「언어연구」, 제23권 제1호(2006), 81～95

Whispering French and Korean: A Comparative Study

Géraldine Vercherand, In-Young Kim and Hi-Yon Yoo

(UPRES EA 33, UFR Linguistique Université Paris 7)

Géraldine Vercherand, In-Young Kim and Hi-Yon Yoo. 2006. Whispering French and Korean: A Comparative Study. *Linguistics Research* 23.1, 81-95. In this paper, we propose to analyze the prosodic realization of two phrasal modalities (declarative and interrogative) through a production and perception study of whispered voice in French and in Korean. We present the analysis of the acoustic parameters that are at stake when producing whispered voice and the results of an identification test. Results show that the most important acoustic cue in differentiating the two modalities in whisper is intensity, which has a higher value at the end of interrogative sentences. Moreover, the results of the identification test show that the two modalities are globally well perceived in whisper, especially while the good identification rate is higher in longer sentences in French whispering. **(UPRES EA 33, UFR Linguistique Université Paris 7)**

Keywords whispering, whispered voice, duration, intensity, F0

1. Introduction

Whispering is a way of speech - even it seems to have no distinctive linguistic function – which is used by most of every language to communicate in a private situation. By simple definition, the whispering voice does not vibrate the vocal folds: the vocal folds remain opening for whispering speech; as a result, there is no fundamental frequency through the whole sentence. Beside this fact, if whispered voice remains as a way of speech, it means that whispering really preserves its' role to be recognized as an intonation for communication despite the absence of F0. This could be problematic, if we consider the F0 as a acoustic key parameter in speech for intonation affects, which holds the key role for phrasal modality perception. If intonation continues to be perceived in whispering voice, then what are the other ways to produce intonation?

The majority of studies concerning whispering voice treats its' acoustic and physiologic aspects: for example, glottal characters [10]; larynx movements [1]; duration maintained by consonants [5], [9] or height of intrinsic segments [4], [7]. The materials which were used for the tests were consisted in words or un-isolated words: see Thomas's work on vowel height perception [11]. But there are very few studies on prosodic realization in whispering voice: i.e., studies on Chinese ton [3], [6] and Vietnamese [8]. Finally, not very much studies in French whispering voice seems to exist : Fonagy [2] is one of the rare work that we can look on. In this study, we propose to analyze the whispering voice produced by reading and by realizing the two modalities: declarative and interrogative modality in French and Korean.

So for this study, we will start with 2 hypotheses: a) duration and intensity is enough to distinguish the phrasal modality; b) rhythmic unit, which could be just 1 or 2 syllables for a sentence- should be enough so as to both fully represent itself and also be understood clearly as a long phrase. We will show the production results analyzed from the whispered voice, and then the identification test's results will come afterwards.

2. Production in French Whispering Speech

2.1 Corpora Material in Whispering Speech Production

The French corpus consists 13 phrases; the phrases are semantically neutralized and had simple syntactic structure. Especially, the corpus had two phrases each for 6 syllable phrase and 7 syllable phrases.

(1) a.	Elle	mange	Ø.	(2 syllables)
	'She	eats.'		
b.	Elle	mange	bien.	(3 syllables)
	'She	eats	well.'	
c.	Anna	mange	bien.	(4 syllables)
	'Anna	eats	well.'	
d.	Anna	mange	une noix.	(5 syllables)
	'Anna	eats	a walnut.'	

e.	Anna	mange	un melon.	(6a syllables)
	'Anna	eats	a melon.'	
f.	Anna	mangea	une noix.	(6b syllables)
	'Anna	ate	a walnut.'	
g.	Anna	mange	des ananas.	(7a syllables)
	'Anna	eats	some bananas.'	
h.	Anna	mangea	un melon.	(7b syllables)
	'Anna	ate	a melon.'	
i.	Annabelle	mange	des ananas.	(8 syllables)
	'Annabelle	eats	some pineapple.'	
j.	Annabelle	mange	une noix du Ghana.	(9 syllables)
	'Annabelle	eats	a walnut from Ghana.'	
k.	Annabelle	mange	un melon du Ghana.	(10 syllables)
	'Annabelle	eats	a melon from Ghana.'	
l.	Annabella	mange	un melon du Ghana.	(11 syllables)
	'Annabelle	eats	a melon from Ghana.'	
m.	Annabella	mange	des ananas du Ghana.	(12 syllables)
	'Annabelle	eats	some pineapple from Ghana.'	

The length was incremented from 2 syllables to 12 syllables, 1 syllable incremented each time. The number of syllables as a parameter for length variation is because French is considered as syllable timing language. Of course, we had no linguistic orthographic marker to differentiate the modalities. Every phrase was capable to be produced as declarative or interrogative phrase. The number of total stimuli becomes 26(13 phrases x 2 modalities). We tried to maintain the CV structure for every French word, maximizing the open vowels and the nasal consonants. This was intended to make every segment voiced and also that we can minimize the intrinsic effects of phones. Every sentence was mixed randomly, written on an A4 paper sheet which was presented to be read. 3 French native speakers were invited to a quiet room at UFR Linguistique, Université Paris 7. The recordings were affected directly into the PC via unidirectional microphone. The speakers were asked to pronounce the sentences natural as possible in two types of speech: normal voice and whispering voice. 5 occurrences obtained from each speaker. The first and the last occurrence were not

integrated for the acoustic analysis; only 3 occurrences per phrase were considered. The duration of each phone and also the intensity were measured.

2.2 Acoustic Analysis Results

The first step of our analysis is comparing the normal voice and the whispering voice. No significant difference was found between the 3 French speakers. The normal voice shows the presence of F0 and high average of intensity: the mean intensity for whispering voice is 22dB inferior to the normal voice. The F0 rises at the final position of the phrase for interrogative sentences and falls at the same final position for declarative sentences. Considering the duration and intensity, the proportion of each segments inside the sentence seemed to be relatively equal between the two modalities. Also, final lengthening via the last syllable was observed and this accentuation was especially remarkable in long sentences than the short ones.

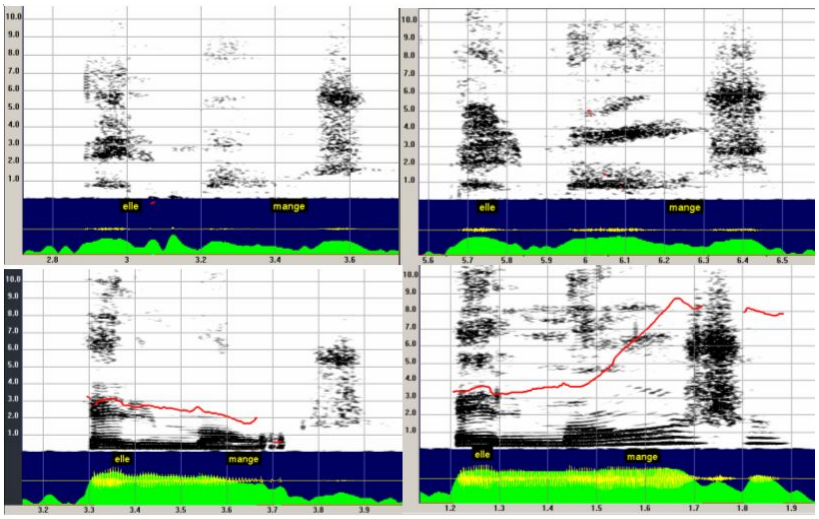

Image 1: Example of French production in 2 syllables:
the 2 upper images are in whispering voice; the 2 lower images are in normal voice
(declarative on the left side and interrogative on the right side).

For our acoustic analysis to verify our hypothesis in whispering voice, first, we measured the duration of each phoneme; the highest intensity of each phoneme was also measured. For the analysis, we calculated the duration of each syllable and we calculated the relative intensity for each syllable; this way of measurement was to measure the highest intensity difference between the consonant and the vowel inside each syllable. This will be so called 'dynamic intensity'.

In French whispering, a duration lengthening over the final syllable of each phrase was found. Further more, remarkable final lengthening for declarative sentences was present. We had faster speech speed for the interrogative French sentences; approximately, 1 syllable less per second was observed. Proportion of segments inside the sentence seemed to be equal for each modality. So the proportion of segments seems not to contribute for modality perception.

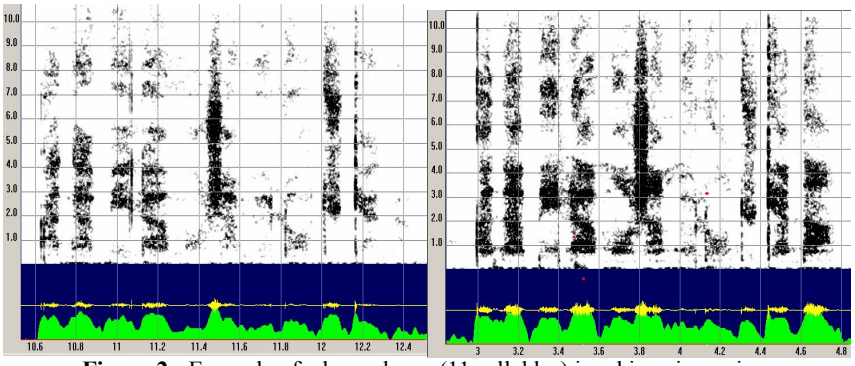

Figure 2 : Example of a long phrase (11 syllables) in whispering voice:
« Annabella mange un melon du Ghana »

The declarative sentence is on left side and interrogative sentence on the right side.

Intensity itself seems to be the most important parameter to distinguish the difference between declarative and interrogative sentences. Higher intensity for interrogative sentence than for declarative sentences was observed. The average of relative intensity for the whispering in French was 4dB, and for the normal speech, they showed 7dB. So the normal voice's dynamic intensity is higher than the dynamic intensity from whispering voice in French.

The concluding remark could be the intensity as a key parameter that permits to

perceive the phrasal modality; distinguishing declarative and interrogative sentences. The final syllables' high intensity in interrogative could be compared to the declarative's final syllables where the declining intensity is maintained. So in French whispering voice, the interrogative sentences have high intensity and fast speed rate as their characteristic acoustic factors.

2.3 Perception in French Whispering Voice

This perception test is to identify the modality perception and also verify the perspectives that we proposed as hypothesis at the beginning. The perspective includes the length effects (the number of syllables) on the perceptive recognition rate of modalities.

For our perception test, the corpus was consisted by 26 stimuli, one phrase per number of syllables and modality. The stimuli from recordings for acoustic analysis were selected for the test. 10 native listeners for each language participated for the perception test. Each stimulus was 5 times, randomly repeated to listeners and they were asked to mark on a answering sheet; they were forced to choose whether it is a declarative sentence or an interrogative sentence. For the analysis, only the last answer from the 5 was used.

2.4 Perception Test Results in French Whispering Voice

In French whispering generally, the modality was fairly recognized. We had 76% homogeneous response as result; 73% for interrogative phrases and 78% for the declarative sentences. So, ANOVA test showed statistically significant effect by length; $F=2$ and $p<0.001$ for declarative sentences; $F=9$ and $p<0.001$ for interrogative sentences.

Confirming our hypothesis, the perception rate correlates with the phrasal length. As the sentence length became longer, the homogeneous reply rate raised. For better results, we regrouped the responses in 3; the first group made by 2 syllable to 5 syllable sentence stimulus showed 68% homogeneous response; the second group in 6

to 8 syllable sentence stimulus showed 72% homogeneous response; the third group in 9 to 12 syllable sentence stimulus reached 88.5% of homogeneous responses (Table 1).

	Identification rate		
Modality	2 to 5 syllables	6 to 8 syllables	9 to 12 syllables
Interrogative	67.5%	64%	91%
Declarative	68.5%	80%	86%
Mean	68%	72%	88.5%

Table 1: The homogenous modality perception rate from total listeners: results rearranged by groups.

Confirming our hypothesis, the perception rate correlates with the phrasal length. As the sentence length became longer, the homogeneous reply rate raised. When the phrases have more that 9 syllables, the homogeneous response rate goes nearly up to 90%. The interrogative sentences show homogeneous data distribution more than 8 syllables (coefficient lower than 15%). It means that modality is unanimously well recognized from 9 syllables.

Concerning declarative modality, the results are less homogeneous. This result seems to be due to one particular listener, so when we discard this listener's data, the results become similar to the interrogative one. This pilot test opens 2 remarkable points;

- the phrasal modality is globally well perceived in whispering voice,
- the ratio for good perception is related to the number of syllables.

3. Production in Korean Whispering Speech

3.1 Corpora Material in Whispering Speech Production

Korean is also considered as a syllable timed language as French, so we can control the phrasal length as a control factor comparing these 2 languages. The corpus

consists 13 phrases; the phrases are semantically neutralized and had simple syntactic structure. Especially, the corpus had three phrases for 6 syllable phrase, in order to see whether the syllabic constitution has effects on production.

The length was incremented from 2 syllables to 12 syllables, 1 syllable incremented each time. Of course, we had no linguistic orthographic marker to differentiate the modalities. Every phrase was capable to be produced as declarative or interrogative phrase. The number of total stimuli becomes 26; 13 phrases x 2 modalities. As in French, the CV or CVC structure for every Korean word are maintained by maximizing the open vowels and the nasal consonants, especially nasal for the offset consonant to make every segment voiced so that we can minimize the intrinsic effects of phones.

Every sentence was mixed randomly, written on an A4 paper sheet which was presented to be read. 3 Korean female native speakers were invited to a quiet local room. The recordings were affected directly into the PC via unidirectional microphone. The speakers were asked to pronounce the sentences natural as possible in two types of speech: normal voice and whispering voice. 5 occurrences obtained from each speaker. The first and the last occurrence were not integrated for the acoustic analysis; only 3 occurrences per phrase were considered. The duration of each phone and also the intensity were measured.

(2) [1] a. 먹어 (2 syllables)
m∧.k∧
eat
'(*Agent* is) eating.' or '(Is *agent*) eating?'

b. 감 먹어 (3 syllables)
kam m∧.k∧
persimmon eat
'(*Agent* is) eating a persimmon.' or '(Is *agent*) eating a persimmon?'

1. The translation for Korean corpus in English could be optional in tense, mood and aspect.

c.	단감	먹어	(4 syllables)
	tan.kam	mʌ.kʌ	
	sweet persimmon	eat	

'(*Agent* is) eating a sweet persimmon.' or '(Is *agent*) eating a sweet pesimmon?'

d.	넌	단감	먹어	(5 syllables)
	nʌ-n	tan.kam	mʌ.kʌ	
	you	sweet persimmon	eat	

'You are eating a sweet persimmon.' or 'Are you eating a sweet persimmon?'

e.	넌	단감	안 먹어	(6 syllables)
	nʌ-n	tan.kam	an mʌ.kʌ	
	you	sweet persimmon	not eat	

'You are not eating sweet persimmon.' or 'Why don't you eat sweet persimmon?'

	내가	단감	먹어	(6 syllables bis)
	nɛ-ka	tan.kam	mʌ.kʌ	
	I	sweet persimmon	eat	

'I am eating a sweet persimmon.' or 'Can I eat a sweet persimmon?'

	너는	단감	먹어	(6 syllables tier)
	nʌ-nɨn	tan.kam	mʌ.kʌ	
	you	sweet persimmon	eat	

'You eat sweet persimmon.' or 'Do you eat sweet persimmon?'

f.	엄마는	단감	먹어	(7 syllables)
	ʌm.ma-nɨn	tan.kam	mʌ.kʌ	
	mum	sweet persiommon eat		

'Mum eats sweet persimmon.' or 'Does mum eat sweet persimmon?'

g. 내 아가는 감 안 먹어 (8 syllables)

nɛ a.ka-nɨn kam an mʌ.kʌ

my baby persimmon not eat

'My baby does not eat sweet persimmon.' or 'Doesn't my baby eat sweet persimon?'

h. 나의 엄마는 단감 먹어 (9 syllables)

na-ɨi ʌm.ma-nɨn tan.kam mʌ.kʌ

my mum sweet persimmon eat

'My mum eats sweet persimmon.' or 'Does my mum eat sweet persimmon?'

i. 아가는 단양단감 안 먹어 (10 syllables)

a.ka-nɨn tan.yang.tan.kam an mʌ.kʌ

baby sweet persimmon of Tan-yang not eat

'Baby does not eat sweet persimmon of Tan-yang.' or 'Doesn't my baby eat sweet persimmon of Tan-yang?'

j. 너의 아가는 단양단감 먹어 (11 syllables)

nʌ-ɨi a.ka-nɨn tan.yang.tan.kam mʌ.kʌ

your baby sweet persimmon of Tan-yang eat

'Your baby eats sweet persimmon of Tan-yang.' or 'Does your baby eat sweet persimmon of Tan-yang?'

k. 너의 아가는 단양단감 안 먹어 (12 syllables)

nʌ-ɨi a.ka-nɨn tan.yang.tan.kam an mʌ.kʌ

your baby sweet persimmon of Tan-yang not eat

'Your baby does not eat sweet persimmon of Tan-yang.' or 'Doesn't your baby eat sweet persimmon of Tan-yang?'

3.2 Acoustic Analysis Results

From this corpus, same methodology used for French whispering analysis was applied for Korean whispering acoustic analysis. In normal voice, Korean has the same F0 contour representing the two modalities.

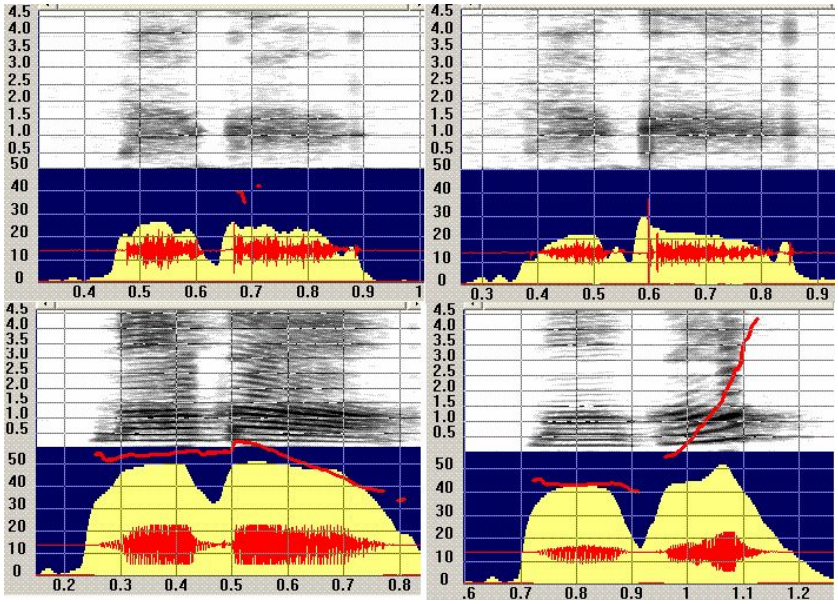

Image 3: Example of Korean production in 2 syllables 'm∧.k∧' (eat):
the 2 upper images are in whispering voice; the 2 lower images are in normal voice
(declarative on the left side and interrogative on the right side).

In normal voice, the F0 rises at the final position: the final syllable of the phrase for interrogative sentences and falls at the same final position for declarative sentences. Considering the duration and intensity, the proportion of each segments inside the sentence seemed to be relatively equal between the two modalities. Also, final lengthening via the last syllable was observed and this accentuation was especially remarkable in long sentences than the short ones for both modalities. A quick view on the proportion of segment duration inside the sentence seemed to be equal between the two modalities; which is common in French.

In Korean whispering, a slightly longer syllable at the end for interrogative sentences than declaratives was found in Korean. Faster speech speed for declarative sentences than interrogatives, which is opposite to French, is also obtained. It means that in Korean, the segments in interrogatives were measured to be longer than the declarative sentence production.

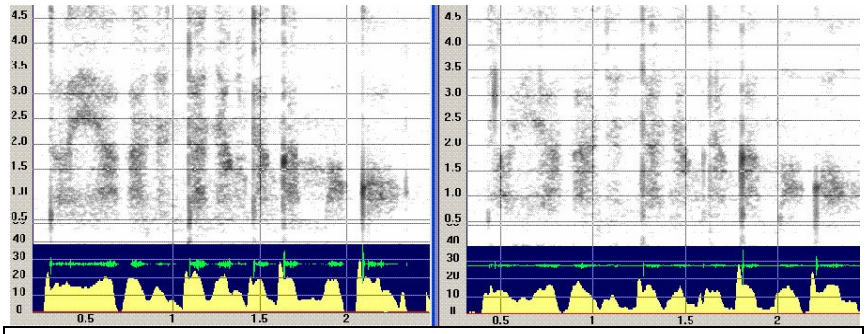

Figure 4 : Example of a long phrase (11 syllables) in Korean whispering voice: « n∧-ɨi a.ka-nɨn tan.yang.tan.kam m∧.k∧ »
The declarative sentence is on left side (*Your baby eats sweet persimmon of Tan-yang*.) and interrogative sentence on the right side (*Does your baby eat sweet persimmon of Tan-yang?*).

Intensity seems to be the most important parameter to distinguish the difference between declarative and interrogative sentences in both languages, but in Korean, higher intensity at the phrasal onset position is measured for declarative sentences. The last 2 syllables' relatively high intensity at the phrasal offset position in declarative sentence was remarkably present. The average of dynamic intensity for the Korean whispering was 5dB, and for normal speech, 4.5dB was measured as dynamic intensity. But in Korean, when the phrase became longer; the difference of dynamic intensity between declaratives and interrogatives became greater.

3.3 Perception in Korean Whispering Voice

This perception test is to identify the modality perception and also verify the perspectives that we proposed as hypothesis at the beginning. The perspective includes the length effects (the number of syllables) on the perceptive recognition rate

of modalities.

For our perception test, the corpus was consisted by 26 stimuli (13 phrases x 2 modalities), one phrase per number of syllables and modality. The stimuli from recordings for acoustic analysis were selected for the test. 10 native listeners for each language participated for the perception test. Each stimulus was 5 times, randomly repeated to listeners and they were asked to mark on an answering sheet; they were forced to choose whether it is a declarative sentence or an interrogative sentence. For the analysis, only the last answer from the 5 was used.

3.4 Perception Test Results in Korean Whispering Voice

In Korean whispering voice, 65% represented the homogeneous response. This rate is not well recognized as French, compared to 75%, but it still shows that the phrasal modality is well recognized in whispering speech. Declarative stimulus showed 64% homogeneous response and 66% for interrogative sentences. According to ANOVA test, the declarative sentences showed significant length effect ($F=5$ and $p<0.001$), even if it doesn't reflects the incrementing correlation between modality perception ratio and number of syllable. On the contrary, the interrogative sentences did not showed the significant length effect on the modality perception ($F=2$ and $p>0.05$)

In Korean, the phrasal modality is perceived uniformly; despite the syllable number; the homogeneous modality perception results are similar. This proposition should be moderated, because some phrase opposes this tendency, and the results seem to be related to the phrase itself, not the syllable numbers. Furthermore, the coefficient of variance does not show homogeneous data distribution (coefficient higher than 15%).

This pilot test permits 2 remarkable points in Korean whispering speech;

- the phrasal modality is globally well perceived in whispering voice,
- the ratio for good perception does not correlates to the number of syllables.

4. Discussion and Conclusion

Through this study on whispering voice, it shows that regardless the absence of F0, intensity and duration are sufficient parameter in modality perception; declarative and interrogative sentences.

In prosodic perspective, characteristic difference in rhythmic structure between Korean and French could be studied further.

In French, as the phrasal length gets longer, the good modality perception ratio increases. It seems the rhythmic factor hold this role; because at least certain number of syllable is necessary to moderate intensity and duration sufficiently, in order to make the listener recognize the modality. In Korean whispering speech, the length did not appear to be relevant. More studies concerning other modalities should be proposed to intensify this perspective. In addition, the two acoustic parameters: duration and intensity does not interrupt in the same way. The intensity is analyzed to be a characteristic acoustic factor between the two modalities, but it appears to be less important in perception.

For further perspectives, our results could be interesting for the ‘cochlear ‘implant domains. It was once reported that the cochlear implanted people facilitates more whispering voice perception than normal voice perception. So the succeeding study that my colleague is thinking of is the synthesised stimuli where the intensity and the duration could be controlled; which is to describe the best application for the cochlear implants.

References

Bonnot J-F.P., « Electromyographical Correlates of Shouted and Whispered Voice », Actes du 11ème Congrès International des Sciences Phonétiques, volume 4, 20-23, Tallinn, Estonie, 1987.

Fonagy, I, Accent et intonation dans la parole chuchotée, Phonetica 20. 177-192, 1969.

Gao M., Tones in whispered Chinese : Articulatory features and perceptual cues, BA, Zhongshan University, 1999.

Harbold. Pitch ratings of voiced and whispered vowels. JASA 30 (7), 600-601, 1958.

Higashikawa, M., K. Nakai, A. Sakakura & H. Takahashi, Perceived Pitch of whispered voice—Relationship with formant frequencies: a preliminary study. Journal of Voice 10.2: 155~158, 1996.

Jensen M.C. Recognition of word tones in whispered speech. Word 14, 186-196, 1958

Mansell, P. An experimental investigation of articulatory reorganization in whispered speech, Working Papers, Institut für Phonetik und spracliche Kommunikation der Universität München Forschungsberichte, 2, 201-253, 1973.

Park M.K., Contribution à l'étude des structures rythmiques en parole lue et spontanée : autour de l'accent focal en coréen et en français, PhD Thesis, University Paris 7 – Denis Diderot, 2000.

Meyer-Eppler, W., Realization of prosodic features in whispered speech. JASA 29: 104~106, 1957.

Tartter V.C What's in whisper? JASA 86 (5), 1678-1683, 1989.

Thomas, I.B., Perceived pitch of whispered vowels. JASA 46 468~470, 1969.

UPRES EA 33 « Atelier de Recherches sur la Parole », UFR Linguistique
Université Paris 7 – Denis Diderot, 2 place Jussieu, 75251 Paris Cedex 05 France
E-mail: vercherand@linguist.jussieu.fr; inyoung.kim@linguist.jussieu.fr; yoo@linguist.jussieu.fr

접수일자: 2006. 05. 13
게재일자: 2006. 06. 12