CONSONANTAL AND PROSODIC INFLUENCES ON KOREAN VOWEL DURATION

*Hyunsong Chung**, *Kyongsok Kim***, *Mark Huckvale**
* Department of Phonetics and Linguistics, University College London
Gower Street, London WC1E 6BT, United Kingdom
** Department of Computer Science, Busan National University
Busan 609-735, South Korea
e-mail: * {hchung, mark}@phon.ucl.ac.uk, ** gimgs@asadal.cs.pusan.ac.kr

ABSTRACT

In this paper, we describe some consonantal and prosodic influences on vowel duration in Korean. In particular, the effect of different obstruents on the duration of following vowels was examined. With a large set of artificial utterances recorded by a single speaker, we studied four factors on vowel duration: the pre-vocalic context, the post-vocalic context, clause-final lengthening and phrase-internal shortening. The results show that tense aspirated obstruents shorten the following vowels most, followed by lax obstruents and tense unaspirated obstruents. The influence of obstruents on preceding vowels is negligible. Phrase-initial vowels are longer than any other phrase-internal ones, while clause-final vowels are longest. We have applied these findings in a Korean speech synthesis system being developed at UCL.

Keywords: Korean, prosody, duration, synthesis

1. INTRODUCTION

Contemporary Korean language synthesis systems such as Hansori (Korea Telecom) or Keulsori (Korean Electronics and Telecommunications Research Institute) [1] make use of the same concatenative signal generation methods used in the major U.S., European and Japanese systems. Such methods readily provide good segmental quality and shift the main research questions in speech synthesis onto the prediction of prosody. But whereas the prosody of the main western languages and Japanese has been extensively studied, the Korean language has been the subject of few formal analyses relevant for speech synthesis.

This paper is concerned with the prediction of vowel durations from their segmental and prosodic contexts. Whereas previous work in this area has simply taken into account the location of the syllables within a phrase, this work also explores influence of pre-vocalic and post-vocalic consonantal contexts. In particular, the effect of different obstruents on the duration of following vowels was examined. Contrary to Peterson and Lehiste's study of English [2] that the effect of syllable-initial consonants on the duration of the following vowel is negligible, we have found that Korean obstruents have considerable effect.

The durational factors we have found are being incorporated in a Korean speech synthesis system which is being developed at UCL based on the ProSynth architecture [3] and the MBROLA system [4]. In this paper, the inherent duration and compression factors of vowels are given. The corresponding values for consonants are not given because we believe that vowel shortening and lengthening play the most important role in describing syllable compression. The overall computational framework parallels the top-down computation of the timing of syllabic components proposed by Local and Ogden [5] and implemented in the YorkTalk system for English.

2. PROCEDURE

2.1 Database

For the study reported, 384 artificial utterances were designed and recorded by a single speaker. The utterances systematically explored both syllable position and syllable composition within a sentence frame containing nonsense monosyllable pairs. For example: /ikʌsɯn V | V soɾita/ was used to investigate the inherent vowel duration; /ikʌsi CV(C) | CV(C) soɾita/ for consonantal influences on vowel duration; /ikʌsi CV | CVCVCVCV/ for prosodic influences on vowel duration. Four durational factors on vowels were explored: the pre-vocalic context, the post-vocalic context, clause-final lengthening and phrase-internal shortening. The recordings were made twice in an anechoic chamber on digital tape using 2 channels at 44,100 samples/sec/channel. Channel 1 was speech signal from microphone, channel 2 was Laryngograph signal. They were resampled to 16 kHz and transferred to disk. In order to make the speaker keep a consistent rate of speech, we used a prompting tool when recording. Sentences were displayed on a monitor screen at five second intervals so that the speaker could read each sentence with a regular rhythm.

Vowel	a	i	u	e	o
DUR	207	160	165	169	186
Vowel	ɛ	ʌ	ɯ	ja	jɛ
DUR	201	144	144	184	210
Vowel	jʌ	je	wa	wɛ	we
DUR	132	215	179	193	165
Vowel	jo	wʌ	we	wi	ju
DUR	176	144	169	156	208
Vowel	ɰi				
DUR	170				

Table 1. Inherent duration of vowels. (msec)

2.2 Annotation

We used the Speech Filing System (SFS) [6] to analyze and annotate the speech data. The segmentation was decided with the reference to three signals: waveform, spectrogram, and Laryngograph signal (Lx). In deciding the boundary between obstruents and vowels, we gave the aspiration portion to the consonants, following Mieko Han [7] and other Korean researchers. They argue that there is a compensatory adjustment in duration between the aspiration of obstruents and the vowels following them. This paper systematically explores this pattern and expresses the extent of the shortening effect of the obstruents on vowels.

2.3 Compression factors

First of all, we chose an inherent duration for each vowel. In this experiment, to reduce the effects of irrelevant variables to the minimum, we assumed that the phrase-initial vowels have the inherent duration. We measured the duration of the vowel in a phrase-initial open syllable with no onset to get this value. To represent the shortening effect of consonants on the vowel, we divided the measured duration of the relevant vowels by the inherent duration. Thus the inherent duration then has a compression factor of 1, while a vowel lengthened in context will have a factor bigger than 1.

3. INHERENT DURATION OF VOWELS

While other researchers have argued that a vowel in citation form has the inherent duration, we avoided this view. In the citation form, the vowel occupies both phrase-final and clause-final position which has more variables than the phrase-initial position. As shown in Table 1, among the short vowels, the low vowel /a/ has the longest duration in Korean. The mid vowels /o, e, ɛ/ follow /a/ in the durational values. The high vowels /i/ and /u/ are shorter than mid and low vowels. Vowels such as /ʌ/ and /ɯ/ have the smallest value. Among diphthongs, we could not find any general pattern. Further investigations are necessary to find the pattern for Korean diphthongs. These findings are not so different from those reported by Mieko Han's [7] and Kim's [8].

Possible consonants in onset position
p', p, p^h, t', t, t^h, k', k, k^h, tɕ', tɕ, $tɕ^h$, s', s, h, n, ɾ, m
Possible consonants in coda position
p, t, k, n, l, m, ŋ

Table 2. Consonants in Korean.

	a	i	u	e	o	AVE
p'	0.72	0.73	0.81	0.71	0.75	0.74
t'	0.72	0.83	0.68	0.87	0.67	0.75
k'	0.64	0.77	0.62	0.86	0.67	0.71
tɕ'	0.76	0.68	0.67	0.77	0.65	0.71
s'	0.65	0.75	0.63	0.77	0.62	0.68
AVE	0.70	0.75	0.68	0.80	0.67	0.72
p	0.68	0.62	0.56	0.84	0.71	0.68
t	0.63	0.67	0.66	0.75	0.62	0.67
k	0.58	0.49	0.54	0.69	0.49	0.56
tɕ	0.65	0.53	0.55	0.63	0.67	0.61
s	0.51	0.51	0.54	0.60	0.54	0.54
AVE	0.61	0.60	0.59	0.72	0.62	0.63
p^h	0.54	0.57	0.47	0.70	0.59	0.57
t^h	0.55	0.52	0.54	0.64	0.46	0.54
k^h	0.40	0.43	0.40	0.53	0.37	0.43
$tɕ^h$	0.45	0.53	0.54	0.54	0.46	0.50
h	0.52	0.60	0.57	0.72	0.59	0.60
AVE	0.49	0.53	0.50	0.63	0.49	0.53

Table 3. Compression factors caused by obstruent influence on the following vowel.

4. CONSONANTAL INFLUENCES ON VOWEL DURATION

4.1 Pre-vocalic obstruent influences

As shown in Table 2, 18 consonants are available in the onset position in Korean. Amongst the Korean stops in prevocalic position, tense aspirated stops shorten the vowel duration most, followed by lax stops and tense unaspirated stops. Table 3 shows that the tense aspirated stops shorten the following vowels most with the compression factors between 0.43 and 0.57. Tense unaspirated stops have the least effect on vowel duration compared to other stops. Average factors are between 0.71 and 0.74. This pattern also applies to the affricates and fricatives in Korean. Amongst the affricates, the tense aspirated affricate /$tɕ^h$/ has the greatest shortening effect, followed by the lax affricate /tɕ/ and the tense unaspirated affricate /tɕ'/. Individual factors are 0.50, 0.61, and 0.71. With fricatives, the tense unaspirated fricative /s'/ has a less shortening effect than the lax one. All other things being equal, prevocalic fricative /s/ shortens vowels more than other unaspirated obstruents. The compression factors for these two fricatives are 0.68 and 0.54. The fricative /h/ has an average factor of 0.60, which is between unaspirated stops and /s/ fricative.

	a	i	u	e	o	AVE
p	0.24	0.59	0.63	0.40	0.51	0.47
t	0.33	0.76	0.49	0.51	0.44	0.51
k	0.38	0.72	0.59	0.34	0.53	0.51
AVE	0.32	0.69	0.57	0.42	0.49	0.50
	ta	ti	tu	te	to	AVE
p	0.37	0.45	0.39	0.48	0.43	0.42
t	0.43	0.47	0.46	0.44	0.45	0.45
k	0.33	0.38	0.41	0.50	0.42	0.41
AVE	0.38	0.43	0.42	0.47	0.43	0.43

Table 4. Compression factors caused by obstruent influence on the preceding vowel.

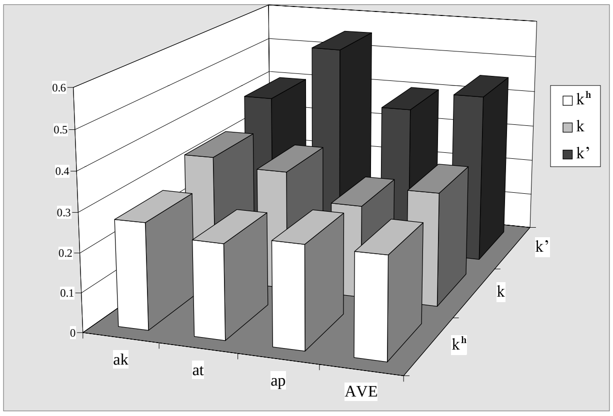

Figure 1. Compression factors caused by pre-vocalic and post-vocalic influences on the vowel: /ak/, /at/, /ap/ are rhymes; /k'/, /k/, /k^h/ are onsets.

When place is taken into account, the bilabial stops and the alveolar stops have the least shortening effect. Amongst the lax stops, bilabial stop /p/ and alveolar stop /t/ influence less on the following vowel duration than the velar stop /k/. The average compression factors of lax stops are 0.68, 0.67, and 0.56. This pattern also applies to the tense aspirated stops. Amongst the tense unaspirated stops, the bilabial stop /p'/ and the alveolar stop /t'/ have less shortening effect than the velar stop /k'/. The factors are 0.74, 0.75, and 0.71.

4.2 Post-vocalic obstruent influences

In post-vocalic (coda) position, only three obstruents (/p/, /t/, /k/) and four sonorants (/l/, /n/, /m/, /ŋ/) can appear in Korean. We failed to find the general shortening effect of the consonants on the preceding vowels (Table 4). However, interestingly enough, we were able to find out that regardless of the post-vocalic stops, the pre-vocalic stops have dominant effects in shortening the relevant vowels. Figure 1 shows that in every case, pre-vocalic tense aspirated stops shorten the vowels most regardless of the post-vocalic consonants.

4.3 Sonorant influences

Both in phrase-initial and clause-final position, /ɾ/ has the least shortening effect on the following vowel,

Phrase-initial						
	a	i	u	e	o	AVE
n	0.71	0.77	0.77	0.76	0.59	0.72
ɾ	0.94	0.93	0.91	0.96	0.92	0.93
m	0.62	0.67	0.50	1.13	0.57	0.70
AVE	0.76	0.79	0.73	0.95	0.69	0.78
Clause-final						
	a	i	u	e	o	AVE
n	1.35	1.69	1.62	1.55	1.43	1.52
ɾ	1.61	1.69	1.86	1.65	1.72	1.71
m	1.21	1.53	1.43	1.41	1.32	1.38
AVE	1.39	1.64	1.63	1.54	1.49	1.54

Table 5. Compression factors caused by sonorant influence on the following vowel.

whereas /m/ has the greatest shortening effect (Table 5). In post-vocalic position, we could not find a pattern. This may be related to the difficulty in segmenting sonorant sounds.

5. PROSODIC INFLUENCES ON VOWEL DURATION

Figure 2 shows that in phrase-initial position, the vowel duration tends to be longer than in all other positions except clause-final. For example, in the phrase-initial position, the factor for /tu/ is 0.82. In the phrase second and third position, it becomes more shortened and in the clause-final position it is longest of all positions. The factors are 0.68 in the second syllable in the phrase, 0.59 in the third syllable, and 0.98 in the clause-final position.

However, there is one exception to the above observation. In Korean, for example, if a vowel in the second syllable in the phrase has a focus, it tends to become longer. This phenomenon only applies to the second syllable. So in some cases, the phrase-initial factor and this factor clash and we can observe the second vowel to be the same duration as or even longer than the vowel in the first syllable. In the recording, the speaker tended to focus on the syllable that is contrastive with other syllables. In this case, if the accent is on the second syllable in the phrase, it will have longer duration than the vowel in the phrase-initial syllable, because it is assigned a focus. For example, the vowel of /te/ in the second syllable has the compression factor of 0.81 while that in the phrase-initial syllable has the factor of 0.80. However, the compression factor of the vowel in the third syllable is 0.70, still shorter than the vowel in the phrase-initial syllable. In order to predict this prosodic effect more correctly, we need to investigate more factors such as lengthening due to emphasis as well as syllable position. We will deal with this matter in the further study.

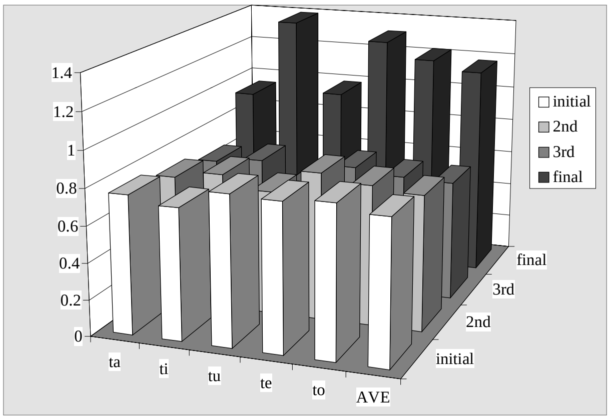

Figure 2. Compression factors on vowels caused by syllable position in the phrase.

6. APPLICATION TO SPEECH SYNTHESIS

6.1 Korean diphone database

Since 1998, we have been developing a Korean diphone database based on the MBROLA system. The database consists of 1,622 Korean diphones as basic units for concatenation of segment strings. So far we have succeeded in making pseudo-natural copy synthesized speech [9]. We have compared this speech with other synthetic speech from KT and ETRI using matching sentences. Most listeners were satisfied with the segmental quality of our synthesized speech. Nearly half of them preferred our synthetic speech to the others. If we can predict the durational values and control the intonation properly, we hope that this good intelligibility and naturalness will be retained.

6.2 Duration rules

We still do not have a fully automatic synthesis system. We also need to investigate more factors such as minimum duration and polysyllabic shortening to predict durations correctly. At the current stage, we aim to implement the result of this paper to predict the duration of syllables and vowels in contexts. We aim to encode the segmental duration rules along the lines of those developed by Dennis Klatt [10]. These rules were originally designed for predicting segmental durations in a linear framework, but in fact they do make use of a number of prosodic contexts. Klatt developed these rules for predicting vowel durations from strings of nonsense syllables spoken in a carrier phrase. In his design, each segment type is allocated an ‘inherent duration’ INHDUR, and a ‘minimum duration’ MINDUR. Each segment is then allocated a duration factor DUR according to a set of context sensitive rules. The final duration of the segment is then just:

Final Duration = MINDUR + (INHDUR - MINDUR) * DUR

Instead of a single DUR value, we will place a DUR feature with a default value of 1 on each node in a prosodic hierarchy. The final DUR value on a segment is then simply the product of the DUR feature values on all superordinate nodes, for example:

$$DUR_{TOTAL} = DUR_{IP} * DUR_{AG} * DUR_{FOOT} * DUR_{SYL} * DUR_{RHYME} * DUR_{CODA} * DUR_{CNS}$$

Then a set of declarative contexts can be specified under which a DUR feature is changed by some multiplicative factor. These factors will be derived from the results of our study.

7. CONCLUSION

In this paper, we have presented a couple of factors operating at segmental and prosodic levels: clause-final lengthening, phrase-internal shortening, and vowel shortening after obstruents. We still need to investigate MINDUR and such factors as polysyllabic shortening, unstressed shortening, lengthening due to emphasis, and shortening in clusters. These factors will operate at different levels of the prosodic structure.

8. REFERENCES

[1] http://zenith.etri.re.kr/SSPTLab/product/products.html

[2] Peterson. G. E. and I. Lehiste, (1960), “Duration of syllable nuclei in English,” *Journal of the Acoustical Society of America* 32, 693-703.

[3] Hawkins, S., J. House, M. Huckvale, J. Local & R. Ogden, (1998), “ProSynth: an integrated prosodic approach to device-independent, natural-sounding speech synthesis,” *Proc. 5th ICSLP ’98*, Sydney, 1707-1710.

[4] Dutoit, T., V. Pagel, N. Pierret, F. Bataille, and O. van der Vreken, (1996), “The MBROLA project: towards a set of high-quality speech synthesizers free of use for non-commercial purposes, ” *Proc. 4th ICSLP '96*, Philadelphia, vol.3, 1393-1396.

[5] Local, J., R. Ogden, (1997), “A model of timing for nonsegmental phonological structure.” In Jan P.H. van Santen, R W. Sproat, J. P. Olive & J. Hirschberg (eds.) *Progress in Speech Synthesis*. Springer, New York. 109-122.

[6] http://www.phon.ucl.ac.uk/resource/sfs.html

[7] Han, Mieko S., (1964), “Duration of Korean Vowels,” *Studies in the Phonology of Asian Languages* 2, Acoustics Phonetics Research Laboratory, University of Southern California, Los Angeles.

[8] Kim, Kong-On, (1974), *Temporal Structure of Spoken Korean: an Acoustic Phonetic Study*, Ph.D. dissertation, University of Southern California.

[9] http://www.phon.ucl.ac.uk/home/hchung/synth.htm

[10] Klatt, D. H., (1979), “Synthesis by rule of segmental durations in English sentences”, *Frontiers of Speech Communication Research*, ed. B. Lindblom & S. Ohman, Academic Press.