Distribution of Korean Syllables by Characters

Soonhyang Lee*

ABSTRACT

This study classifies Korean syllables into various types and investigates the distribution of syllables by each type. Korean syllables are classified into four or eight types. In this study, they are classified into thirty-two types based on character combination in order to evaluate the intelligibility of Korean synthetic syllables. Among those Korean syllables derived from the possible combinations of Korean characters, only currently used syllables were selected. Based on this classification and distribution, representative and diagnosable testing materials can be made. These testing materials can be applicable to intelligibility tests of Korean synthetic syllables.

Keywords: Syllable Distribution, Korean Synthetic Syllable, Character Combination, Testing Material, Intelligibility Test.

1. Introduction

This study examines the types of Korean syllables in order to evaluate the intelligibility of Korean synthetic syllables. The evaluation of synthetic speech quality includes intelligibility, naturalness, understandability, overall impression, voice quality, speech rate, etc (Gerit & Portele, 1988). Intelligibility can be evaluated at the syllable, word, phrase, and sentence level. Syllables are the fundamental units for the intelligibility test. The most important part in intelligibility test is to establish test materials. The intelligibility test materials should be representative, equivalent, diagnosable, sensitive, and efficient (Zhang et al., 1988). Following these principles, we should establish the intelligibility test lists, which should be selected from all the possible combinations of Korean syllables.

Korean syllables are classified into 4 types and 8 types on the basis of phoneme. But this paper is concerned with the evaluation of synthetic speech, and the input syllables into speech synthesis systems are not phoneme-compounding syllables, but character-compounding syllables. Considering this point, this study focuses on the distribution of Korean syllables by character combination, not by phoneme combination. All character-compounding syllables include both used syllables and unused syllables. If we select test

* Dept. of Secretarial Administration, Bucheon College

materials (syllables) among all character-compounding syllables including unused syllables, the materials will not represent the phonotactic structure of the Korean language. So we excluded character-compoundable but unused syllables (e.g., 곃), and investigated the distribution of the rest of them. The total number of character-compounding syllables is 19,530, but the number of currently used syllables in the dictionary must be fewer than it. So the number of currently used syllables and their distribution by types were investigated to make lists for an intelligibility test of synthetic syllables.

2. Types of Korean Syllables

Korean syllables can be classified into 4 types: "vowel", "vowel+consonant", "consonant+vowel", and "consonant+vowel+consonant". Based on phonemes, they can be classified into 8 types as follows;

Table 1. 8 types of Korean syllables

	Types of syllables	Example
1	monophthong	/a/
2	diphthong	/ye/
3	consonant+monophthong	/ga/
4	consonant+diphthong	/hyo/
5	monophthong+consonant	/al/
6	diphthong+consonant	/yal/
7	consonant+monophthong+consonant	/gok/
8	consonant+diphthong+consonant	/hyuŋ/

i) Monophthongs are /ㅏ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ, ㅐ/.
ii) Diphthongs are /ㅑ, ㅕ, ㅛ, ㅠ, ㅢ, ㅒ, ㅘ, ㅝ, ㅞ, ㅟ/.
iii) Consonants are /ㄱ, ㄲ, ㄴ, ㄷ, ㄸ, ㄹ, ㅁ, ㅂ, ㅃ, ㅅ, ㅆ, ㅇ, ㅈ, ㅉ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ/.

Huh (1965), Kim (1992), Lee (1996), and Lee & Choi (1997) claimed that the Korean monophthongal system is composed of 10 vowels; Oh (1993) claimed 10 vowels; and Bae (1996) claimed 8 vowels. Shin (2000) considered the monophthongal system to be composed of 7 vowels, in which /ㅟ/ and /ㅢ/ are classified as diphthongs, and /ㅐ/ and /ㅒ/ should be integrated into one phoneme. This study follows 7 vowel system.

Table 2. 32 Types of Korean syllables

4 types	8 types	32 types	examples
1	1	V1	아
	2	V2	야
2	3	C1 + V1	가
		C2 + V1	까
		C3 + V1	ㄱㅏ*
	4	C1 + V2	겨
		C2 + V2	껴
		C3 + V2	ㄱㅕ*
3	5	V1 + C1	악
		V1 + C2	았
		V1 + C3	않
	6	V2 + C1	연
		V2 + C2	왔
		V2 + C3	얇
4	7	C1 + V1 + C1	각
		C1 + V1 + C2	갔
		C1 + V1 + C3	값
		C2 + V1 + C1	꼭
		C2 + V1 + C2	꺾
		C2 + V1 + C3	끊
		C3 + V1 + C1	ㄱㅏㄴ*
		C3 + V1 + C2	ㄱㅏㅆ*
		C3 + V1 + C3	ㄱㅏㅄ*
	8	C1 + V2 + C1	멱
		C1 + V2 + C2	켜
		C1 + V2 + C3	갼*
		C2 + V2 + C1	뼉
		C2 + V2 + C2	꼈
		C2 + V2 + C3	깫*
		C3 + V2 + C1	ㄱㅕㄴ*
		C3 + V2 + C2	ㄱㅕㅆ*
		C3 + V2 + C3	ㄱㅕㅄ*

* V is for vowel and C is for consonant. V1 is for monophthongs, V2 is for diphthongs, C1 is for simple consonants, C2 is for twin consonants, and C3 is for cluster consonant.

This paper deals not with phoneme-compounding syllables, but character-compounding syllables. The results of this study will be used as the basis of test materials of synthetic speech, and the input syllables into speech synthesis system for test are not phoneme based syllables, but character based syllables. For example, "외", "왜", "웨" are phonetically realized as one syllable /we/, but all three syllables can be the test materials of synthetic speech. So considering that results of this study can be utilized in synthetic speech test, Korean syllables were subdivided into 32 types as in Table 2. This subdividing helps develop speech synthesis systems to make a convenient diagnosis of their systems. Table 3 shows the corresponding consonant and vowel characters.

Table 3. Vowels and consonants in the 32 types of Korean syllables.

notation	name	example
V1	monophthong	ㅏ, ㅐ, ㅓ, ㅔ, ㅗ, ㅜ, ㅡ, ㅣ
V2	diphthong	ㅑ, ㅒ, ㅕ, ㅖ, ㅘ, ㅙ, ㅚ, ㅛ, ㅝ, ㅞ, ㅟ, ㅠ, ㅢ
C1	simple consonant	ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ
C2	twin consonant	ㄲ, ㄸ, ㅃ, ㅆ, ㅉ
C3	cluster consonant	ㄳ, ㄵ, ㄶ, ㄺ, ㄻ, ㄼ, ㄽ, ㄾ, ㄿ, ㅀ, ㅄ

3. Types of Korean Syllables

This study used the *New Korean Dictionary* (Nam, 1994). All consonants and all vowels in Table 3 were combined into 19,540 syllables. Used syllables were distinguished from unused syllables out of all possible combined syllables by a programing language, C++. Used syllables extracted from the dictionary were reclassified into 32 types. And distribution of these used syllables were investigated by each type. Korean syllables in the dictionary have the distribution shown in Table 4.

Table 4. Distribution of Korean syllables in a dictionary

32 types	example	Total Syllables	Used Syllables	Distribution (%)
V1	아	8	8	0.4
V2	야	13	13	0.7
C1 + V1	가	104	104	5.8
C2 + V1	까	40	40	2.2
C3 + V1	ㄳㅏ*	88	0	0
C1 + V2	겨	169	100	5.5
C2 + V2	껴	65	30	1.7
C3 + V2	ㄳㅑ*	143	0	0
V1 + C1	악	112	68	3.8
V1 + C2	았	40	3	0.2
V1 + C3	않	88	18	1.0
V2 + C1	연	182	59	3.3
V2 + C2	왔	65	4	0.2
V2 + C3	왏	143	3	0.2
C1 + V1 + C1	각	1456	730	40.5
C1 + V1 + C2	갔	520	24	1.3
C1 + V1 + C3	값	1144	62	3.4
C2 + V1 + C1	꼭	560	222	12.3
C2 + V1 + C2	꺾	200	10	0.6
C2 + V1 + C3	꾾	440	13	0.7
C3 + V1 + C1	ㄳㄴ*	1232	0	0
C3 + V1 + C2	ㄳㅆ*	440	0	0

32 types	example	Total Syllables	Used Syllables	Distribution (%)
C3 + V1 + C3	ㄳ*	968	0	0
C1 + V2 + C1	먹	2366	236	13.1
C1 + V2 + C2	겪	845	10	0.3
C1 + V2 + C3	갃*	1859	0	0
C2 + V2 + C1	뻑	910	42	2.3
C2 + V2 + C2	꼈	325	4	0.2
C2 + V2 + C3	깛*	715	0	0
C3 + V2 + C1	ㄳ*	2002	0	0
C3 + V2 + C2	ㄳ*	715	0	0
C3 + V2 + C3	ㄳ*	1573	0	0
TOTAL		19,530	1,803	100

i) "Total Syllables" means the number of all syllables where all the possible combinations of consonants and vowels are combined.
ii) "Used Syllables" means all existing syllables in the dictionary.
iii) "Distribution" means percentages of corresponding type of syllables of all currently used syllables.
iv) "*" means the corresponding types of syllables not currently used (not in the dictionary).

First, "C1+V2+C1" type of syllables presented the largest percentage of all character compounding syllables (12.1%), while in used syllables, "C1+V1+C1" type of syllables occupied the largest percentage (40.5%).

Table 5. Proportion ranking of all types

Type	Percentage of Total Syllables		Percentage of Used Syllables	
C1 + V1 + C1	② 7.5%	1,456	① 40.5%	730
C1 + V2 + C1	① 12.1%	2,366	② 13.1%	236
C2 + V1 + C1	③ 2.9%	560	③ 12.3%	222

Second, in the distribution of four types of syllables (Table 6), in the case of "Consonant+vowel+consonant" type, the proportion of used syllables (75%) appeared to be lower than that of total syllables (93.6%). But in case of the other types, the proportion of used syllables was higher than the total syllables.

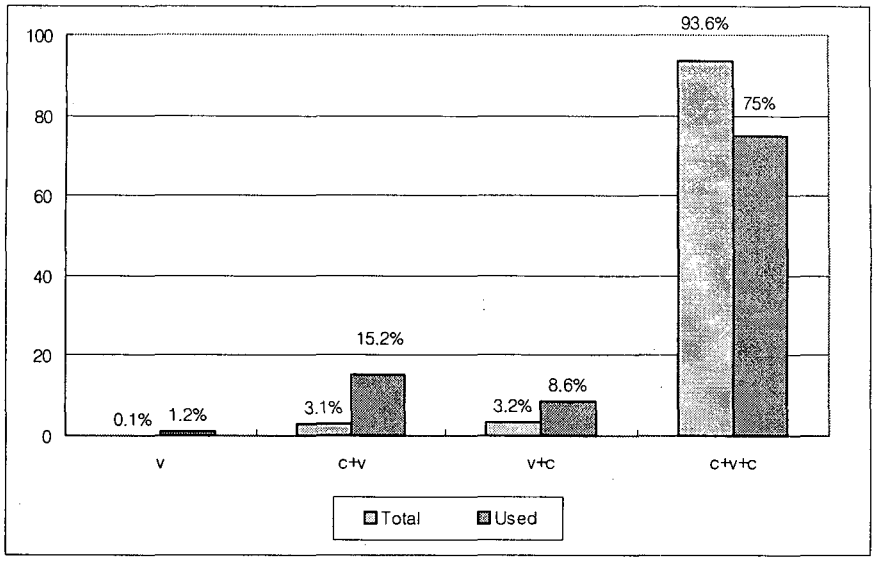


Figure 1. Distribution of 4 types of Syllables

Third, closed syllables (83.6%) occupied much more, proportionately, than open syllables (16.4%). Closed syllables with simple consonants in the final position, are the most common (90% of all closed syllables) as in Table 6.

Fourth, in eight types of syllables, "C+V1+C" proportionately was the most common (59%). And "C+V2+C" was second (16%), as is shown in Figure 2.

Table 6. Distribution of syllables by consonant types

Type of syllables	Number of Syllables	Distribution (%)	Type of Syllables	Number of Syllables	Distribution (%)
open syllables	295	16.4	V	21	1.2
			CV	274	15.2
closed syllables	1508	83.6	_C1	1357	75.3
			_C2	55	3.1
			_C3	96	5.3

* "-C1" means the final consonant of a syllable is a simple consonant, "-C2" means the final consonant of a syllable is a twin consonant, and "_C3" means the final consonant of a syllable is a cluster consonant.

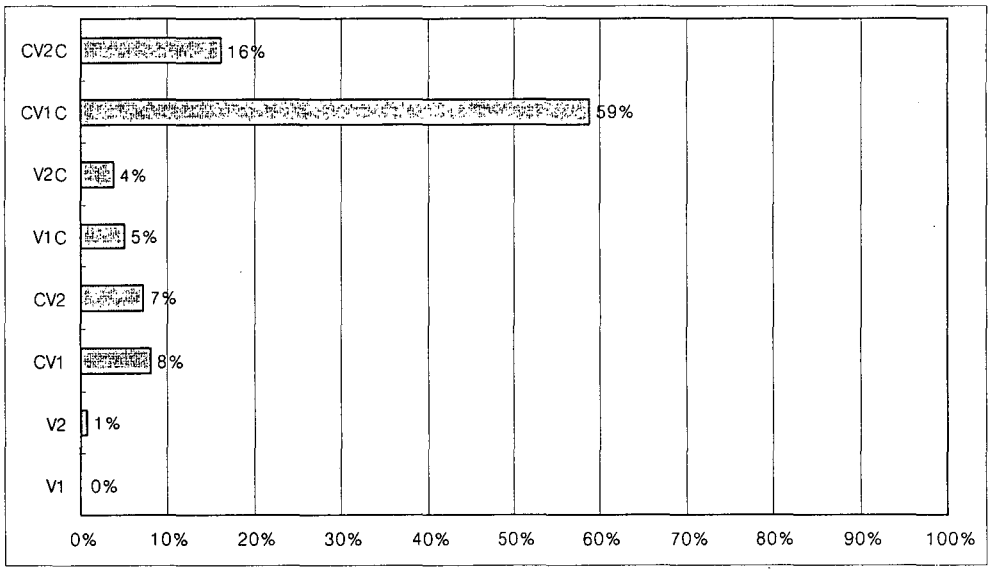
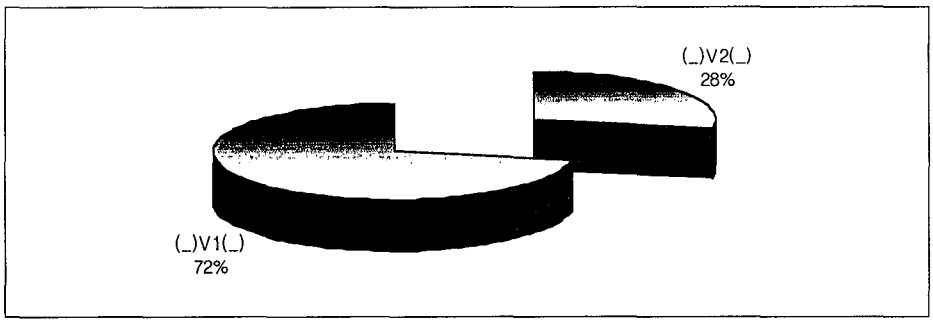


Figure 2. Distribution of eight types of syllables

Fifth, the proportion of syllables with a monophthongal nucleus was 72%, whereas the proportion of syllables with diphthongal nucleus was 28%.



* - means all possible consonants.

Figure 3. Distribution of syllables by nucleus types

4. Summary and Limitation

We investigated the distribution of Korean syllables by 32 types. The results of this study will be the basic materials for an intelligibility test of Korean synthetic speech. When we make test materials for the evaluation of synthetic speech, the testing syllables can be selected by the proportion of corresponding types based on the types of syllables. Test materials made by this method will be representative and reasonable. This study

used only the lexical entries of the New Korean dictionary by Nam(1994), so we should expand the data by observing additional dictionaries.

In this study, syllables were classified into 32 types by how their characters are compounded. We should further study the new classification of syllables, focusing on the manner of articulation. When the articulation features of consonants and vowels are employed to classify the types of syllables, the test materials can be helpful in finding find which features of consonant and vowel cause problems in speech synthesis.

References

Bae, J. 1996. *Introduction to Korean Phonology* Seoul: Shingu Publications.

Sonntag, Gerit P. & Thomas Portele. 1988. "Comparative evaluation of synthetic prosody with the PURR method." *ICSLP* Vol. 5, 1715-1718.

Huh, U. 1965. *Korean Phonology.* Seoul: Jeongeumsa.

J, Zhang., Dong S., & G. Yu. 1988. "Total Quality evaluation of speech synthesis systems." *ICSLP* Vol. 5, pp. 1711-1714.

Kim, M. 1992. *Korean Phonology.* Seoul: Hanshin Publications.

Lee, B. & M. Choi. 1997. *Korean Phonology,* Korea National Open University Press.

Lee, H. 1996. *Korean Phonetics.* Seoul: Taehaksa.

Nam, Y. 1994. *New Korean Dictionary.* Seoul: Seongandang.

Oh, J. 1993. *Modern Korean Phonology.* Seoul: Hyungseol Publications.

Shin, J. 2000. *Understanding Sound Speech.* Seoul: Hangook Publications.

Received: Jan. 31, 2002.
Accepted: Feb. 27, 2002.

▲ Soonhyang, Lee
Department of Secretarial Administration, Bucheon College
424 Simgok-dong, Wonmi-gu, Bucheon, Gyeonggido, 420-735
Tel: +82-32-610-3433 (O)
E-mail: shlee@bc.ac.kr