PROSODY CONTROL BASED ON SYNTACTIC INFORMATION IN KOREAN TEXT-TO-SPEECH CONVERSION SYSTEM

Yeon-Jun Kim, Yung-Hwan Oh

Department of Computer Science, Korea Advanced Institute of Science and Technology
373-1 Kusong-dong Yusong-gu Taejon 305-701 Korea

ABSTRACT Text-to-Speech(TTS) conversion system can convert any words or sentences into speech. To synthesize the speech like human beings do, careful prosody control including intonation, duration, accent, and pause is required. It helps listeners to understand the speech clearly and makes the speech sound more natural.

In this paper, a prosody control scheme which makes use of the information of the function word is proposed. Among many factors of prosody, intonation, duration, and pause are closely related to syntactic structure, and their relations have been formalized and embodied in TTS.

To evaluate the synthesized speech with the proposed prosody control, one of the subjective evaluation methods - MOS(Mean Opinion Score) method has been used. Synthesized speech has been tested on 10 listeners and each listener scored the speech between 1 and 5. Through the evaluation experiments, it is observed that the proposed prosody control helps TTS system synthesize the more natural speech.

1. INTRODUCTION

When human talks about something or reads sentences, the prosodic informations such as tone height, accent and duration are contained in the speech. To make TTS system synthesize the speech like human voice, the prosody control is necessary. The synthesized speech with prosodic information can be understood clearly and sounds more natural.

There are two methods in the prosody control. One method is based on the syntactic analysis, the other one is on the semantic analysis. The latter is a very difficult way, because the lexicon and the complicated natural language processing are needed for the semantic analysis. While the semantic approach is complex and time consuming, the syntactic method is easier.

Especially, the syntactic analysis in Korean is easier than that in English or Japanese, because the function words locate at the end of word-phrases which construct a sentence. A *word-phrase* is a larger grammatical unit than a word, containing a content word and one or more function words. The function words at the end of a word-phrase include the case information, so that we can find out the case of the word-phrase with checking the function word. Also the function word determines the prosody of the word-phrase.

In this paper, we investigated the relations between the function word and the prosody factors relevant to the syntactic structure of the sentence, such as intonation, duration, and pause. And we improved the naturalness of the synthesized speech by using these relations.

2. CHARACTERISTICS OF KOREAN

Korean is Ural-Altaic from the genealogical point of view. One of the common features of Ural-Altaic is that all the Ural-Altaic are agglutinative languages. Like another Ural-Altaic, a Korean sentence is composed of larger grammatical units than word, and we call it word-phrase. A word-phrase is a sequence of words uttered/written as a group starting with a content word, such as noun, verb, adjective, conjunction etc., with or without inflectional endings and attached words, such as auxiliaries, particles and so forth.

우리는	민족중흥의	역사적	사명을	띠고
we	for the national restoration	historic	with the mission	being charged

이 땅에	태어났다.
in this country	were born

(We were born in this country, being charged with the historic mission for the national restoration.)

(a) a sentence

우리는 =	우리	+	는
	we		subjective auxiliary
	(content word)		(function word)

(b) a word-phrase

Figure 1 - An example of Korean sentence

As Figure 1-(a) shows, a Korean sentence consists of several word-phrases. A word-phrase is composed of a content word and some function word as Figure 1-(b). The case of the first word-phrase of the sentence in Figure 1 is subjective. The reason why we can tell the case of word-phrase is not that it is the first one in the sentence, but that its function word is a subjective auxiliary.

Generally, the structure of English sentences is fixed with subject, verb and object/complement in that order. In a Korean sentence, the word order is considerably free, while that is quite strict in English and another European language. Since the word order of Korean is not fixed, the auxiliaries and the endings in the end of word-phrase play an important role in understanding a sentence. We can find out not only the meaning of a sentence but also its structure with the function words. For the reason mentioned above, a man has a tendency to pronounce the function words distinctly. If we look at the pitch contour and the syllabic durations of a sentence, we can see the phenomenon that each function word has its own prosodic pattern.

In this paper, we find out the prosodic patterns according to the function word and make use of them to improve the naturalness of TTS system.

3. PROSODY CONTROL

3.1 Intonation

In Korean, the intonation cannot change the meaning of a sentence, i.e. Korean is not a tonal language, but the various forms of intonation are shown according to the structure of sentence and the emphasis. Therefore, to synthesize the speech which is not clumsy, the intonation control is very important.

In a normal sentence, the declination of intonation is shown in clauses which are composed of a subjective phrase, several objectives or complement phrases and a predicate phrase. In the compound sentence, the baseline-resetting is shown. We can position the boundary of a clause only by checking the ending of a predicate.

As the result, the ending of a predicate, '~다(/da/)', causes the intonation to descend.

Also the ending of a linking, such as '~고(/go/)', '~며(/mj ^/), '~ ㅕ(/j ^/)', which lies at the boundary of clauses in the compound sentence causes the declination of intonation. And the baseline-resetting is shown in the start point of clause following the ending of a linking.

Each word-phrase which is contained in sentences has its own intonation pattern according to its auxiliary. As Figure 2 shows the result of analysis, the patterns keep the baseline before the auxiliary, but at the auxiliary they are divided into three groups, [A] steep ascents, [B] ascent and [C] descent.

The auxiliaries which are contained group [A] are subjective '은/n/', '는/nun/', and modifier '의/ui/'. They make the intonation pattern steep ascent, of which frequency are around 20 Hz higher than that of baseline. Most auxiliaries belong to the group [B] which ranges 10 Hz higher than the baseline. In group [C] which is about 10 Hz lower than the baseline, '(으)로/(u)ro/', a adverbial auxiliary, is contained.

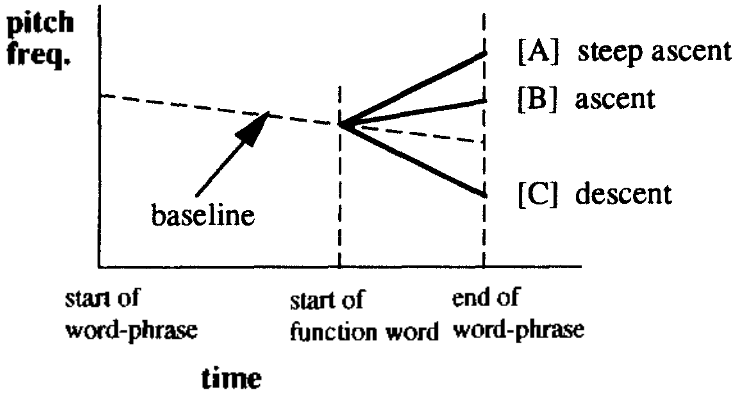


Figure 2 - Intonation patterns according to auxiliary

3.2 Duration

The duration of a word-phrase is determined by the number of syllables forming the word-phrase. Each syllable forming the word-phrase doesn't have the same duration. It is well known that the duration of a syllable at the end of a word is longer than that of initial or medial syllable. The duration of a syllable which is at the end of a word is almost the same as that of an isolated syllable.

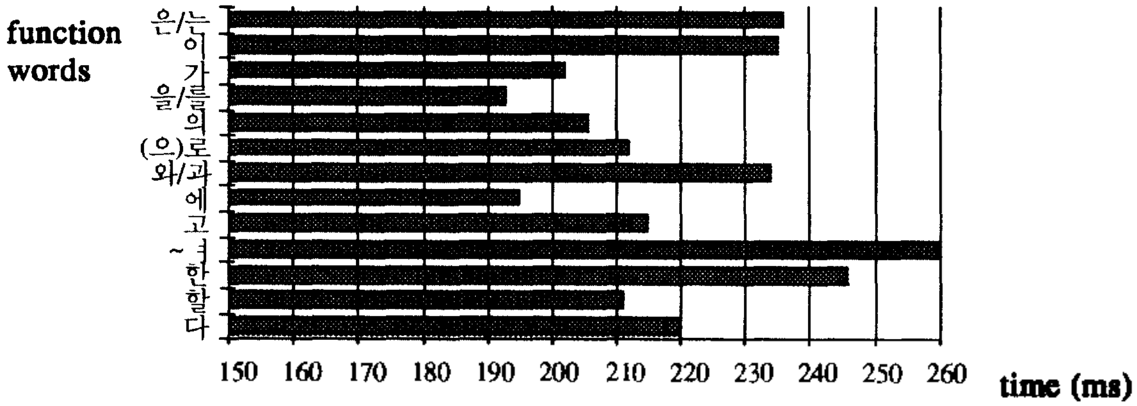


Figure 3 - Durations of each function word

In Korean, this phenomenon, pre-pausal lengthening, is remarkable, because the function word is usually located at the end of word-phrase which is the grammatical unit to utter or write. The duration of the function word at the end of word-phrase may be uttered much longer to clarify the case of word-phrase. Figure 3 shows the means of durations of the function words in sentences which are spoken fluently.

As the results of the analysis, the subjective auxiliary, such as '은/n/', '는/nun/', '이/i/', '가/ga/', has the longest duration. But '가/ga/' starting with the plosive consonant is shorter than other subjective auxiliaries. As compared with the means of duration of a syllable, 150 ms, it is shown that durations of the other function words are much longer.

3.3 Pause

Pauses occur within sentences, especially between word-phrases which are not syntactically related. Longer pauses are frequently seen at the end of an embedded clause or before a prepositional phrase that does not modify the following word. It seems that this difference in lengthening is due to the syntactic structure in the continuous speech.

As the results of the analysis, the duration of a pause depends on its position. When a pause is placed between two sentences, the average duration is about 1,000 ms. When a pause is placed between two clauses, the average duration is 450 ms. When a pause is placed between two word-phrases, the average duration is 90 ms. When a pause is placed in front of a syllable which has a fortis, the duration of pause is lengthen as long as 70 ms. When a pause is placed behind a syllable ending with the stop, such as /p/, /t/, /k/, the duration of pause is lengthen as long as 50 ms.

Table 1 - Durations of pause according to the position

position of pause	duration (ms)
between sentences	1,000
between clauses	450
between word-phrases	90

* begining with fortis : +70, ending with stop : +50

4. EXPERIMENTAL RESULTS

To extract the prosody control rule, we took the standard Korean speech spoken by five men. The extracted prosody control rules were embedded in TTS system and evaluated by listening to the speech synthesized with the TTS system.

There are two methods for evaluating the quality of the speech synthesized by various ways, the subjective evaluation and the objective evaluation. The former measures the subjective quality, including naturalness and ease of listening, whereas the latter measures how accurate phonetic information can be transmitted.

In this paper, one of the subjective evaluation method, mean opinion value (MOS), is used to evaluate the speech synthesized with the prosody control. In the opinion tests, the quality is on the basis of speech uttered by a man and measured by subjective scores (usually by five levels : 5 is excellent, 4 good, 3 fair, 2 poor, and 1 bad). Then MOS is determined by 10 listeners.

Figure 4 shows the sentences used to evaluate the synthesized speech. We have verified the improvement of the quality in synthesized speech by comparing synthesized speeches with and without applying the rules of prosodic information.

A. 우리는 민족 중흥의 역사적 사명을 띠고 이 땅에 태어났다.

B. 조상의 빛난 얼을 오늘에 되살려, 안으로 자주 독립의 자세를 확립하고, 밖으로 인류 공영에 이바지할 때다.

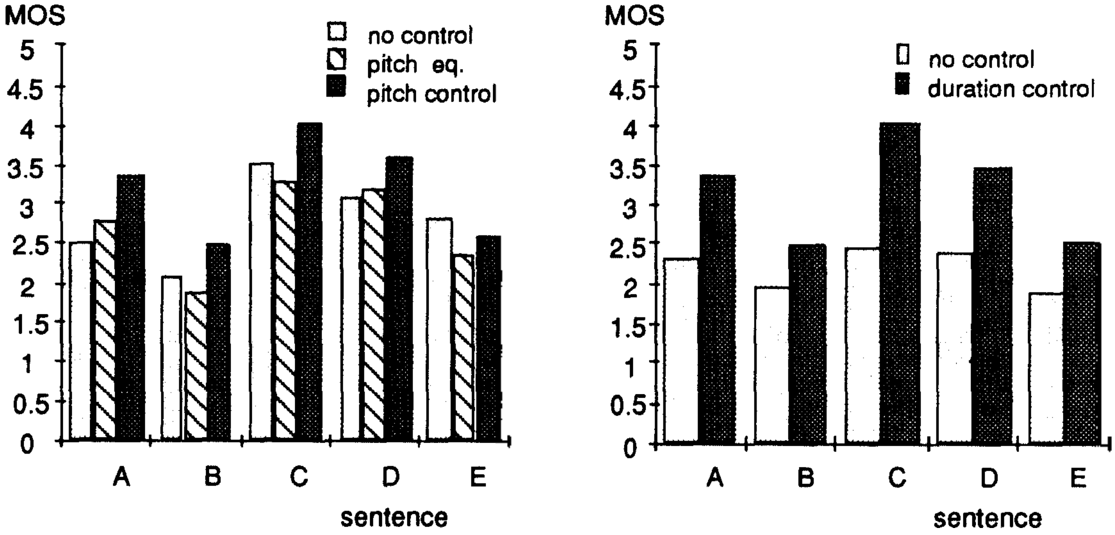
C. 이에, 우리의 나아갈 바를 밝혀 교육의 지표로 삼는다.

D. 공익과 질서를 앞세우며 능률과 실질을 숭상하고, 경애와 신의에 뿌리박은 상부 상조의 전통을 이어받아, 명랑하고 따뜻한 협동 정신을 북돋운다.

E 반공 민주 정신에 투철한 애국 애족이 우리의 삶의 길이며, 자유 세계의 이상을 실현하는 기반이다.

Figure 4 - Sample sentences for test

Through the evaluation experiments, it is observed that the proposed prosody control helps TTS system to synthesize the more natural speech. As Figure 5 shows, the duration control has more effect on the naturalness of speech synthesized. On the other hand, the pause control has less effect on the naturalness of reading sentences.



(a) pitch control (b) duration control

Figure 5 - MOS evaluation results

5. CONCLUSIONS

In this paper, we described the prosody control based on the prosody patterns of function words. By examining Korean speech spoken naturally, we have obtained the prosodic information which consists of intonation, duration and pause. The rules for intonation, duration and pause are extracted by analyzing these prosodic information. The rules of prosodic information are embedded on Korean TTS system to improve the naturalness.

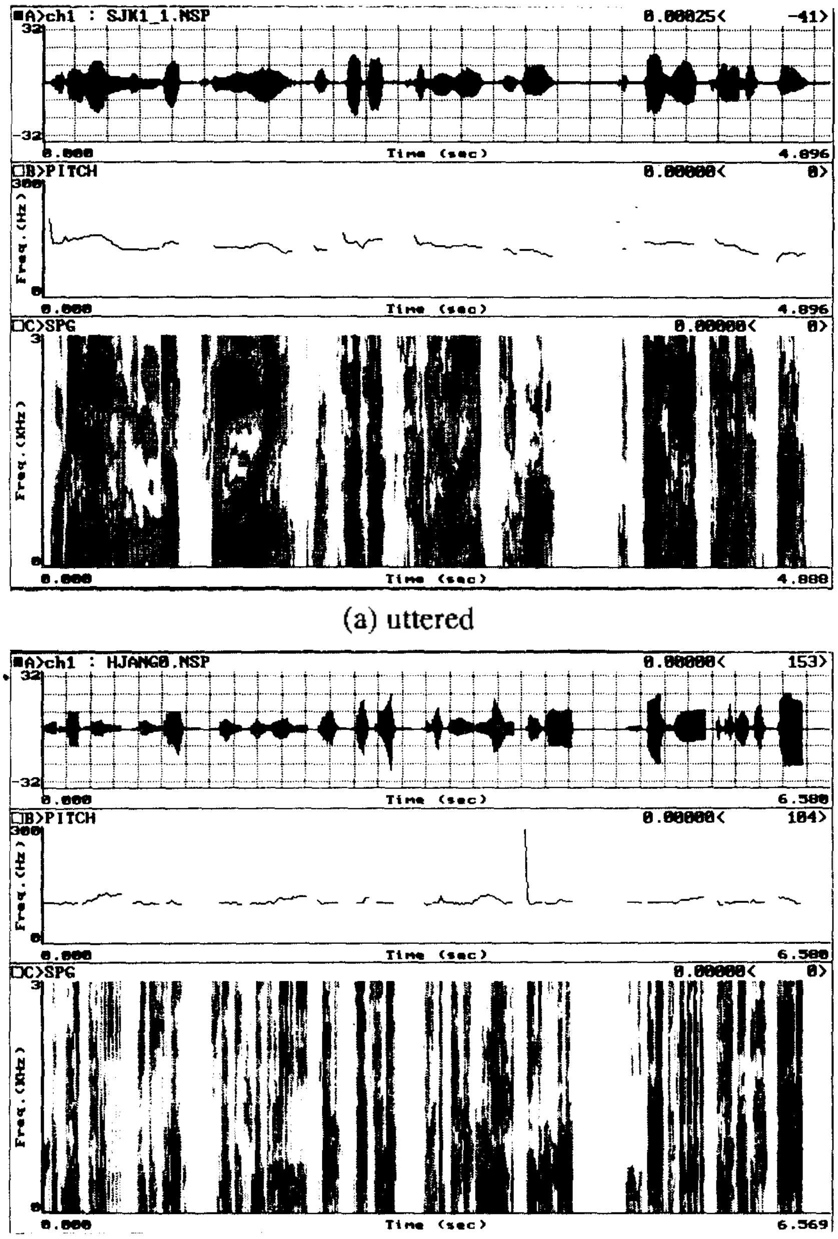
We have verified the improvement of the quality in synthesized speech by comparing synthesized speeches with and without applying the rules of prosodic information. Through the evaluation experiments, it is observed that the proposed prosody control helps TTS system to synthesize the more natural speech.

REFERENCES

1. Yanghee Lee, Sathosi Imai; Extraction and Concatenation of Demisyllables for Allophone in Korean Speech Synthesis-by-Rule, IEICE, pp 1020~1027, Jun. 1988

2. H. W. Kim, Y. K. Lee, K. K. Jung, & S. K. Ahn, A Paser for the Control of Prosody in Korean Text-to-Speech System, Workshop for speech communication and signal processing, pp 218~223, 1992

3. Sangyong Kim, Jungsoo Kim, Phonology and Prosody Control of Speech Synthesis using Syntactic Analysis, Jounal of the Korean Inst. of Telematics and Electronics, pp 508~514, May 1993



(a) uttered

(b) synthesized

Fig 6 - Speech waveforms and spectrograms