

ON PHONETIC CHARACTERISTICS OF PAUSE IN THE KOREAN READ SPEECH

Yong-Ju Lee

Department of Computer Engineering, Wonkwang University

Sook-hyang Lee

Department of English Language and Literature, Wonkwang University

344-2 Shinyong-Dong, Iksan, Chonbuk, Korea, 570-749

ABSTRACT

To guarantee a good quality of synthesized speech in a rule-based synthesis, detailed and accurate segmental and suprasegmental rules for a given text are necessary. Especially, much work on the prosodic structure is needed. The goal of this study is to examine roles of syntactic information, intonation pattern, and phrase final lengthening in formation of prosodic units or perception of a phrase boundary. It investigates (a) whether location of (perceived) pause or a phrase boundary, one of the phonetic aspects of prosody, is syntactically determined and (b) how phonetic features such as a silence interval, intonation pattern, and lengthened vowel duration interacts with one another in formation of prosodic units. Radio news of two male announcers were recorded. Perception test of a phrase boundary was performed on the recorded data and at the same time both syntactic and acoustic analyses of them were done. The results showed that perceived pause were accompanied by at least one of the three phonetic features. Many cases were observed where the major syntactic boundary was not accompanied by any of the three phonetic features indicating that syntactic information does not play a crucial role in defining prosodic units. Rather, intonation pattern was proved to play a crucial role in formation of prosodic units.

1. INTRODUCTION

For a rule-based speech synthesis, segmental and suprasegmental rules for given text in a language are necessary and quality of the synthesized speech will depend on how much they are detailed and accurate. Especially, much work on formation of the prosodic units, one of the most controversial issues in phonetics and phonology, is needed for speech synthesis in Korean. Prosodic phonologists emphasize the role of syntactic information in formation of prosodic units [11, 8, 4] while some phoneticians emphasize importance of intonation pattern [9, 5]. One way to weigh the degrees of contribution of syntactic information and intonation pattern to formation of prosodic units is to perform a perception experiment on recorded speech data and see where

listeners perceive a pause or a phrase boundary in a recorded sentence. By acoustic analysis of the same data, we will be able to figure out what kinds of phonetic feature accompany the perceived phrase boundary and see how they interact with one another as cues to the perception of boundaries. Based on a number of production and perception study of syntactic information and a phrase boundary, phonetic cues to the perception of boundaries can be expected to be pure silent interval [3], lengthened vowels at the end of (syntactic) phrases [6, 10] and intonation pattern [1]. This study, as a preliminary study for building computational models of pause assignment in speech synthesis of Korean, focuses on figuring out roles of such phonetic features in perception of phrase boundaries in Korean. It investigates (a) what kinds of phonetic features determine the location of (perceived) pause (b) how the three phonetic information interacts with one another in the perception of a phrase boundary.

2. METHODS

In order to collect the spontaneous read speech data, radio news of two male announcers (A1, A2) were recorded. Eighteen sentences from announcer A1 and sixty four sentences from announcer A2 were recorded. Syntactic and acoustic analyses were done on eighty two recorded sentences. Syntactic hierarchy between constituents of each sentence was examined in relation to location of perceived pauses in it. Acoustic analysis was done using CSL 4300B with 16kHz sampling rate. We examined duration of silence intervals, intonation pattern, and the phrase-final lengthening by displaying recorded data as waveform, pitch contour, and wide-band spectrogram on three windows synchronized with one another. Duration of silence intervals and vowels were measured using both waveform and spectrogram. Silence intervals were divided into two categories, 'long' and 'short' with the boundary between them being 100 ms. (and 150ms. when the following segment is a stop or an affricate in order to exclude the interval of stop closure of the segment.). Analysis of intonation pattern in this paper was based on the prosodic hierarchical structure proposed by a couple of studies [7, 5].

At the same time, perception test of pause was performed to see what kind of phonetic features they used to

perceive a pause in speech. One male and one female graduate students at Wonkwang University served as subjects. They were asked to listen to the same data analyzed above as many times as they wanted and decide the degree of strength of a pause between 'strong' and 'weak'.

3. RESULTS AND DISCUSSION

Generally, sentences were long and syntactically complicated. Twenty one out of eighty two sentences consisted of less than ten words. Various duration of silence intervals were observed within sentences. One hundred and sixteen long silence intervals in total were observed from the eighty two utterances. Long sentences tended to show long silence intervals while only 3 out of 21 short sentences have long ones. Almost 50 % of the long sentences (24 out of 52) did not have a long silence interval until the first half of utterances while they appeared quite frequently in the second half. Length and complexity of subject NP in a long sentence seemed to affect position of a silence interval. In general, when sentences consisted of short NP and long VP, even short silence intervals did not appear at the end of NP, i.e., at the major syntactic boundary. Instead, it appeared somewhere within the following VP (26 out of 32 sentences). For example, if NP consisted of a noun itself or an adjective plus a noun, first silence interval did not appear at the end of NP but somewhere within the following VP. On the contrary, when NP consisted of the relative clause plus the modified subject noun, it tended to appear right after the relative clause or at the end of the NP.

Examination of location of silence intervals with respect to depth in the syntactic hierarchy of sentences showed that silence intervals generally did not appear between constituents that are within the same maximal category low in the syntactic hierarchy, e.g., between constituents within NP or VP close to the bottom of the hierarchy. Generally, in the first half of the sentences, it appeared one node down from the top of the hierarchy while in the second half, it appeared up to three nodes down. Exceptions to this tendency were also observed. As shown in (1), silence interval appeared between constituents that are within the same maximal category low in the hierarchy. Object NP in this sentence is syntactically complicated: it consisted of the relative clause and noun. Silence interval (354 ms.) appeared between constituents within the relative clause that is located low in the hierarchy.

- (1)
 (354ms.) [[ih-chun hwechang-i kachikUpkUm hjEngshikUlo
 'Choon-Lee President-Subj.' 'in the form of pay in advance'
 sampung-kEnsEl-esE ppaenae-n] Relative Clause (310ms.)
 'from Sampoong construction comp.' 'had withdrawn-mod.'
 suEgwEnUi ipchulkUm naeyEk-UI] NP] VP
 'record of deposit and withdrawal of several 10 million Won-Obj.'

Examination of sentences which showed similar pattern to (1) showed that length of the whole sentences were generally long and the maximal category to which those constituents belong was long or syntactically complicated. These results could be interpreted as

an indication that appearance of silence interval within utterances does not entirely depend on syntactic information while it helps us predict not only where silence interval can appear but also where it cannot appear.

In the perception test of pause, two subjects gave the similar response to the long silence interval. Both subjects, in general, identified a long silence as a strong pause as can be seen in Table 1. Subject 1 identified ninety eight out of one hundred and sixteen long silence intervals as a strong pause, twenty six as a weak pause, and two as no pause. Subject 2 identified eighty as a strong pause, thirty a weak pause, and six as no pause.

	Strong pause	weak pause	no pause
subj. 1	98	26	2
subj. 2	80	30	6

Table 1. Response of two subjects to 116 the long silence intervals.

Intonation analysis and duration analysis showed that almost all phrases followed by a long silence interval which were identified as a strong pause ended with a boundary tone (usually L%, HL%, and sometimes H%), and the vowel or syllable at the end of the phrases were long compared to other vowels or syllables in the middle of the phrases. Furthermore, The first segment of the following phrase were fully articulated. For example, stops (phonemically all of them are voiceless in Korean) showed a clear and long silent closure with a strong burst (and long VOT for the aspirated stops).

Phrases followed by a long silence interval but was identified as a weak pause did not end with any boundary tone or/and lengthened vowel. Instead, they ended with an accentual tone (LH- or sometimes L-). Phrase final vowel lengthening seemed not to occur, either, which is consistent with the results from a study[5] that vowel did not seemed to be lengthened at the end of an Accentual Phrase in Korean. Finally, phrases followed by a long silence interval which was not used as a cue to the perception of a phrase boundary were accompanied by different phonetic features from sentence by sentence. Some of the phrases were accompanied by a boundary tone and/or lengthened vowel at the end of the phrase while in other phrases long silence intervals seemed to be due to hesitation or disfluency of the announcer.

There were some cases observed where listeners perceived phrase boundaries even between constituents where long silence intervals did not appear. As shown in Table 2, subject 1 perceived a phrase boundary at sixty eight positions while subject 2 at two positions, showing much disagreement in perceptual judgement of a phrase boundary between the two subjects. Intonational and durational analyses of the data showed that thirty phrases were accompanied by only a boundary tone, two phrases by only a lengthed vowel, nineteen phrases by both a short silence interval and a boundary tone, one phrase by both a short silence interval and a lengthed vowel, eight phrases by

both a boundary tone and a lengthed vowel, and six phrases by all the three cues for Subject 1. For Subject 2, both phrases were accompanied by both a boundary tone and a lengthened vowel. No phrase boundary was observed from either of subjects which was accompanied by only a short silence interval. While almost all the data (65 out of 68 in Subject 1, 2 out of 2 in Subject 2) were accompanied by a boundary tone.

	total	SSI	BT	LV	SSI/ BT	SSI/ LV	BT/ LV	SSI/ BT/ LV
subj .1	68	0	32	19	1	8	6	
subj .2	2	0	0	0	0	2	0	

Table 2. Total number of phrases after which a strong pause were perceived even though they were not accompanied by a long silence interval and number of phrases which were accompanied by a short silence interval, boundary tone, lengthened vowel phrase finally, or more than one of them (except for a long silence interval). *** SSI: short silence interval IP: change in intonation pattern LV: lengthened vowel ***

These results could be interpreted as evidence for a crucial role of intonation pattern to the perception of a pause in utterances. One thing worthy of note is that although a perceived phrase boundary was not followed by a silence interval, phonetic value of the first segment of the immediately following phrase was still fully realized, i.e., not reduced. For example, when the following phrase began with a lenis stop, it was realized as a voiceless stop with a long closure and a strong burst noise instead of becoming voiced or being reduced to the fricative or sonorant, which could be expected if it was surrounded by vowels within an Accentual Phrase in Korean [5].

4. CONCLUSION

We examined roles of syntactic information, intonation pattern, and phrase final vowel lengthening in formation of prosodic units, or perception of a phrase boundary as a preliminary study for building a computational model of pause assignment in speech synthesis of Korean. The results showed that syntactic information did not play a primary role in formation of prosodic units but helped us to predict where a phrase boundary can appear and it cannot appear: Appearance of silence intervals within utterances depended on length and syntactic complexity of the sentences. Generally, short sentences were read without a silence interval. Many cases were observed where silence interval did not appear during the first half of the sentence while it appeared frequently during the second half. Silence intervals did not appear even at the major syntactic phrase boundary, i.e., between the subject noun phrase and the following verb phrase of the sentences. In general, silence interval was not observed between syntactic constituents of the maximal category which were placed deep in the syntactic hierarchy. Almost all the perceived phrase

boundaries were accompanied by one of the boundary tones even when the other phonetic features did not appear, indicating that intonation pattern played a crucial role in defining prosodic units. Only the relationship between the perceived strong pause and its related phonetic features were discussed in this study. Further study is needed on the relationship between the perceived weak pause and its related phonetic features for building a more perfect computational modeling of a pause assignment in synthesis of Korean.

5. REFERENCES

1. W.E. Cooper, J.M. Sorensen, "Fundamental frequency contour at syntactic boundaries," *J. Acoust. Soc. Am.* 62, 683-692, 1977.
2. J.P. Gee & F. Grosjean, "Performance Structures: A psycholinguistic and linguistic appraisal," *Cognitive Psychology* 15, 411-458, 1983.
3. F. Goldman-Eisler, "Pauses, clauses, sentences," *Lang. Speech* 15, 103-113, 1972.
4. D.H. Klatt, "Vowel lengthening is syntactically determined in a connected discourse," *Journal of Phonetics* 3, 129-140, 1975.
5. S.-A. Jun, *The Phonetics and Phonology of Korean Prosody*, Ph.D. Dissertation, The Ohio State University, 1993.
6. S.-h. Lee, "Intonational domains of the Seoul dialect of Korean," paper presented at 117th meeting of the Acoustical Society of America, Syracuse, N.Y., 1989.
7. M. Nespore and I. Vogel, *Prosodic Phonology*, Dordrecht: D. Reidel, 1986.
8. J. Pierrehumbert and M. Beckman, *Japanese Tone Structure*, MIT Press, Cambridge, MA, 1988.
9. D.R. Scott, "Duration as a cue to the perception of a phrase boundary," *J. Acoust. Soc. Am.* 71(4), 996-1007, 1982.
10. E. Selkirk, *Phonology and Syntax: The Relation between Sound and Structure*, MIT Press, Cambridge, MA, and London, England, 1984.