Durational Characteristics of Korean Lombard Speech

Sunhee Kim

Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology, South Korea
shkim@ee.kaist.ac.kr

Abstract

This paper aims to examine the durational characteristics of Korean Lombard speech using speech data, which consist of 500 Lombard words and 500 normal words of 10 speakers (5 males and 5 females). Each word was segmented and labeled manually and the duration of each segment and each word was extracted for comparison. The durational change of Lombard speech in comparison with normal speech was analyzed using a statistical method. The results show that durational characteristics of Korean Lombard speech mostly correspond to the results reported in previous research. The duration of words in Lombard speech is increased in comparison with normal speech, and the average unvoiced consonantal duration is decreased while the average vocalic duration is increased. Female speakers show a stronger tendency to lengthen the duration in Lombard speech, but the difference between females and males was not statistically significant. Finally, this study also shows that the speakers of Lombard speech could be classified according to their different duration rates, which demonstrates speaker dependent characteristics.

1. Introduction

The phenomenon caused by the *articulatory changes made by speakers in order to be more intelligible in a noisy environment*, so-called Lombard effect, is known as one of the causes which triggers recognition performance degradation [1, 2, 3, 4]. Previous studies related to Lombard speech are focused on the analysis of the relevant features [1, 4], or on the development of algorithms to improve the performance of a speech recognizer based on those analyses [2, 3]. More studies on this subject were presented and summarized within the framework of automatic speech recognition in [1, 2, 3, 4].

As was reported by previous researchers on the Lombard effect, many acoustic and perceptual characteristics depend on the speakers. Despite the differences such as the recording conditions and the number and characteristics of speakers, similar tendencies are observed across languages. According to [1], the main acoustic changes between normal speech and Lombard speech are: (i) increase in fundamental frequency (F0); (ii) shift in energy from low frequency bands to middle or high bands; (iii) increase in level; (iv) increase in vowel duration; (v) spectral tilting; and (vi) shift in formant center frequencies for F1(mainly) and F2. General methods used to improve the performance of speech recognition algorithms for Lombard speech were summarized in three general areas[3]: (i) those of robust features; (ii) those of equalization methods, (iii) those of model adjustment or training methods.

Studies on Lombard speech of particular languages have been conducted for American English, French, Spanish and Japanese [1]. Lombard speech was found to be distinct from normal speech in similar ways for four languages. The results for French and English are very similar, while those for Spanish include two particularities. They show some differences in the variation of second formant and significant differences between male and female speakers in the areas of the pitch and the spectral energy distribution. In Japanese, tendencies of a shift for the first and second formants were reported in certain conditions.

For the Korean language, [5] proposed a Lombard effect compensation model simulating the variations of formant location, formant bandwidth, pitch, spectral tilt, and energy in each frequency band. [6] proposed a robust feature extraction method using Lombard effect compensation filter. Both [5] and [6] focused on the compensation, while acoustic-phonetic characteristics of Korean Lombard speech have rarely been studied in depth.

This paper presents the durational characteristics of Korean Lombard speech among the features characterizing the Lombard effect in Korean. Focusing on a particular characteristic, the duration, this study will show, first of all, in what class of phonetic units of the language the characteristic in question is relevant, and furthermore, how different speakers use this characteristic as Lombard effect, which could be a speaker dependent characteristic.

This paper is organized as follows: Section 2 presents Korean phonetic units and Section 3 describes the speech database used for the analysis. Section 4 presents the experimental results and Section 5 discussion of the analysis.

2. Phonetic units of Korean

Table 1 and *Table 2* present Korean consonants and vowels. The IPA of each phonetic unit is presented with its corresponding computer readable phonetic alphabet in parentheses proposed in [8].

		labial	dental	palatal	velar	glottal
stop	lenis	p (b)	t (d)		k (g)	
	fortis	p' (B)	t' (D)		k' (G)	
	aspirated	**p^h (p)**	**t^h (t)**		**k^h (k)**	
affricate	lenis			c (z)		
	fortis			c' (Z)		
	aspirated			**c^h (c)**		
fricative	lenis		s (s)			h (h)
	fortis		s' (S)			
nasal		m (m)	n (n)			ŋ (N)
liquid			l /r (l/r)			

Table 1: Korean consonants

	anterior		posterior	
	unround		unround	round
high	i (i)		ɯ (ʊ)	u (u)
middl	e (e)		ʌ (ʋ)	o (o)
low			a (a)	

			ɯy	wi	
ye	yʌ	yu		we	wʌ
	ya	yo			wa

Table 2: Korean vowels and diphthongs

As the word list for the speech data consists of 50 words for PC command, it is not completely phonetically balanced and the 5 units shaded in the tables, which do not influence the results of the current analysis, were not included.

3. The Speech Database

The acoustic-phonetic analysis was performed on a portion of the speech data designed in [6]. In this study, 500 Lombard words and 500 normal words of 10 speakers (5 males and 5 females) were used, while the whole speech data consist of 9000 words of 90 speakers. Each speaker pronounced the list of words twice in normal speech and three times in Lombard speech. After evaluating the recorded utterances, one utterance for each style per speaker was saved. All 10 speakers selected for this study are natives of the same linguistic area.

50 isolated words for PC command were pronounced in normal speech and in Lombard speech. They are composed of 4 one-syllable words, 36 two-syllable words, 6 three-syllable words and 4 four-syllable words. *Table 3* shows the complete list of the words classified by the number of syllables. The words are transcribed in the computer readable phonetic alphabet with their gloss.

1-Syll W	GUd 'end', ye 'yes', on 'on', zib 'home'
2-Syll W	gvmseg 'search', nalZa 'date', daUm 'next', dadGi 'close', dwilo 'backword', menyu 'menu', meil 'mail', banbog 'repetition', bogi 'see', bollyum 'volume', svlZvN 'set up', sUtob 'stop', sizag 'start', amho 'password', yvngyvl 'connect', yvlgi 'open', ozvn 'morning', opU 'off', ohu 'afternoon', izvn 'before', ilZvN 'schedule', zamgUm 'lock', zeseN 'play', zvzaN 'save', zvnsoN 'send', zvNzi 'stop', zoNnjo 'end', zuso 'address', zuNzi 'stop', cvUm 'first', cuga 'add', cugSo 'reduce', toNhwa 'call', heze 'cancel', hwagDe 'enlarge', hwagin 'confirm'
3-Syll W	GUnnegi 'finish', dwegamGi 'rewind', anio 'no', apUlo 'forward', zebalSin 'recall', pUllei 'play'
4-Syll W	bollyumnadcum 'turn up', bollyumnopim 'turn down', ilSizvNzi 'pause', ilSizuNzi 'pause'

Table 3: Word list

A total of 256 phonetic units appeared in the word list, which consisted of 41 stops, 17 fricatives, 28 affricates, 41 nasals, 19 liquids, 95 vowels and 15 diphthongs.

The recording was conducted in a quiet room located on campus. In order to obtain normal utterances, each speaker was asked to pronounce each word appearing on a screen with 3 second of pause between two words. A unidirectional microphone was placed at a distance of 0.5 m from the speaker for the recording.

The acquisition of Lombard speech in this study differs from that of most previous experiments [1, 2, 3, 4], in which a speaker is asked to pronounce tokens while listening to a certain level of noise transferring through headphones. As the speaker increases his/her voice level to be more intelligible in the noisy environment, we assume that he/she shows a similar tendency when he/she is asked to control the PC at a certain distance in voice. In that case, the speaker is also expected to increase his/her voice level, although not as much as shouted speech, so that the resulting utterances have the same effect as Lombard speech [7]. For the recording, a voice command controlled PC was placed at a distance of 3 m and the utterances were recorded through the same kind of unidirectional microphones used for the recording of normal speech at a distance of 0.5 m, 1 m, 2 m, and 3 m simultaneously. In this study, the recordings captured by microphone at a distance of 0.5 m form the speaker, that is, at the same distance for normal speech from the speaker, were analyzed in order to compare the acoustic features of Lombard speech with those produced in normal speech.

Each utterance was segmented and labeled manually using Praat 4.2.21 according to the segmental and prosodic labeling convention proposed in [8].

4. Results

4.1. Durational analysis

4.1.1. Duration of words

Table 4 shows the average difference of duration between normal speech and Lombard speech, and the percentage variation (100*(Lombard average – Normal average)/Normal average) for each class of words.

	Average difference of duration (SD)	Percentage variation
1-Syll W	74.25 (50.83)	25.28
2-Syll W	78.6 (105.09)	13.9
3-Syll W	79.63 (99.33)	12.24
4-Syll W	-3.07 (99.45)	-0.3
Average	57.35 (77.66)	9.08

Table 4: Average difference of duration (ms) with its standard deviation (SD) and percentage variation (%) for words

As is shown in *Table 4*, the duration of word of Lombard speech increases by 9.08% on average compared with normal speech. The percentage variation is highest in 1-syllable words, and it decreases when the number of syllables varies to 2 to 3. In case of 4-syllable words, the duration decreases by less that 1%. The results of the paired sample t-test indicated a statistically significant difference between normal

speech and Lombard speech for all groups of words except 4-syllable words. For 1-syllable words, the t=4.61, p<.001; for 2-syllable words, the t=2.63, p<.05; and for 3-syllable words, the t=2.81, p<.05; while no statistically significant difference exists between normal speech and Lombard speech for 4-syllable words.

Table 5 shows the average difference of duration with its standard deviation and the percentage variation depending on the gender of speakers. Although the tendencies for both genders are almost same as those in *Table 4*, the percentage variation in female speakers is higher than that in male speakers. But, contrary to our expectations, the results of the paired sample t-test indicated that there was no statistically significant difference between male speakers and female speakers in spite of the quite higher percentage variation of female speakers.

	Male		Female	
	D (SD)	PV	D (SD)	PV
1 Syll W	51.6 (35.84)	19.9	96.9 (56.97)	29.53
2 Syll W	44.33 (56.62)	8.7	112.88 (136.77)	18.16
3 Syll W	35.73 (49.08)	5.91	123.53 (122.37)	17.73
4 Syll W	-4.15 (126.23)	- 0.44	- 2.0 (79.47)	- 0.18
Average	31.87 (64.00)	5.52	82.82 (88.61)	12.07

Table5: Average difference of duration (D: ms) with its standard deviation (SD) and percentage variation (PV: %) for words for male and female speakers

4.1.2. Duration of segments

The duration of segments of consonants of Lombard speech decreases compared to that of normal speech, while that of vowels increases as in *Table 6*. The mean percentage variation of consonants is –9.10%, and those of vowels and diphthongs are 34.28% and 24.71% respectively.

	D (SD)	PV
Stops	-14.54 (14.41)	-15.59
Fricatives	-16.16 (17.87)	-15.65
Affricates	-8.53 (23.96)	-7.42
Nasals	-10.11 (23.17)	-7.77
Liquids	.58 (7.71)	.94
Vowels	44.09 (30.48)	34.28
Diphthongs	45.21 (39.09)	24.71

Table 6: Average difference of duration (D: ms) with its standard deviation (SD) and percentage variation (PV: %) for segment classes

The results of the paired sample t-test indicated a statistically significant difference between normal speech and Lombard speech for unvoiced consonants and vowels, whereas, in voiced consonants, nasals and liquids, the segment duration of Lombard speech is not significantly different from that of normal speech. For stops, the t=5.87, p=<.001, for fricatives, the t=2.74, p<.05, for affricates, the t=3.33, p<.01; for vowels, the t=-4.57, p<.01; and for diphthongs, the t=-3.65, p<.01.

Table 7 shows the average difference of duration with its standard deviation and the percentage variation depending on the gender of speakers for each phone class. The decrease rate of consonants in male speakers is higher than that in female speakers, while the increase rate of vowels and diphthongs for females is higher than that for males. As was mentioned above, the statistical analyses do not result in showing any significant difference.

	Male		Female	
	D (SD)	PV	D (SD)	PV
Stops	-17.45 (8.32)	-19.95	-11.63 (19.4)	-15.74
Fricatives	-24.74 (21.54)	-22.99	-7.57 (8.38)	-7.68
Affricates	-18.64 (17.17)	-16.41	-11.61 (17.25)	-9.76
Nasals	-3.23 (25.1)	-2.96	-16.98 (21.44)	-11.24
Liquids	-1.86 (8.75)	-3.3	3.03 (6.49)	4.48
Vowels	31.59 (11.68)	27.37	56.58 (39.55)	39.9
Diphthongs	32.34 (31.23)	19.68	58.07 (45.26)	28.81

Table 7: Average difference of duration (D: ms) with its standard deviation (SD) and percentage variation (PV: %) for segment classes for male and female speakers

4.2. Classification of speakers based on durational analysis

In this section, the speaker dependent characteristics of Lombard speech will be examined by classifying the speakers according to their durational changes in percentage variation.

In *Table 8*, the percentage variation of words of 10 speakers was presented for each group of words. The 4-syllable words were excluded here, as they were proved to be statistically insignificant. For 1-syllable words, all the speakers pronounced the tokens longer in Lombard speech than in normal speech. For 2- and 3- syllable words, the duration of Lombard speech increased for seven speakers, even though the rate varied depending on each individual. Three speakers pronounced the Lombard speech with a small reduction of duration rate, which was shown shaded in the *Table 8*.

	1 Syll-W	2 Syll-W	3 Syll-W
M1	12.02	20.4	14.37
M2	51.49	23.94	18.53
M3	22.98	9.82	7.74
M4	43.49	17.21	13.34
M5	10.92	-0.23	-3.39
F1	16.73	-13.29	-3.59
F2	9.35	-2.90	-0.83
F3	47.73	44.11	40.27
F4	25.76	25.77	24
F5	12.32	18.44	12.16

Table 8: Classification of speakers by durational analysis of word classes
(M1~M5: Male speakers, F1~F5: Female speakers)

Table 9 shows the 10 speakers and their percentage variations for each class of segments. When the percentage variation value is positive, in other words, in case when the Lombard speech is pronounced longer than normal speech, the cells are shaded. Like the global tendencies of the durational changes we have already seen in *Table 6*, for all speakers the percentage variations of vowels and diphthongs were increased. The speakers could be classified in three groups based on their tendencies of lengthening or shortening of segment duration. The first group (M1, M2 and F2) distinguishes non-voiced classes from voice classes, and the second group (M3, M4, M5, F1 and F5) consonants from vowels. For the third group (F3 and F4) the sonorants are divided, and nasals show consonant-like behavior, whereas liquids show vowel-like behavior.

	S	F	A	N	L	V	D
M1							
M2							
M3							
M4							
M5							
F1							
F2							
F3							
F4							
F5							

Table 9: Classification of speakers by durational analysis of segment classes (M1~M5: Male speakers, F1~F5: Female speakers, S: Stops, F: Fricatives, A: Affricates, N: Nasals, L: Liquids, V: Vowels, D: Diphthongs)

Table 8 and *9* present a typology of the speakers according to the strategies each speaker employs for Lombard speech, which reflects the general tendencies as well as speaker dependent variations.

5. Discussion

This paper examined the durational characteristics of Korean Lombard speech compared to the normal speech using speech data of 1000 words. The durational change of Lombard speech in comparison with normal speech was analyzed within words and classes of segments.

According to the durational analysis of the words and segments, the duration of words in Lombard speech is increased in comparison with normal speech, and the average unvoiced consonantal duration is reduced while the average vocalic duration is increased. The increase of vowel duration has been pointed out as an important characteristic of Lombard speech in previous works [1, 2, 3, 4], which also applies in the case of Korean. As for the decrease of consonants, [2] and [4] show the same phenomenon in both English and Spanish. Especially, in the case of Spanish, only two consonants out of 18 resulted in a statistical significance and the decrease rate was less than 2%. Consequently, Korean unvoiced consonants are prominent examples that show the decrease of duration in Lombard speech. Furthermore, the results allow us to infer that the increase of the duration of words in Korean Lombard speech is provoked by the durational increase of vowels and diphthongs.

The influence of the gender of speakers was not statistically significant in the durational characteristics despite the higher percentage variation found in female speakers compared to that of male speakers.

The speaker dependent characteristics of Lombard speech was also examined by classifying the speakers according to their durational changes in percentage variation. With regard to the percentage variation of words, all the speakers pronounced the tokens longer in Lombard speech than in normal speech for 1-syllable words. For 2- and 3-syllable words, the duration of Lombard speech increased for seven speakers, while three speakers pronounced the Lombard speech with a small reduction of duration rate. In respect to the percentage variation in each phone class, speakers could be classified in three groups based on their tendencies of lengthening or shortening of segment duration. In sum, speaker dependent characteristics of Lombard speech was demonstrated through this classification of speakers according to the strategies each speaker employs for durational change in Lombard speech compared to normal speech.

6. Acknowledgements

This work was supported by Brain Korea 21 Project, the School of Information Technology, KAIST in 2005.

7. References

[1] Junqua, J.-C., "The Influence of Acoustics on Speech Production: A Noise-Induced Stress Phenomenon as the Lombard Reflex," *Speech Communication*, Vol. 20, pp. 13-22, 1996.

[2] Hansen, J., "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech Recognition," *Speech Communication*, Vol. 20, pp. 151-173, 1996.

[3] Bou-Ghazale, S. E., J. Hansen, "A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress, " *IEEE Transactions on Speech and Audio Processing*, Vol. 8-4, pp. 429-442, 2000.

[4] Castellanos, A., J.-M. Benedi, F. Casacuberta, "An Analysis of General Acoustic-Phonetic Features for Spanish Speech Produced with the Lombard Effect", *Speech Communication*, Vol. 20, pp. 23-35, 1996.

[5] Chi, S., Y-.H. Oh, "Lombard Effect compensation and noise suppression for Noisy Lombard Speech Recognition", *Proc. ICASSP*, pp. 2013-2016, 1996.

[6] Woo, S.-Y., Robust feature extraction using Lombard effect compensation filter, Master's thesis, Korea Advanced Institute of Science and Technology, 2003.

[7] Linéard, M.-G., "Effect of vocal effort on spectral properties of vowels", J. Acoust. Soc. Am. Vol. 106-1, pp. 411-422, 1999.

[8] Lee, S.-H. et al, "Some considerations on SiTEC segmental and prosodic labeling convention for Korean", *Malsori*, Vol. 46, pp. 127-143, 2003.