AN EXPERIMENTAL PHONETIC STUDY OF THE INTERRELATIONSHIP BETWEEN PROSODIC PHRASE AND SYNTACTIC STRUCTURE

Cheol-jae Seong & Sanghun Kim***
*Dept. of Linguistics, Chungnam Nat'l University
**Spoken Language Processing Section
Electronics & Telecommunications Research Institute
Tel. +82-42-821-6395, E-mail: scj@zenith.etri.re.kr

ABSTRACT

The boundaries found in the target 100 travelling domain dialogue sentences were labelled automatically according to the relative 9-stepped phonetic depth. The text database was also tagged with syntactic information. Having established 4 kinds of acoustic features, we arranged the prosodic aspect which can be depicted as a continuous change of duration and intonation across the penultimate, boundary, and post-boundary syllables along the X-Y two dimensional scale. Majority of the syntactic pairs seemed to have a characteristic that the intonation tends to fluctuate from rising to falling and, simultaneously, the duration showed of a short-long-short or a long-short-short pattern in the same syllable string of penultimate-boundary-post_boundary.

1. INTRODUCTION

The current speech technology has been aiming to acquire much clearer and more natural synthetic speech sound. The naturalness among the referred two aims can be developed by an adequate phrasing of target sentence, of course, which seems to be strongly related to both syntactic and phonetic aspect simultaneously.

The present study aims to describe the interrelationship between syntactic structure and prosodic phrasing through spontaneous dialogue speech. We also tried to establish a suitable phrasing pattern with respect to the purpose of getting more natural synthetic speech sound. This kind of researches have been progressed with statistical method in the basis of large database in many countries.

The Korean prosodic phrase, here, means a prosodic unit which can be clearly identified as having an evident break boundary at its final position in a sentence in the sense of both perceptual and acoustical viewpoint. The end of each prosodic phrase is, accordingly, marked as the point of major boundary in a sentence.

In order to get statistically valuable data which will be practically available in the coding process, it seems to be needed to get a phonetic database which is composed of at least more than 1,000 sentences.

It may be, however, so difficult not only to discriminate the prosodic boundary through perceptual way but also to expect some coherent result in the time axis when we think of the perceptual limit of humankind. These difficulties, therefore, would lead us to find an effective automatical way to search a prosodic boundary.

After we get a good result from such a process, we may proceed to the next step which will relate the prosodic boundary with the syntactic informations.

From now on, we will propose a methodology on automatic detection of prosodic phrase and do a research plan which relates prosodic unit with syntactic construction in about 100 colloquial Korean dialogue sentences.

2. AUTOMATIC DETECTION OF PROSODIC PHRASE

2.1 Procedure

Using 156 read sentences, which were collected by ETRI, we measured following two aspects from which was made the total data structure. The phonetic database was recorded by a female announcer, which was composed of various text material and mixed each other to get rid of content-interference between sentences.

Recording was performed in sound proof room. The segmentation and labeling were done by CSL 4300 and, specifically, pitch lines were marked in the voiced part pitch-synchronously. The pitch values firstly detected by automatic procedure were corrected by manual on the errors.

① average segment duration
- [medial(+final)]
- discrimination between word final syllables and other than final syllables

Since vowel part may be more sensitive related with durational aspect in universal language, we measured the [medial(+final)], so to speak vowel+coda. The reason why we differentiated between word final syllables and others was that the final lengthening effect might cause durational variation even with the same segment in accordance with the positions.

② Fo measurement on [vowels] in each sentence
- The Fo values were measured on the vowel part with frame_average, in a sentence.
- Using such results, we marked the ratio for each vowel in a sentence to the sentential average.

After measuring the above values, we tried to classify each prosodic boundary automatically as one of the 10 patterns which had been discriminated by the degree of durational lengthening or shortening and that of intonational rising or falling comparative to the reference value. Table 1 shows the whole set of such a 10 patterns.

Label [0] applies to the case that the silent pause in the prosodic boundary exceeds 50 msec

Label [1]-[3] applies to the case that when the durational lengthening is 10 % over the average and the relative pitch ratio falls 10 % below the sentential average and the relative pitch ratio falls 10 % below the sentential average, relatively

level, which means the ratio was included in the section from -10% to +10% of sentential average, and pitch ratio rises 10 % over the sentential average, respectively.
Label [4]-[6] applies to the case that when the durational variation is relatively level, which means that the measured duration is included in the domain from -10% to +10% fo average value and, *again*, the relative pitch ratio falls 10% below the sentential average, relatively level(which means the ratio was included in the section from -10% to +10% of sentential average), and pitch ratio rises 10% over the sentential average. In this case, the conditions for pitch are the same as those in the label [1]-[3].
Label [7]-[9] applies to the case that when the durational shortening is 10% below the average and, *again*, the relative pitch ratio falls 10% below the sentential average, relatively level, and pitch ration rises 10% over the sentential average, respectively. Here, the conditions for pitch are same as above.

[0] pause > 50 msec
(369)
[1] durational lengthening > 10%
intonational falling <10%
(29)
[2] durational lengthening > 10%
intonational level(-10% < pitch < 10%)
(294)
[3] durational lengthening > 10%
intonational rising > 10%
(87)
[4] durational lengthening or shortening(-10% < dur < 10%)
intonational falling < 10%
(13)
[5] durational lengthening or shortening(-10% < dur < 10%)
intonational level(-10% < pitch < 10%)
(144)
[6] durational lengthening or shortening(-10% < dur < 10%)
intonational rising > 10%
(110)
[7] durational shortening < -10%
intonational falling < 10%
(92)
[8] durational shortening < -10%
intonational level(-10% < pitch < 10%)
(514)
[9] durational shortening < -10%
intonational rising > 10%
(392)

Table 1. Whole set of 10 boundary patterns * The number between parentheses means that of occurrence in database.

Following sample is an application of this labeling criteria. The Korean letters were romanized along with the phonetic values. The syntactic depth almost seems to conform to the prosodic labeling depth.
* Juzeonza-neun[9] bueok-e[3] jungyoha-n[8] yoso-igo[0] maeil[2] yeoreo[8] charye[6] sayongha-neun[2] jubang yongpum-ida[0]. (A pot is an important element in kitchen and it has been used a couple of times in a day)

3. THE INTERRELATIONSHIP BETWEEN PROSODIC BOUNDARY AND SYNTACTIC STRUCTURE

3.1. Prosodic and Text Database

The prosodic database was composed of approximately 100 dialogue sentences selected from the ETRI corpus, which was spoken by Korean, Japanese, and Korean interpreter. Among which, the last has been rightly the target for analysis. This is why the synthesizer has a role of interpreter. The A/D conversion has been carried out by Sun sparc 5 workstation under the condition of 16 kHz sampling rate and 16 bit resolution. Segmentation and labeling proceeded to the phoneme level.
On text database, the syntactic informations were marked automatically using the tag sets which has been adopted for ETRI synthesizer. Since the data collected from spontaneous speech data tends to give a lot of difficulty to auto-tag syntactic information, so we manually corrected the resulted errors.

3.2. Prosodic Boundary and Syntactic Construction

3.2.1 Prosodic Boundary

On the target prosodic database, we auto-labeled the boundary information according to the criterion referred from above. The labels from 0 to 2 were classified as major boundary and the rest were unified as minor one. The auto-lable [0], which symbolizes the silent interval, was merged into [lable 1].
With reference to durational aspect, the average value calculated from 156 read database was regarded as the reference.
On Fo manipulation, we calculated the relative Fo ratios for all the vowels involved in the section between long pauses exceeding 150 msec to the mean value in the target area. Be sure to remind that in the case of 156 read database we calculated the mean Fo in a sentence but, here, the section between long pauses were considered.
The following table 2 shows a sample of text database in which the syntactic informations and boundary informations are tagged with. Korean was also romanized and upper case letters between parentheses means the syntactic information.

United(PK)[2] hanggong-imnida(EG).[1]
(This is United Airline)
sam-wol(PN)[8] par-il(PK)[9] gat-eumyeoneunyo(EG+EC)[2]
(About March 8th...)
----------> Long Pause(341 msec)
Jeongo(PN)[3] yeoldu-si(PN)[3] bihaenggi-hago(PJ)[2]
(The flight at noon and...)
----------> Long Pause(616 msec)
Eo(EI)[1] ohu(PN)[9] nesi(PN)[9] bihaenggi-ga(PS)[7] it-seumnida(EG).[2]
(Uhm, there is a flight at 4 p.m.)
Nagoya-eseo(PP)[2] Seoul(PK)[3] gat-eumyeoneun-yo(EG+EC)[3]
(In the case from Nagoya to Seoul...)
----------> Long Pause(463 msec)
sam-wol(PN)[2] sipsa-il(PN)[2]
(March 14th)
----------> Long Pause(189 msec)

Table 2. A sample of dialogue text database tagged with syntactic and boundary information.

3.2.2. On syntactic informations

The analysis was focused to the syntactic informations paired each other at both sides of the major boundary. These are the tag sets which has been used for ETRI synthesizer and our approach simultaneously, which can be divided by 3 main parts, that is, particle, ending, and others. Each symbol stands for the syntactic and slightly semantic meaning, which, as you already observed in table 2, were marked at the word final position in the text database.

1) particle
PT(subjective: eun, neun), PS(subjective: i, ga), PO(objective: eul, reul), PP(locative: eseo), PD(dative: ege), PR(dative: e), PF(locative: ro), PE(commitative: wa, gwa), PC(modifying: majeo, k'aji, jocha, do, ina, irado, inama), PN(adnominal: euy), PJ(conjunctive: hamyeo, hago, ina), PK(nouns arranged...), PV(etc.: instrumental, vocative, comparative, ...)
2) ending
EG(declarative), EQ(interrogative), EI(imperative including interjections without sense), ER(propositive), EX(exclamatory), EJ(subordinating clause), EC(coordinating clause), EN(nominal clause), ED(adnominal clause)
3) others
AD(general adverb), D(determiner: geureon, geu), SA(conjunction), AA(special adverb: jeuksi, aju, ohiryeo,...), N(nouns alone)

Table 3. Tag sets related to syntactic & semantic information
* Korean is romanized.

We also present all the cases of such pairs with the number of occurrences in the table 4.
Table 4 shows the resulted count frequency of two adjacent grammatical indices which were split by prosodic major boundaries. The number between parentheses means the count frequency.

AA*EC(1), AD*EI(7), EC*D(1), EC*ED(3), EC*EI(7), EC*PF(1), EC*PN(1), ED*EI(3), EG*AD(6), EG*EG(4), EG*N(1), EG*PK(2), EG*PN(3), EG*PR(1), EG*PV(1), EG+EC*EI(12), EG+EC*PJ(1), EG+EC*PK(2), EG+EC*PN(5), EG+EC*SA(1), EI*PR(7), EJ*EI(5), EJ*PK(2), N*N(1), PF*EI(8), PF*ER(1), PJ*EI(3), PJ*PN(1), PK*D(1), PK*EQ(2), PK*PK(7), PK*SA(2), PN*EI(7), PO*PD(1), PO*PS(1), PP*PS(1), PR*ED(1), PR*EI(8), PR*PK(1), PR*PN(4), PR*PP(1), PR+EC*EI(2), PS*EI(11), PS*PK(2), PT*EI(13), PT*PC(1), PT*PK(2), PT*PN(3), PT*PR(1), PV*EI(1), SA*AA(1), SA*ED(3), SA*EI(8)

Tabel 4. Syntactic information indices pairs and the number of occurrences

The prosodic information of boundaries can be characterized by the duration and Fo fluctuation in the successive 3 syllables, that is, penultimate, boundary, and post_boundary syllables. The penultimate and boundary syllables are included in the preceding word and post_boundary syllable is naturally a part of the following word. Of course, two adjacent words are split by major boundary, as referred. In table 5, concrete 4 features which can prosodically characterize successive 3 syllables are presented.

Target
..S(penultimate syllable)S(boundary syllable) #(pause)* S(post-boundary syllable)..
① penultimate syllable/ boundary syllable(Fo)
② penultimate syllable/ boundary syllable(duration)
③ boundary syllable/ post-boundary syllable(Fo)
④ boundary syllable/ post-boundary syllable(duration)

Table 5. 4 acoustic features characterizing prosodic boundary
unit: ratio

For example, the first feature can be paraphrase as that ratio for the Fo of penultimate syllable to that of boundary syllable. The implemented prosodic pattern using these features in the successive 3 syllables can be presented as follows.

1) penulti/bnd(Fo:<1) & bnd/post_bnd(Fo:<1)
penulti-bnd-post_bnd: continuous rise
2) penulti/bnd(Fo:>1) & bnd/post_bnd(Fo:<1)
penulti-bnd-post_bnd: fall-rise
3) penulti/bnd(Fo:<1) & bnd/post_bnd(Fo:>1)
penulti-bnd-post_bnd: rise-fall
4) penulti/bnd(Fo:>1) & bnd/post_bnd(Fo:>1)
penulti-bnd-post_bnd: continuous fall

5) penulti/bnd(dur:<1) & bnd/post_bnd(dur:<1)
penulti-bnd-post_bnd: short-long-long
6) penulti/bnd(dur:>1) & bnd/post_bnd(dur:<1)
penulti-bnd-post_bnd: long-short-long
7) penulti/bnd(dur:<1) & bnd/post_bnd(dur:>1)
penulti-bnd-post_bnd: short-long-short
8) penulti/bnd(dur:>1) & bnd/post_bnd(dur:>1)
penulti-bnd-post_bnd: long-short-short

Table 6. Implemented prosodic patterns using 4 acoustic features

Now, we'd like to suggest sort of a convenient diagram to catch the brief idea of coupling the presented prosodic patterns in the boundary with syntactic information. Both prosodic aspect which can be depicted as a continuous change of duration and intonation across the penultimate, boundary, and post-boundary syllables and syntactic informations can be arranged along with the X-Y two dimensional scale simultaneously.
The X-axis refers to 4 prosodic patterns related to Fo and Y-axis means 4 durational features. It can be, therefore, possible to speak that the potential syntactic activity and prosodic activity can be grouped together to represent their own prosodic pattern. Table 3 shows the result.
Majority of the syntactic pairs(coordinates(3.c) in Table 5) seemed to have a characteristic that the intonation tends to fluctuate as rise-fall and, simultaneously, the duration showed of a short-long-short or a long-short-short pattern in the same 3 syllable strings of penultimate-boundary-post_boundary. While we can further suggest a couple of opinions about each coordinate or grammatical pairs for its own purpose, it might be, however, preferable to end at this point because we want to suggest kind of a methodology for the current research.

dur\Fo	1)	2)	3)	4)
5)			SA*AA	
6)				
7)	EG*PR EG+EC*PJ EG+EC*PK PJ*PN	PR*PP	EC*EI EC*PF EC*PN ED*EI EG*AD EG+EC*EI EG+EC*PN EJ*EI EJ*PK PJ*EI PK*EQ PO*PD PO*PS PP*PS PR*EI PR*PN PR+EC*EI PS*PK PT*EI PT*PC PT*PK PT*PN PT*PR PV*EI	
8)	EG+EC*SA PF*ER PK*D PK*SA PR*PK SA*ED	AA*EC EG*EG EG*PK EG*PN EG*PV EI*PR	AD*EI EC*D PF*EI PK*PK PN*EI PR*ED PS*EI SA*EI	EG*N N*N

Tabel 7. Distribution of Syntactic information indices pairs split by major prosodic boundary - The x/y scales are referred as variation of Fo and duration respectively.

4. CONCLUSION

Up to now we tried to establish a proper prosodic phrasing pattern for given text and secondarily to correlate prosodic characteristics with syntactic construction. Of course this is kind of a suggestion to show the possibility of coupling prosodic phenomena with syntactic one in the limited database. The results, therefore, can be corrected if the database would be enlarged. In order to make these works fruitful, the level of Natural language processing should be enhanced much more relative to current state.

5. REFERENCES

[1] Beckman, M. & Hirschberg, J. 1994. "The ToBI Annotation Conventions", manuscript. Ohio State University.

[2] Campbell, N. 1993. "Automatic Detection of Prosodic Boundaries in Speech", Speech Communication 13. 343-354.

[3] Hirschberg, J. 1995. "Acoustic and Prosodic Cues to Speaking Style in Spontaneous and Read Speech", In Symposium on Speaking Style. Proceedings of 1995 ICPhs. Stockholm. Sweden.

[4] Jun, S.A. 1993. The Phonetics and Phonology of Korean Prosody. Ph.D. dissertation. The Ohio State University.

[5] Lee, H.Y. 1990. The Structure of Korean Prosody. Ph.D. Thesis. University College London. University of London.

[6] Nespor, M. and Vogel, I. 1986. Prosodic Phonology. Foris. Dordrecht.

[7] Ross, N. 1995. Modelings of Intonation for Speech Synthesis. Ph.D. dissertation. Boston University.

[8] Seong, C.J. & Hahn,M.S. 1996. "The Prosodic Analysis of Korean Dialogue Speech-through a Comparative Study with Read Speech-", Proceeding of ICSLP 96. Philadelphia. 1037-1040.

[9] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. 1992. "ToBI: a Standard for Labelling English Prosody", Proceedings of ICSLP 92, vol 2. 867-870.

[10] Strangert, E. & Zhi, M.J. 1989. "Pause Patterns in Swedish: A Project Presentation and Some Data", STL-QPSR 1/1989. 27-31.

[11] Wightman, C.W. & Ostendorf, M. 1994. "Automatic Labeling of Prosodic Patterns", IEEE Transactions on Speech and Audio Processing vol.2 - no 4. 469-481.