7 (63.6)	11 (100)	4 (36.4)	0	11 (100)	11 (100)
Normal	4 (33.3)	9 (75.0)	8 (66.7)	3 (25.0)	12 (100)	12 (100)
Fast	6 (50)	5 (41.7)	6 (50.0)	7 (58.3)	12 (100)	12 (100)

Table 3. Frequency of *t* ([t], [d]) realisations and frequency in percentages (%) within a rate and utterance type. [d] is from [acəs'i*d*o] [k'iga] and [t] is from [acəs'i] [*t*ok'iga].

The realisations of the alveolar stops are not as varied as those of the velar stops; the only allophones of *t* attested in the speech data are a voiceless stop or a fully voiced stop, as shown in Table 3. This should not be generalised, given that the results are from only two types of utterances without any variations of the segmental contexts. However, that the word-medial position is a weak position for consonants is also shown by the alveolar stop. In Table 3, a higher percentage of instances of word-initial *t*s are produced as voiceless stops which is the stronger form in this case, at slow and normal rates.

In addition, Table 2 and Table 3 show that as speech rate increases, the underlying lenis stops tend to have more lenited forms. In each column of the tables, the percentage of strong allophones tends to increase as speaking rate decreases, whereas the percentage of weak allophones shows the reverse pattern, although these patterns are less clear with the word-medial [d] in Table 3.

3   Perception Experiment

It has been reported that word segmentation ambiguity is resolved easily both by

children aged 3 or 4 years and adults who are native Korean speakers, with naturally produced utterances (Choi & Mazuka, 2003). The experiment discussed here investigates *how*, since the phonetic details of what makes the disambiguation possible have not been comprehensively studied.

Although much of the acoustic information is present in speech signal, the experiment focused on the segmental level of information (i.e., the properties of the potential word-initial consonant) and durational cues (i.e., the domain-initial consonantal lengthening and the domain-final vowel lengthening, see Cho & Keating 2001, and Chung, 2002 for a review).

3.1 Creation of Experimental Stimuli
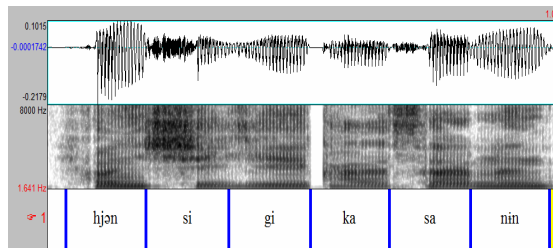
Figure 1. Spectrogram of the default stimulus used for the Canonical condition in Set A.
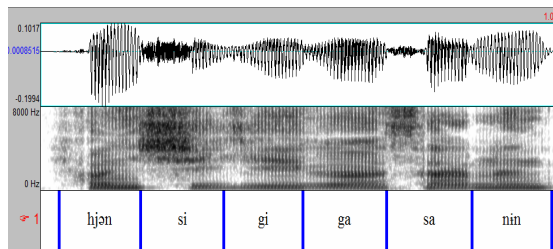

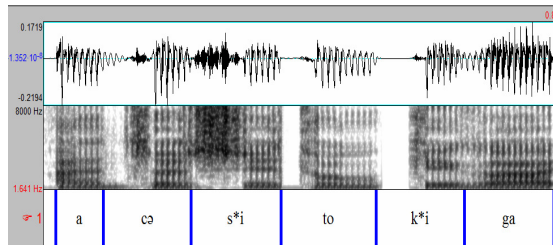**Figure 2. Spectrogram of the default stimulus used for the Lenited condition in Set A.**

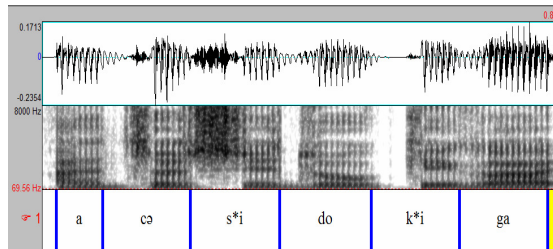**Figure 3. Spectrogram of the default stimulus used for the Canonical condition in Set B.**

Figure 4. Spectrogram of the default stimulus used for the Lenited condition in Set B.

Two default stimuli with only the first half of each utterance were created from recordings of the two male speakers for each experimental set (Set A: *Hyen.sik.i.ka.sa.nun*; Set B: *A.ce.ssi.to.kki.ka*). One utterance from each pair spoken at the normal rate was selected to create the base utterances for the resynthesis of the experimental stimuli.

There were two experimental conditions regarding the segmental property, a canonical potential-phrase-initial segment and a lenited potential-phrase-medial segment (Canonical and Lenited henceforth). In the Canonical, the strong allophone (e.g., voiceless stop) selected from the same speaker's speech was embedded in the base utterance with the following vowel (e.g., [ka] in Set A), and in the Lenited, the weak allophone (e.g., an approximant in Set A) was embedded with the following vowel (e.g., [ga] in Set A), as illustrated in Figure 1-4. A fully voiced stop, rather than an approximant, was used in Set B to reflect the realisation of the alveolar stop in the production data. It is important to note that not just the consonant but the whole syllable was replaced in creating the experimental stimuli, although the experimental conditions were named as the Canonical and the Lenited, depending on the type of the embedded consonants in the experimental stimuli. This was in order to create more natural stimuli, since the properties of the vowels following different types of consonants are not the same. For example, the vowel following a voiceless stop with a stop burst and aspiration period has breathy quality, whereas voicing goes through the CV syllable when the consonant (C) is a fully voiced approximant.

All default stimuli were resynthesised so that all the CV syllables had the same duration and were monotonous with Praat ver.5.0.06, using PSOLA method. The mean syllable duration and F0 values for the default stimuli were intended to be as close as possible to the mean values of the original speech data for each speaker (161.29ms and 149.86Hz in the original speech for Set A; 132.29ms and 118.08Hz in Set B). The intended F0 of Set A was 150Hz, and set B was 120Hz. Although the intrinsic duration of segments (i.e., fricatives are longer than stops) was not taken into account, Korean listeners on whom the experimenter informally tested the experimental stimuli did not comment that the speech timing was unusual. Although the monotonous experimental stimuli did not sound like natural utterances, the resynthesising process did not affect

6

intelligibility.

		Syllable Duration (ms)	F0 (Hz)	Intensity (dB)
Set A	Canonical	158.06 (2.41)	150.03 (0.97)	61.17 (2.81)
	Lenited	157.45 (5.00)	150.06 (0.10)	61.19 (2.77)
Set B	Canonical	118.48 (25.90)	119.86 (1.24)	64.85 (1.65)
	Lenited	117.80 (25.70)	120.13 (0.24)	64.83 (1.60)

Table 4. Means (standard deviations) of syllable duration, F0, and intensity for the default stimuli.

For duration manipulation, each segment in the three target syllables (e.g., *k, i, k, a, s, a* in Set A) and two VC sequences around the potential phrase boundary (e.g., *ik* and *as* in Set A) were lengthened by 20%, 40%, 60%, 80%, and 100% independently. Therefore, there was either one lengthened segment or one lengthened VC in a listening stimulus, and the duration of the rest of the utterance was kept the same as the default stimuli. Segments and VC sequences were lengthened by five degrees to determine the point which listeners begin to show certain reaction patterns and to provide a reference point for further experiments with more types of acoustic cues. For each utterance pair, there were 41 stimuli (i.e., 1 default stimulus + 6 segments lengthened in 5 degrees + 2 VC sequences lengthened in 5 degrees) in both the Canonical and the Lenited conditions.

The average values of the measured acoustic parameters of the default stimuli are summarised in Table 4. Since the listening materials were resynthesised from real speech, the variables could not achieve the intended values exactly.

In the given experimental conditions, it was hypothesised that 1) if a vowel were lengthened, listeners would interpret it as a pre-junctural vowel, 2) if a consonant were lengthened, listeners would interpret it as a post-junctural consonant, and 3) if there were no duration cue, listeners would interpret the syllable with a strong consonant (i.e., voiceless stops in the experiment) as a post-junctural syllable onset, and the syllable with a weak consonant (i.e., voiced stops or approximants in the experiment) as a pre-junctural syllable onset.

3.2   Experimental Procedure

Seven people who are native speakers of Korean (5 male, 2 female, aged between 27 and 32) without hearing problems participated in the experiment which was run with Praat ver.5.0.06.

In Set A and Set B, each stimulus occurred randomly three times, with the proviso that the same stimulus was not allowed to occur twice in succession. Therefore, 246 randomly arranged stimuli were played in each set (41 stimuli × 2 conditions × 3 repetitions). The presentation order of Set A and Set B was counterbalanced, and the main experiments were preceded by practice sessions

with 24 stimuli which were created from different speakers' data.

After each sound was played through the headphones, participants were asked to choose whether the utterance should be phrased into 4+2 syllables or 3+3 syllables on the PC screen. The answers for the forced-choice questions were presented in the romanised transcription.

3.3 Results

Listeners' response rates for the 3+3 syllable phrasing in percentages of the total responses are presented in Figure 5 - Figure 10, and they were the dependent variables in all statistical analyses. Firstly, three-way repeated measures ANOVAs were performed using the SEGMENTAL LENITION (Canonical vs. Lenited), the SEGMENTAL IDENTITY (*k, i, k, a, s, a* in Set A; *ss, i, t, o, kk, i* in Set B), and the LENGTHENING (0, 20, 40, 60, 80, 100%) in each set. The lengthening of the VC sequences (*ik, as* in Set A; *it, okk* in Set B) was analysed separately. The significance level of all the statistical tests was set to 0.05, and only significant results are reported.

In Set A, only the SEGMENTAL IDENTITY led to a significant effect [$F(5, 30)$ = 5.89, $p < 0.005$]. There were no other significant main effects or interactions. Moreover, when exploring the main effect of the SEGMENTAL IDENTITY in more detail by comparing the results for each segment with every other segment using Sidak *post hoc* tests, no significant results occurred. The significant main effect of the SEGMENTAL IDENTITY seems to be caused by the *(k)i* lengthening which elicited the highest percentage of the 3+3 syllables phrasing responses (M = 54.36, SE = 6.18). This is supported by the significant difference between *(k)i* lengthening and *(k)a* lengthening (with the lowest 3+3 syllables phrasing response rate, $M = 36.10$, $SE = 3.69$) [dependent *t*-test, $t(6) = 2.91$, $p < 0.05$], but the non-significant result of the *k(i)* (the second highest 3+3 syllables phrasing response rate, $M = 43.65$, $SE = 3.41$) vs. *(k)a* comparison [$t(6) = 1.93$, *ns*]. With lengthening of the VC sequences, no statistically significant main effects or interactions were observed.
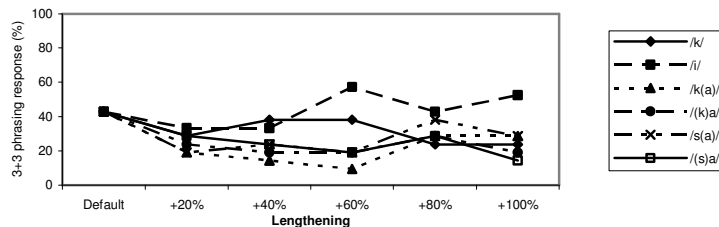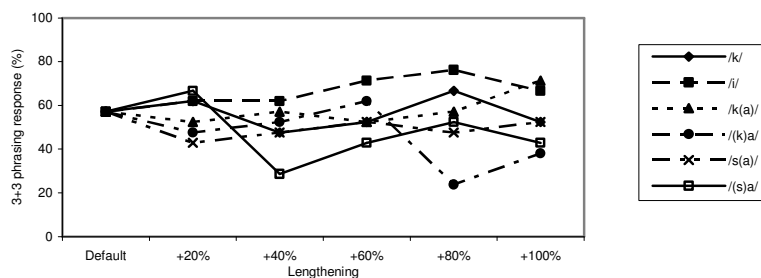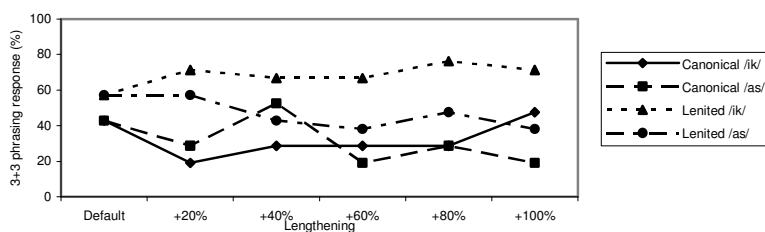

Figure 5. 3+3 syllables phrasing response (%) to the segmental lengthening for the Canonical condition, Set A. X-axis, Y-axis, and each line show the degrees of lengthening, 3+3 response, and lengthened segment, respectively.
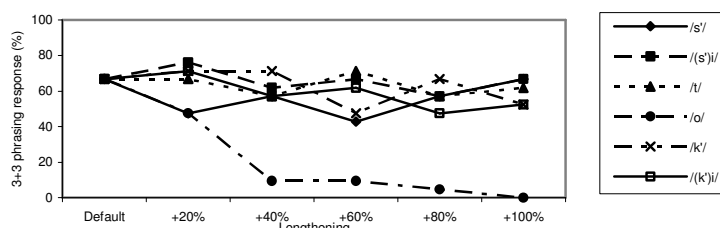
Figure 6. 3+3 syllables phrasing response (%) to the segmental lengthening for the Lenited condition, Set A. X-axis, Y-axis, and each line show the degrees of lengthening, 3+3 response, and lengthened segment, respectively.

Figure 7. 3+3 syllables phrasing response (%) to the VC lengthening, Set A. X-axis, Y-axis, and each line show the degrees of lengthening, 3+3 response, and lengthened VC, respectively.

Figure 8. 3+3 syllables phrasing response (%) to the segmental lengthening for the Canonical condition, Set B. X-axis, Y-axis, and each line show the degrees of lengthening, 3+3 response, and lengthened segment, respectively.
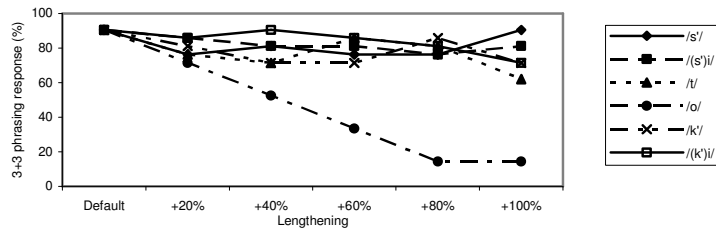
Figure 9. 3+3 syllables phrasing response (%) to the segmental lengthening for the Lenited condition, Set B. X-axis, Y-axis, and each line show the degrees of lengthening, 3+3 response, and lengthened segment, respectively.
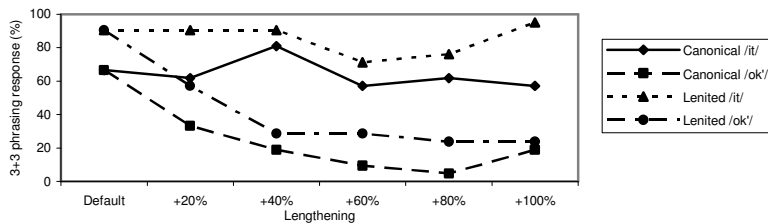

Figure 10. 3+3 syllables phrasing response (%) to the VC lengthening, Set B. X-axis, Y-axis, and each line show the degrees of lengthening, 3+3 response, and lengthened VC, respectively.

By contrast, for Set B, the main effects of the SEGMENTAL LENITION [$F(1, 6) = 10.70$, $p < 0.005$], the SEGMENTAL IDENTITY [$F(5, 30) = 19.90$, $p < 0.001$], and the LENGTHENING [$F(5, 30)=6.02$, $p < 0.005$] were statistically significant. Also, the interaction of SEGMENTAL IDENTITY × LENGTHENING [$F(25, 150) = 4.15$, $p < 0.001$] was significant, although no significant interaction effects were found in any other combinations of the independent variables. Results with the VC lengthening showed a similar picture to the segmental lengthening; there were significant main effects of the SEGMENTAL LENITION [$F(1, 6) = 15.82$, $p < 0.01$], the VC IDENTITY [$F(1, 6) = 44.29$, $p < 0.005$], the LENGTHENING [$F(5, 30) = 6.82$, $p < 0.001$], and a significant interaction of VC IDENTITY × LENGTHENING [$F(5, 30) = 4.44$, $p < 0.005$].

Two interesting points in Set B are 1) the main effect of the SEGMENTAL LENITION, 2) the significant interactions of SEGMENTAL IDENTITY × LENGTHENING, and also VC IDENTITY × LENGTHENING. Firstly, *post hoc* tests with Sidak corrections revealed that, surprisingly, the effect of the SEGMENTAL LENITION is the reverse of the prediction – listeners preferred 4+2 syllables phrasing in the Canonical condition (for 3+3 syllables phrasing, $M = 55.16$, $SE = 8.35$, with segmental lengthening; $M = 44.84$, $SE = 5.74$ with VC lengthening) compared to the Lenited condition ($M = 75.14$, $SE = 4.52$ with segmental

10

lengthening; $M = 64.29$, $SE = 5.27$ with VC lengthening, $p < 0.01$ for both segmental lengthening and VC lengthening).

Lengthening	20%	40%	60%	80%	100%
df			6		
t	-1.43	-3.39	-3.83	-8.56	-8.36
p	ns	< 0.05	< 0.05	< 0.001	< 0.001

Table 5. *t*-statistics table for each level of lengthening (*o* lengthening vs. lengthening of other segments)

Secondly, the interactions between the SEGMENTAL/VC IDENTITY and the LENGTHENING suggest that the lengthening of different parts of the utterances affected the listeners' response differently, as hypothesised. It seems listeners only responded differently to the *o* lengthening in Set B (Figure 8 - 9), and this is supported by further statistical analyses. In another two-way repeated measures ANOVA with the SEGMENTAL IDENTITY excluding *o* (*ss, i, t, kk, i* in Set B) and the LENGTHENING, no significant effects were observed; however, a one-way repeated measures ANOVA with LENGTHENING only in the *o* segment showed a significant effect [$F(5, 30) = 16.89$, $p < 0.001$]. Next, since listeners' responses to one segment, *o*, differed from all the others, data were collapsed so that there were two levels of the segmental identity (*o* lengthening vs. lengthening of all the other segments), and dependent *t*-tests were performed at each of the five levels of lengthening (20-100%) with Bonferroni corrections. In the *t*-tests, it was revealed that listeners chose significantly more 4+2 syllable phrasing when the vowel *o* was lengthened by more than 40% compared to when all other segments are lengthened (Table 5).

Lengthening	20%	40%	60%	80%	100%
df			6		
t	2.77	6.59	3.42	4.83	3.09
p	< 0.05	< 0.005	< 0.05	< 0.005	< 0.05

Table 6. *t*-statistics table for each level of lengthening (*it* lengthening vs. *okk* lengthening)

Similar patterns were revealed in the analyses of VC lengthening. *Post hoc* tests with Sidak corrections showed that lengthening of *it* ($M = 75.01$, $SE = 5.72$) elicited significantly more 3+3 syllable phrasing responses than *okk* lengthening ($M = 34.12$, $SE = 5.90$, $p < 0.005$), whereas no significant results were observed with the LENGTHENING. When data were collapsed so that there were two independent variables (*it* lengthening vs. *okk* lengthening), dependent *t*-tests with Bonferroni corrections performed at each of the five levels of lengthening (20-100%) showed significant differences between the *it* and *okk* lengthening at all levels (Table 6).

4 General Discussion

The main findings of the production experiment are 1) post-word junctural underlying lenis stops tend to have stronger allophones compared to those in word-medial position, and 2) consonantal lenition is affected by speaking rate – speaking rate and the degree of lenition seem to be positively correlated.

This suggests that the production (and possibly the detection) of speech junctural phenomenon at least on the level of segments or syllables may be highly context-sensitive. That is, speakers may not necessarily produce speech segments with concrete phonological features, and what helps listeners to segment speech is probably the relative strength of segments or syllables within certain domains, such as prosodic phrases or utterances.

In the perception experiment, although it was predicted that the Canonical condition would elicit more 3+3 syllable phrasing responses, the results show the opposite finding than was predicted. Listeners show a greater relative preference for the 4+2 syllable phrasing in the Canonical condition, as opposed to in the Lenited condition.

One possible explanation of the unexpected result is that the utterances were actually parsed into three phrases of 3+1+2 syllables by listeners in the Canonical condition, with an emphasis on the potentially monosyllabic particle, *ka* and *to*, in the experimental materials. When choosing the phrasing pattern in the experiment, listeners then seemed to choose to group these particles with the subject, due to the syntactic constraint that the nominative case marker or the inclusion particle should be together with the preceding subject. Two participants actually commented that some part of the utterances sounded emphasised after the experiment.

To elaborate, in the case of the 4+2 syllable phrasing (i.e., *Hyensikika sanun* in Set A and *Acessito kkika* in Set B), the first three syllables should be interpreted as subjects, the fourth syllable as a nominative case particle or an inclusion particle (meaning *also*) respectively. In the 3+3 syllable phrasing (i.e., *Hyensiki kasanun* in Set A and *Acessi tokkika* in Set B), however, the first three syllables are subjects, the following two syllables are objects, and the last syllables are interpreted as particles (see Section 2.1).

It seems that the monosyllabic particles can be phonetically focused when they are emphasised or when the preceding constituents are emphasised. It is therefore likely that listeners treated the strong syllables containing the canonical onsets as indicators of emphasis. This point is noteworthy regarding the status of the function words in different languages; in languages like English, function words are often phonetically reduced. However, considering that such function words are frequently absent or contracted in Korean colloquial speech, the addition of these function words in utterances may inherently result in emphasis.[3] Moreover, the two particles used in the experiment, *ka* and *to,* may not have the

same semantic weight – one of them may be more likely to be perceived as focused in the experimental setting where similar listening stimuli are repeated, and this might be the reason why more statistically significant results were found with Set B. It is important in any case to note that the morphological structure of the experimental stimuli can affect listeners' judgements.

When the duration of different parts of the utterances was manipulated in the experiment, listeners seemed to take the lengthened vowel as a marker of phrase-finality. As the statistical analyses and the listeners' response changes indicate, the response rate of *Acessito kkika* significantly increases with the lengthening of the vowel *o*. Although statistically significant changes in listeners' response were not found in Set A, the general patterns in Figure 5 to 10 show the divergence between lengthening of the two different vowels in the third and fourth syllables.

Consonantal lengthening, however, does not seem to affect the response patterns of listeners. In the resynthesis procedure, the duration of consonants (i.e., the closure duration for stops and frication duration for fricatives) was increased by 20% to 100%, with the same ratio for vowel lengthening. Unlike vowel lengthening, even 100% consonantal lengthening led to an increase of less than 25ms for the lenis stop closure, and less than 40ms for the fortis stop closure, and less than 50ms for the fricative. However, considering that the durational contrast between the default stimuli and the stimuli with 100% consonantal lengthening were clearly audible, it seems that listeners are not as sensitive to consonantal lengthening as they are to vowel lengthening and the changes in consonant allophones.

The response percentage lines for the two VC sequences around the potential phrase juncture display similar patterns to those of the lengthening of the relevant vowel (which is in the VC sequence). However, despite the similar patterns, the effects of lengthening seem to be more notable with the vowels. It may be because the actual vowel duration to which listeners seem sensitive to was shorter in the VC sequence lengthening conditions.

5   Conclusion

In the production experiments, where two pairs of utterances with different phrase juncture locations were examined, it was shown that word-initial position is a stronger position for lenis consonants than word-medial position, and the lenition process seems to be influenced by speaking rate variations.

The simple perception experiments where listeners were asked to divide up speech without intonational cues showed that information on the segmental or the syllable level can affect listeners' identification of speech junctures, and that vowel lengthening is interpreted as a pre-junctural phenomenon by Korean listeners.

In addition, it is important to consider the relationship between the

morphological and semantic structure of an utterance and the prosody, which posed a problem in the perception experiment. This language-specific morphological and semantic effect should certainly taken into account by researchers investigating prosody when designing experiments and analysing results. Further research is needed to clarify the relationship between morphology, the semantic weight of the morphological constituents and prosodic manifestation, and how they contribute to speech communication.

Notes

1. Korean intervocalic lenis stops are usually phonetically voiceless when initial in the Accentual Phrase, and voiced when medial. The degrees of voicing within the stop closure are influenced by a number of factors, such as speaking rate and segmental contexts (see Jun, 1996 and references therein).

2. Two utterances in slow rate were excluded due to speakers' mistakes.

3. For example, when a monosyllabic particle is attached to a stem with an open final syllable, components other than the syllable coda of the particle may not be realised in casual speech (e.g., [hak'jo] (*school*) + [lɨl] (ACC) → [hak'jo lɨl] → [hak'jol]).

References

Cho, Taehong., & Keating, Patricia. A. 2001. Articulatory and Acoustic Studies on Domain-initial Strengthening in Korean. *Journal of Phonetics* 29. 155-190.

Choi, Youngon. & Mazuka, Reiko. 2003. Young Children's Use of Prosody in Sentence Parsing. *Journal of Psycholinguistic Research* 32. 197- 217.

Chung, Hyungsong. 2002. *Analysis of the Timing of Spoken Korean with Application to Speech Synthesis.* Unpublished PhD thesis. UCL.

Jun, Sun-Ah. 1996. *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure.* Garland Publishing Inc., New York: NY.

Phonetics Laboratory, Department of Linguistics
University of Cambridge, Sidgwick Avenue
Cambridge, CB3 9DA, UK
e-mail:hsj24@cam.ac.uk