Journal of The Korea Society of Computer and Information
Vol. 20, No. 2, February 2015
www.ksci.re.kr
http://dx.doi.org/10.9708/jksci.2015.20.2.149

# 음성 합성 시스템의 품질 향상을 위한 한국어 문장 기호 전처리 시스템

이 호 준*

# Korean Sentence Symbol Preprocess System for the Improvement of Speech Synthesis Quality

Ho-Joon Lee *

## 요 약

본 논문에서는 한국어 문장 기호의 처리를 통해 자연스러운 음성 합성 결과를 생성하는 방법에 대해서 논의한다. 이를 위해 한국어 위키피디아 문서를 분석하여 문장 기호의 사용을 8가지 형태로 분류하고, 11개의 정규표현식 규칙으로 문장 기호를 처리하는 방안을 제시한다. 그 결과 63,000 문장에 대해 56%의 정확도와 71.45%의 재현율을 달성하였으며, 문장 기호 처리 결과를 SSML 기반의 음성 합성 표현으로 변환하여 음성 합성 결과의 품질을 향상시키는 방법을 제안한다.

▸ Keywords : 문장 기호 처리, 음성 합성, 품질 향상, 한국어 문장 기호, 전처리

## Abstract

In this paper, we propose a Korean sentence symbol preprocessor for a SSML (speech synthesis markup language) supported speech synthesis system in order to improve the quality of the synthesized result. After the analysis of Korean Wikipedia documents, we propose 8 categories for the meaning of sentence symbols and 11 regular expression for the classification of each category. After the development of a Korean sentence symbol preprocess system we archived 56% of precision and 71.45% of recall ratio for 63,000 sentences.

▸ Keywords : sentence symbol processing, speech synthesis, quality improvement, Korean sentence symbols, preprocessing

•제1저자 : 이호준 •교신저자 : 이호준
•투고일 : 2015. 1. 20, 심사일 : 2015. 1. 30, 게재확정일 : 2015. 2. 11.
* 영동대학교 스마트IT학과(Department of Smart IT, Youngdong University)
※ This research was partly supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (grant number: NRF-2012R1A1A1013389) and by the IT R&D program of MSIP/KEIT (10044577, Development of Knowledge Evolutionary WiseQA Platform Technology for Human Knowledge Augmented Services).

## I. 서 론

음성은 글로 쓰여진 문장(텍스트, text)을 소리내어 읽은 문장의 청각적 표현이지만, 문장 부호 등에 대한 처리가 제대로 되지 못하면 텍스트로 표현된 문장을 그대로 읽었을 경우 음성 표현이 매우 부자연스러울 수 있다. 예를 들어 아래 예제 (1)의 경우에는 텍스트에 표현된 문장 기호인 홑화살괄호〔1〕를 무시하고 모든 텍스트를 그대로 발화해도 되지만, 동일한 방법으로 예제 (2)를 발화한다면 매우 부자연스러운 음성 표현이 될 수 있다. 특히 예제(3)과 같이 문장 기호를 이용하여 부가적인 정보가 길게 표현된 경우에는 이러한 문제점이 더욱 심각하게 나타난다.

(1) 〈모나리자〉는 레오나르도 다빈치의 대표적인 미술 작품이다.
(2) 컴퓨터의 부품〈예: CPU, RAM, HDD〉은 고온의 열에 매우 취약하기 때문에 컴퓨터의 온도를 적정 수준으로 유지하는 것은 매우 중요하다.
(3) 세종대왕(世宗大王, 1397년 5월 7일 (음력 4월 10일) ~ 1450년 3월 30일 (음력 2월 17일), 재위 1418년 ~ 1450년)은 조선의 제4대 왕이다.

문장 기호는 실제 텍스트에서 매우 빈번하게 나타나는데, 한글 위키피디아에서 무작위로 63,000 문장을 추출하여 분석한 결과 약 40%인 24,945개의 문장에서 하나 이상의 문장 기호가 사용되고 있는 것을 확인할 수 있었다. 데이터 분석 결과 전체 24,945 문장에서 사용된 문장 기호 중에서 예문 (2)나 (3)과 같이 음성 발화를 하는 경우 텍스트에 대한 적절한 처리가 필요한 경우가 약 88%인 것으로 조사되었다. 따라서 범용적인 음성 합성기를 이용하여 위키피디아 문서를 음성으로 변환하는 경우 텍스트에 표현된 문장 기호를 적절히 처리하지 않으면 전체 문장의 1/3 이상은 듣는 사람이 이해하기 어려운 형태가 될 수 있다.

최근 스마트 기기 등에서 음성 표현을 이용한 상호 작용이 활발히 이루어지면서 음성 합성이나 음성 인식 기술이 다양한 분야에 적용되고 있는데〔2, 3, 4, 5, 6, 7〕, 아직까지 텍스트에 표현된 문장 기호의 적절한 처리에 대한 연구는 전무한 실정이다. 또한 글로 표현된 문장을 청각 신호로 변환하는 음성 합성 기술의 경우 시각 장애를 가진 장애인들에게 매우 중요한 의사소통 수단이 되기 때문에 텍스트로 이루어진 문장을 자연스러운 음성으로 변환하는 기술의 개발은 매우 중요하다고 할 수 있다.

본 논문에서는 한국어 텍스트에서 나타나는 문장 기호를 음성 합성에 적합한 형태로 처리하여 음성 합성 결과의 품질을 향상시키는 방안에 대해서 논의한다. 정규표현식을 이용하여 문장 기호 전처리 시스템을 구축한 결과 56%의 정확도(precision)와 71.45%의 재현율(recall)을 달성하였으며, 이러한 문장 기호 처리 결과를 활용하여 자연스러운 음성 합성 결과를 생성하는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2절에서는 음성 합성 결과의 품질을 향상시키는 기존 연구에 대해서 살펴보고, 3절에서는 위키피디아 문서를 대상으로 텍스트에서 나타나는 문장 기호의 유형과 의미를 분석한다. 4절에서는 이렇게 분석된 정보를 활용하여 입력으로 들어온 문장에 존재하는 문장 기호의 유형과 의미를 자동으로 파악하는 문장 기호 처리 시스템에 대해 살펴보고, 5절에서는 본 연구를 통해 개발된 문장 기호 처리 시스템을 음성 합성 시스템에 적용하는 방법에 대해서 다룬다.

## II. 관련연구

한국어 텍스트에서 나타나는 문장 기호의 종류와 사용 방법은 한글 맞춤법 일부 개정안〔1〕에서 자세하게 다루어지고 있는데, 한글 맞춤법 일부 개정안에 따르면 문장 기호는 크게 마침표, 쉼표, 따옴표, 묶음표, 이음표, 드러냄표, 안드러냄표의 7가지로 구분된다. 마침표는 온점(.) 및 고리점(゜), 물음표(?), 느낌표(!)의 세 유형으로 다시 구분되며, 쉼표는 반점(,) 및 모점(、), 가운뎃점(•), 쌍점(:), 빗금(/)의 네 유형으로 구분된다. 따옴표에는 큰따옴표(“ ”) 및 겹낫표(『 』), 작은따옴표(‘ ’) 및 낫표(「 」)가 있고, 묶음표에는 소괄호(( )), 중괄호({ }), 대괄호([ ]) 등이 있다. 이음표는 줄표(—), 붙임표(-), 물결표(~) 등으로 구성되어 있고, 드러냄표에는 드러냄표(ˊ, ˚)와 안드러냄표에는 숨김표(××, ○○), 빠짐표(□), 줄임표(……) 등으로 구성되어 있다. 각 문장 기호의 의미는 한글 맞춤법 일부 개정안에 예제를 포함하여 자세하게 설명되어 있으며 그 의미를 간략하게 정리해보면 강조, 대화, 인용, 기호, 빈자리, 설명, 원어, 연대 등의 8가지로 구분지어 볼 수 있다.

이러한 한국어 문장 기호는 특히 자연언어처리 과정이나 음성 합성 과정에서 중요하게 다뤄져야 하지만, 아직까지 대부분의 시스템에서는 정교한 처리 과정보다는, 단순한 형태의 전처리로 문장 기호를 처리하고 있다. 따라서 음성 합성 시스템의 품질 향상을 위한 연구는 주로 합성 문장의 정확한 운율

구조를 파악하여 자연스러움을 향상시키는 방안에 대한 연구가 대부분이고 문장 기호의 적절한 처리에 대한 연구는 그 중요성에 비해 아직 관련 연구가 미비한 실정이다.

한국어 TTS 시스템을 위한 운율의 트리 기반 모델링[8]에서는 예측하기 쉬운 트리 기반의 구조를 이용하여 한국어 운율 구조에 영향을 미치는 정보를 추출하고, 여기에 bootstrapping aggregation과 born again tree 기술을 적용하여 운율 요소들을 정확히 예측하는 방법을 제안하고 있다. 이러한 과정을 통해 매우 정밀하게 한국어의 운율 구조를 모델링하고, 이를 합성하는 방법을 제시하고 있다.

감정 표현 방법[9]에 관한 연구에서는 자연스러운 음성 합성 결과를 생성하기 위해 음성 합성 결과에 감정을 표현할 때 운율 및 음질의 역학을 분석하고 있다. 이 연구에서는 6명의 발화자에 의해 기쁨, 슬픔, 화남, 공포, 중립의 5가지 감정 상태로 표현된 60개의 데이터를 이용하여 감정에 따른 운율과 음질의 변화를 분석하여 이를 음성 합성 시스템에 적용하는 방법에 대해 논의하고 있다.

HMM기반 한국어 TTS 자연성 향상 연구[10]에서는 대용량 음성 데이터베이스로부터 생성된 tri-phone 정보를 이용하여 복잡한 운율정보를 상승, 평탄, 하강의 억양 변화로 단순화하여 자연스러운 합성결과를 생성하는 방안을 제안하고 있다.

## III. 데이터 분석

텍스트에서 문장 기호에 의해 둘러싸인 내용은 자연스러운 음성 표현을 생성하기 위해 문장 기호만 제거하고 내용은 유지하여 발화하거나 문장 기호와 내용 모두를 삭제하고 발화해야 한다. 앞서 살펴본 예제 (1)의 경우는 문장 기호만 제거하고 내용은 유지하여 발화하는 경우에 해당하고, 예제 (2)와 (3)은 문장 기호와 내용 모두를 삭제하는 경우에 해당한다. 따라서 테스트에서 사용된 문장 기호의 유형은 유지와 삭제의 두 가지로 나누어 볼 수 있다. 또한 최근 발표된 한글 맞춤법 일부 개정안[1]에 따르면 텍스트에서 사용된 문장 기호의 의미는 크게 강조, 대화, 인용, 기호, 빈자리, 설명, 원어, 연대 등의 8가지로 구분지어 볼 수 있다. 한글 맞춤법 일부 개정안의 문헌 자료를 분석한 결과 강조, 대화, 인용, 기호, 빈자리의 다섯 가지 의미는 주로 유지의 유형과 연관성이 높고, 설명, 원어, 연대의 세 가지 의미는 주로 삭제 유형과 연관성이 높은 것으로 분석되었다.


그림 1. 문장 기호 분석 워크벤치
Fig. 1. Sentence Symbol Annotation Workbench

본 연구에서는 한국어 위키피디아 문서를 대상으로 텍스트에서 나타나는 문장 기호의 유형과 의미를 분석하였다. 이를 위해 HTML5와 JavaScript를 이용하여 문장 기호의 유형과 의미를 기술(annotation)할 수 있는 워크벤치(그림 1)를 개발하였다. UTF-8로 인코딩 된 텍스트 문서를 워크벤치를 통해 읽으면 웹 브라우저에 해당 문서의 내용이 표현되며, 마우스를 클릭하여 문장 기호로 표현된 부분을 선택할 수 있다. 문장 기호로 표현된 부분을 선택하면 유형으로는 삭제(remove)와 유지(keep) 중 하나를 선택할 수 있으며, 의미는 원어(ori), 연대(date), 설명(exp) 중 하나를 선택하거나 대화(conv), 인용(quot), 강조(emp), 기호(symb), 빈자리(blnk) 중 하나를 선택할 수 있다. 만약 선택한 영역이 이러한 8가지 기본 의미 구조 외에 다른 의미라고 판단되면 Comment 항목에 해당 내용을 직접 입력할 수 있다.

문장 기호 분석 워크벤치를 이용하여 기술된 정보는 XML과 유사한 형태로 저장되는데 그 결과는 그림 2와 같다. 그림 2의 첫 번째 결과는 소괄호로 이루어진 내용을 분석한 것으로 유형은 제거(remove)이며 의미는 원어(ori)임을 나타낸다.

은퇴한 후 1962년에서 1969년까지 8년 동안 미네소타 구단에서 일한 그는 1962년~1964년에는 스카우터를, 1965년에서 1968년 6월 중순까지는 3루 베이스 코치를, 그리고 1968년 남은 시즌에는 미네소타의 AAA 팀인 덴버 베어스<tag-3 label="remove ori">(Denver Bears)</tag-3>을 맡았다.

영국은 1793년 <tag-14 label="keep emp">'제1차 대프랑스 동맹'</tag-14>을 결성하여 프랑스에 선전 이후 프랑스 사이에 전쟁 상태에 있었다.

그림 2. 문장 기호 분석 결과
Fig. 2. Sentence Symbol Annotation Result

또한 두 번째 결과는 작은 따옴표로 이루어진 내용을 분석한 것으로 유형은 유지(keep)이며 의미는 강조(emp)임을 나타내고 있다. 문장 기호를 분석하는 과정에서는 태그가 중첩될 수 있기 때문에 표준 XML과는 다르게 각각의 태그에 ID를 결합하여 표현하였다.

이와 같은 방법으로 6명의 분석가(annotator)가 총 63,000 문장의 한국어 위키피디아 문서에 주석(annotation)을 달았으며, 24,945 문장에서 하나 이상의 문장 기호가 사용되고 있는 것으로 분석되었다. 이는 전체 문장 중 39.6%의 문장은 적어도 하나 이상의 문장 기호를 포함하고 있다는 것을 의미하며 63,000 문장에서 나타난 모든 문장 기호에 대해서 유형을 기준으로 분류하면 표 1과 같다.

표 1. 기호 유형에 따른 분류 결과
TABLE 1. Classification Result based on Symbol Types

| 유형 | 개수 | 비율 |
|---|---|---|
| 유지 | 3,147 | 12.26% |
| 삭제 | 22,510 | 87.74% |
| 총 문장 기호 | 25,657 | 100% |

문장 기호의 중첩 사용 정도를 살펴보면, 전체 25,657개의 문장 기호에서 단독으로 문장 기호가 사용된 경우는 25,090개였고, 2중 기호는 565개, 3중 기호는 2개로 나타났다. 특히 3중으로 표현된 문장 기호는 전부 컴퓨터 프로그램의 일부로, 3중 표현은 일반적인 상황에서는 거의 발생하지 않는 것으로 분석되었다. 유지로 분류된 문장 기호의 의미를 살펴보면 표 2와 같고, 삭제로 분류된 문장 기호의 의미를 살펴보면 표 3과 같다.

표 2. 유지 기호의 의미 분류 결과
TABLE 2. Semantics of Symbols Annotated as Keep

| 유지 의미 | 개수 | 비율 |
|---|---|---|
| 강조 | 1,952 | 62.03 |
| 인용 | 168 | 5.34 |
| 기호 | 144 | 4.58 |
| 설명 | 51 | 1.62 |
| 빈자리 | 35 | 1.11 |
| 대화 | 29 | 0.92 |
| 연대 | 4 | 0.13 |
| 기타 | 764 | 24.28 |
| 총 개수 | 3,147 | 100.00 |

표 3. 삭제 기호의 의미 분류 결과
TABLE 3. Semantics of Symbols Annotated as Remove

| 삭제 의미 | 개수 | 비율 |
|---|---|---|
| 설명 | 11,538 | 51.26 |
| 원어 | 8,347 | 37.08 |
| 연대 | 1,981 | 8.80 |
| 기타 | 644 | 2.86 |
| 총 개수 | 22,510 | 100.00 |

표 4. 유형별 문장 기호 출현 빈도
TABLE 4. Frequency of Symbols with Different Types

| 유지 | | 삭제 | |
|---|---|---|---|
| 기호 | 개수 | 기호 | 개수 |
| ' ' | 2,290 | ( ) | 21,288 |
| " " | 361 | 〔〔 〕〕 | 309 |
| ( ) | 196 | 〔 〕 | 246 |
| 〔 〕 | 39 | ( ) | 81 |
| 〔〔 〕〕 | 35 | ' ' | 61 |
| { } | 30 | {{ }} | 61 |
| < > | 7 | { } | 32 |
| ‘ ’ | 6 | (' ') | 17 |
| 《 》 | 6 | " " | 16 |
| ≪ ≫ | 5 | ( ) | 8 |

총 25,657개의 문장 기호 중 유지와 삭제의 유형으로 많이 나타난 문장 기호를 빈도순으로 각각 10개씩 뽑아서 정리해보면 표 4와 같다. 한글 맞춤법 일부 개정안에서 설명하고 있는 문장 기호와 비교해볼 때 겹낫표(『 』)나 홑낫표(「 」)는 위키피디아 문서에서 나타나지 않았고 대신 두 개 이상의 문장 기호가 결합되어 사용되는 경우를 확인할 수 있었다. 또한 동일한 작은 따옴표나 소괄호도 표 4에서와 같이 다양한 형태로 나타나는 것을 확인할 수 있었다.

표 4의 내용을 표 1에서 정리한 전체 문장 기호에 대한 비율로 분석해보면 그림 3과 같다. 그림 3을 살펴보면 유지의 경우 첫 번째 기호인 작은 따옴표가 전체 문장 기호의 73%를 차지하고 있으며, 다섯 번째 기호인 중괄호까지 포함하면 전체의 93%, 그리고 10개의 기호까지 모두 포함하면 전체의 95%를 포함하는 것으로 분석되었다. 삭제의 경우에는 첫 번째 기호인 소괄호가 전체 문장 기호의 95%를 차지하고 있으며, 10개의 기호를 모두 포함하면 전체의 98%를 포함하는 것으로 분석되었다.

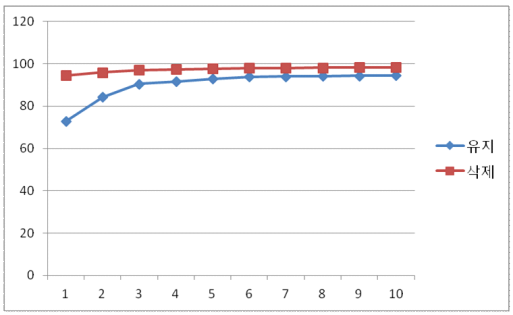


그림 3. 문장 기호 분석 결과
Fig. 3. Sentence Symbol Annotation Result

본 연구에서는 문장 기호의 의미를 원어, 연대, 설명, 대화, 인용, 강조, 기호, 빈자리의 8가지로 설정하였으며, 각 의미에서 주로 나타나는 상위 3개의 문장 기호 패턴을 정리해보면 표 5와 같다.

표 5. 의미별 문장 기호 출현 빈도
TABLE 5. Frequency of Symbols with Different Semantics

| 의미 | 유형(개수) | 패턴 | 개수 | 커버리지 |
|---|---|---|---|---|
| 원어 | 삭제 (8,347) | ( ) | 8,227 | 98.56 |
| | | { } | 68 | 99.38 |
| | | ' ' | 17 | 99.58 |
| 연대 | 유지 (4) | {{ }} | 2 | 50.00 |
| | | ' ' | 1 | 75.00 |
| | | ( ) | 1 | 100.00 |
| | 삭제 (1,981) | ( ) | 1,927 | 97.27 |
| | | {{ }} | 33 | 98.94 |
| | | { } | 18 | 99.85 |
| 설명 | 유지 (51) | ( ) | 26 | 50.98 |
| | | ' ' | 16 | 82.35 |
| | | {{ }} | 3 | 88.24 |
| | 삭제 (11,538) | ( ) | 10,869 | 94.20 |
| | | {{ }} | 271 | 96.55 |
| | | { } | 159 | 97.93 |
| 대화 | 유지 (29) | ' ' | 27 | 93.10 |
| | | ‘ ’ | 1 | 96.55 |
| | | " " | 1 | 100.00 |
| 인용 | 유지 (168) | ' ' | 149 | 88.69 |
| | | ‘ ’ | 15 | 97.62 |
| | | ( ) | 2 | 98.81 |
| 강조 | 유지 (1,952) | ' ' | 1,593 | 81.61 |
| | | ‘ ’ | 274 | 95.65 |
| | | ( ) | 25 | 96.93 |
| 기호 | 유지 (144) | ( ) | 97 | 67.36 |
| | | ' ' | 21 | 81.94 |
| | | { } | 14 | 91.67 |
| 빈자리 | 유지 (35) | ( ) | 21 | 60.00 |
| | | ‘ ’ | 7 | 80.00 |
| | | { } | 3 | 88.57 |

표 5에서 각 의미별로 상위 3개의 기호가 해당 의미의 대부분을 차지하고 있음을 알 수 있다. 이는 소수의 정교한 규칙으로도 문장 기호 대부분의 의미를 파악할 수 있다는 가능성을 보여준다.

## Ⅳ. 문장 기호 처리 시스템

문장 기호가 실제 텍스트에서 어떤 의미로 사용되었는지에 대한 정확한 판단을 위해서는 문장 기호가 어떤 통사적 역할(syntactic role)을 하고 있는지와 어떤 의미(semantics)를 내포하고 있는지를 분석해야 한다. 이를 위해서는 일반적으로 입력 문장에 대한 형태소 분석, 구문 분석, 의미 분석의 자연언어처리 과정이 순차적으로 수행되어야 한다.

그러나 구문 분석을 위해서는 먼저 문장 기호로 표현된 내용이 문장 성분(constituent)으로 사용되는 것인지 아닌지에 대한 판단이 이루어져야 한다. 앞서 살펴본 예제 (1)의 경우에는 문장 기호만 제거하면 문장 기호 내부의 내용을 하나의 문장 성분으로 처리할 수 있지만 예제 (2)나 (3)의 경우에는 문장 기호 내부의 내용을 유지한 상태로는 구문 분석이 불가능하다.

따라서 문장 기호의 의미 파악 및 처리 과정은 구문 분석 이전 단계에서 이루어져야하기 때문에 본 연구에서는 문장 기호의 처리를 자연언어처리의 전 단계에서 수행하고자 한다. 그러므로 문장 기호의 전처리 과정에서는 텍스트에 표현된 문장 기호와 문장 기호 내부에서 사용된 어휘의 일반적 특성만을 이용하여 문장 기호의 유형과 의미를 분석하고자 한다.

3절의 데이터 분석 내용을 바탕으로 원어, 연대, 설명, 대화, 인용, 강조, 기호, 빈자리에 대한 대상 문장 기호와, 문장 기호 내부에서 사용된 어휘의 일반적인 특성을 정리해보면 다음 표 6부터 표 13까지와 같다.

표 6. 강조의 특성 및 정규표현식
TABLE 6. Characteristics of Pattern Emphasis

| 의미 | 강조 |
|---|---|
| 예제 | '황제 직할령' |
| 문장 기호 | {}, {{}}, ◇ , « », 《》 , ‘ ’, “ ”, 「 」, 『 』 |
| 내부 내용 | 2음절 이상 2어절 이하 |
| 정규표현식 | {\{ 《« 《“「『]\S{2,}(\s\S+)?{\}〉»》”」』} |

표 7. 대화의 특성 및 정규표현식
TABLE 7. Characteristics of Pattern Conversation

| 의미 | 대화 |
|---|---|
| 예제 | '그럼 제가 카즈하 누나랑 결혼해도 되는거죠?' |
| 문장 기호 | ‘ ’, “ ”, 「 」, 『 』 |
| 내부 내용 | 3어절 이상<br>문장 마지막에 !, ?, . 등 구두점 존재 |
| 정규표현식 | {“「『]\S+\s\S+\s\S{^!?\.}*{!?\.}{”」』} |

표 8. 인용의 특성 및 정규표현식
TABLE 8. Characteristics of Pattern Quotation

| 의미 | 인용 |
|---|---|
| 예제 | '민주화운동 기념 사업회법' |
| 문장 기호 | ‘ ’, “ ”, 「 」, 『 』 |
| 내부 내용 | 3어절 이상<br>문장 마지막에 구두점 없음 |
| 정규표현식 | {“「『]\S+\s\S+\s\S{^!?\.}*{”」』} |

표 9. 기호의 특성 및 정규표현식
TABLE 9. Characteristics of Pattern Symbol

| 의미 | 기호 |
|---|---|
| 예제 | (P) |
| 문장 기호 | {}, (), ◇, « », 《》 |
| 내부 내용 | 1음절 어휘 존재 (예: (1), (ㄱ) 등) |
| 정규표현식 | {\(\( «《]\S{\}\)\〉»》} |

표 10. 빈자리의 특성 및 정규표현식
TABLE 10. Characteristics of Pattern Blank

| 의미 | 빈자리 |
|---|---|
| 예제 | ( , ) |
| 문장 기호 | {}, {}, () |
| 내부 내용 | 괄호 안에 아무것도 없음<br>괄호 안에 공백 하나 존재<br>괄호 안에 공백, 쉼표, 공백이 연속으로 존재<br>괄호 안에 공백, 쉼표, 공백, 쉼표, 공백이 연속으로 존재 |
| 정규표현식 | {\({\(]\s?{\}}\)}<br>{\({\(]\s,\s{\}}\)}<br>{\({\(]\s,\s,\s{\}}\)} |

표 11. 설명의 특성 및 정규표현식
TABLE 11. Characteristics of Pattern Expression

| 의미 | 설명 |
|---|---|
| 예제 | (복잡한 체계의 전체는, 단지 각 부분의 기능의 총합이 아니라 각 부분을 결정하는 통일체라는 관점) |
| 문장 기호 | {}, (), ◇ |
| 내부 내용 | 3어절 이상 |
| 정규표현식 | {\({\(]\S+(\s\S+)?{\}\)〉} |

표 12. 원어의 특성 및 정규표현식
TABLE 12. Characteristics of Pattern Original

| 의미 | 원어 |
|---|---|
| 예제 | (李天君) |
| 문장 기호 | {}, {}, (), ◇, « », 《》, ‘ ’, “ ”, 「」, 『』 |
| 내부 내용 | 한글, 숫자 외에 다른 언어로 시작 |
| 정규표현식 | {\({\(]〈« 《“「『]〔^0-9ㄱ-ㅎㅏ-ㅣ가-힣〕+{\}}\)〉»》”」』} |

표 13. 연대의 특성 및 정규표현식
TABLE 13. Characteristics of Pattern Date

| 의미 | 연대 |
|---|---|
| 예제 | (2000~2001.10.14.) |
| 문장 기호 | {}, {}, (), ◇, « », 《》, ‘ ’, “ ”, 「」, 『』 |
| 내부 내용 | 문장 시작, 시작 다음 어절에서 숫자와 년으로 시작<br>문장 시작, 시작 다음 어절에서 숫자와 월로 시작<br>문장 시작, 시작 다음 어절에서 숫자와 일로 시작<br>문장 시작, 시작 다음 어절에서 숫자와 시로 시작<br>문장 시작, 시작 다음 어절에서 숫자와 분으로 시작 |
| 정규표현식 | {\({\(]〈« 《“「『]\d+{년월일시분}{\}}\)〉»》”」』}<br>{\({\(]〈« 《“「『]\S+\s\d+{년월일시분}{\}}\)〉»》”」』} |

표 6부터 표 13까지의 내용을 살펴보면, 3절에서 분석된 데이터를 이용하여 각 패턴에서 사용된 문장 기호를 분석하였고, 문장 기호의 내부 내용을 추상적으로 표현하여 이를 정규표현식으로 변환하였다. 이때 각각의 정규표현식은 겹치는 부분이 없도록 조정되었다.

이렇게 만들어진 정규표현식 기반 규칙을 이용하여 3절에서 분석된 모든 데이터에 대해 각 의미를 예측한 결과는 표 14와 같다.

표 14. 전처리 결과
TABLE 14. Preprocessing Results

| 의미 | precision | recall | F-score |
|---|---|---|---|
| 강조 | 7.21 | 65.47 | 13.0 |
| 대화 | 18.7 | 68.97 | 29.42 |
| 인용 | 11.56 | 40.48 | 17.98 |
| 기호 | 10.0 | 52.38 | 16.79 |
| 빈자리 | 37.78 | 97.14 | 54.4 |
| 설명 | 43.92 | 70.69 | 54.18 |
| 원어 | 64.42 | 96.3 | 77.2 |
| 연대 | 45.48 | 90.58 | 60.56 |
| **전체** | **56.0** | **71.45** | **62.79** |

이 때 정확도(precision), 재현율(recall), F-score는 일반적인 의미와 동일하게 사용되고 있는데, 정확도는 문장 기호 시스템에 의해 예측된 결과 중에서 올바른 답을 예측한 비율을 의미하고, 재현율은 올바른 답 중에서 예측된 비율을 의미하며, F-score는 정확도와 재현율의 조화 평균을 의미한다. 따라서 문장 기호 처리 시스템에서 해당 의미만을 예측하도록 규칙을 매우 협소하고 한정적으로 제시한다면 정확도는 거의 100%에 도달할 수 있겠지만, 규칙 생성에 사용하지 않은 입력에 대한 처리(재현율)는 거의 0%가 될 수 있다. 따라서 문장 기호 처리 시스템에서는 정확도와 재현율이 모두 중요한 요소라고 볼 수 있으며, F-score가 시스템 성능의 좋은 판단 기준이 될 수 있다.

8개 항목 전체에 대한 정확도는 56%이고 재현율은 71.45%였으며 F-score는 62.79로 3절에서 예상한 바와 같이 적은 수의 규칙만으로도 대부분의 의미 파악이 가능하다는 것을 보여주고 있다. 강조와 대화, 인용 및 기호는 재현율에 비해 정확도가 낮게 나타났는데, 특히 강조, 대화, 인용은 3절의 데이터 분석 과정에서도 정확한 의미 구분이 어려웠던 만큼 이들에 대한 명확한 가이드라인의 설정이 요구된다.

## V. 문장 기호 처리 기반 음성 합성 시스템

일반적인 텍스트-음성 변환(Text-to-Speech, TTS) 시스템은 입력 문장을 음성 합성 결과로 변환하기 때문에 음성

합성 결과에 특별한 운율 정보(prosodic information; 음의 높낮이, 세기, 길이, 휴지 등)를 추가하거나 수정 하는 것이 거의 불가능하다.

그렇지만 TTS 시스템이 SSML(Speech Synthesis Markup Language)의 입력을 지원한다면, XML 형태의 표준화된 표현 방법을 이용하여 음성 합성 결과에 추가적인 운율 정보를 표현할 수 있다. SSML은 W3C에서 제안하는 음성 합성기를 위한 XML 기반의 마크업 언어로 VoiceXML의 일부로 활용되기도 하며 애플의 음성 명령이나 Microsoft의 SAPI와 유사한 형태를 보인다. SSML을 통해서 제어할 수 있는 정보는 emphasis(강조), break(휴지), prosody(운율), voice(목소리) 등이 있는데 해당 정보가 사용된 예제는 아래 그림 4와 같다.

```
<?xml version="1.0"?>
<speak xmlns="http://www.w3.org/2001/10/synthesis"
       xmlns:dc="http://purl.org/dc/elements/1.1/"
       version="1.0">
  <metadata>
    <dc:title xml:lang="en">Telephone Menu: Level 1</dc:title>
  </metadata>

  <p>
    <s xml:lang="en-US">
      <voice name="David" gender="male" age="25">
        For English, press <emphasis>one</emphasis>.
      </voice>
    </s>
    <s xml:lang="es-MX">
      <voice name="Miguel" gender="male" age="25">
        Para español, oprima el <emphasis>dos</emphasis>.
      </voice>
    </s>
  </p>

</speak>
```

그림 4. SSML 입력 예제
Fig. 4. SSML Input Example

4절에서 개발한 문장 기호 처리 시스템을 통해 입력된 문장에 대해서 각 문장 기호 패턴에 적합한 음성 합성 결과를 생성하는 방법을 살펴보면 아래 그림 5와 같다.

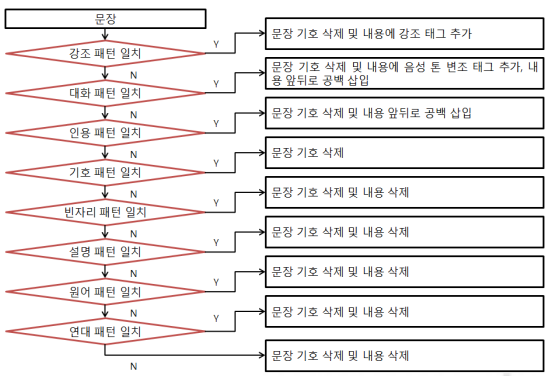


그림 5. TTS 시스템 일반도
Fig. 5. General TTS Architecture

강조 패턴으로 인식된 내용에는 SSML의 강조 태그를 삽입하여 다른 부분에 비해 강하게 음성 합성 결과를 생성하게 되며, 대화 패턴의 경우에는 대화문 앞뒤에 충분한 휴지(pause)와 함께 대화문을 적절한 성별로 변환하여 읽어주고, 인용의 경우에는 인용문 앞뒤에 충분한 길이의 휴지를 삽입하게 된다. 기호 패턴의 경우에는 문장 기호만 삭제하고, 문장 기호 내부의 내용은 음성 합성 결과에 포함시키며, 그 외의 패턴들은 음성 합성 결과 생성 시 자연스러움을 저해하는 요소가 되므로 문장 기호와 내용을 모두 삭제하여 음성 합성 결과를 생성하게 된다.

지금까지 문장 기호 처리 시스템을 이용하여 이를 음성 합성 시스템에 적용하는 방안에 대해서 다루었는데, 실제로 문장 기호가 포함된 문장에 대해서 적절한 문장 기호 처리를 통해 SSML 문서를 생성한 결과를 살펴보면 그림 6과 같다.

```
입력 문장:
핫토리 모친은 범인의 자살을 용감하게 막아내고, '당신 다시 살인할 생각이에요?
자살은 자신을 죽이는 살인이에요.'라고 설득했다.

SSML 출력:
<?xml version="1.0"?>
<speak version="1.1" xmlns="http://www.w3.org/2001/10/synthesis"
       xml:lang="ko-KR">
  <p>
    <s xml:lang="ko-KR">
      <voice name="junwoo" gender="male">
        핫토리 모친은 범인의 자살을 용감하게 막아내고,
      </voice>
      <break/>
      <voice name="sujin" gender="female">
        당신 다시 살인할 생각이에요? 자살은 자신을 죽이는 살인이에요.
      </voice>
      <break/>
      <voice name="junwoo" gender="male">
        라고 설득했다.
      </voice>
    </s>
  </p>
```

그림 6. 문장 기호 처리 SSML 결과
Fig. 6. SSML Result of Sentence Symbol Processing

## Ⅵ. 결론

본 논문에서는 한국어 위키피디아 텍스트에서 나타나는 문장 기호를 처리하여 음성 합성 결과의 품질을 향상시키는 방안에 대해서 논의하였다. 이를 위해 총 63,000 문장의 한국어 위키피디아 문서를 분석하여 문장 기호의 사용 패턴 및 의미를 분석하였으며, 이를 정규표현식으로 변환하여 문장 기호 전처리 시스템을 구축하였다. 또한 이렇게 만들어진 문장 기호 전처리 시스템을 한국어 음성 합성기와 연동하여 자연스러운 음성 합성 결과를 생성할 수 있는 방안을 제시하였다.

11개의 정규표현식을 이용하여 문장 기호를 8가지 형태의 의미로 분석한 결과 56%의 정확도와 71.45%의 재현율을 기록하였으며, 분석된 문장 기호의 의미에 적합한 형태로 SSML 문서를 생성하여 음성 합성기의 입력으로 사용하였다.

현재 시스템에서는 강조, 대화, 인용에 대한 처리가 다소 미흡한 것으로 분석되었는데, 향후 연구로 이들을 명확하게 구분할 수 있는 가이드라인의 제시 및 이를 기반으로 시스템의 성능 향상을 진행할 예정이다.

## REFERENCES


[1] Revised Guidelines on Korean Orthography, Ministry of Culture, Sports and Tourism, 2014.

[2] Jin-Hyung Kim, So-Young Park, "Rule-based Speech Recognition Error Correction for Mobile Environment," Journal of the Korea Society of Computer and Information, vol. 17, no. 10, pp. 25-33, October 2012.

[3] Gyeongyong Heo, Woo-Young Jang, Jun-Pyo Park, "Digital Doorlock with Voice Recognition," Proceedings of the Korean Society of Computer Information Conference, pp. 269-270, July 2012.

[4] Seong Jin Cho, Seongho Lee, Sungyoung Lee, "Design of Emotion Recognition system utilizing fusion of Speech and Context based emotion recognition in Smartphone," Proceedings of the Korean Society of Computer Information Conference, pp. 323-324, July 2012.

[5] Kee-Beak Kim, Jong-Ho Choi, "Contents Navigation System using Speech Recognition," KSCI Review, vol. 15, no. 1, pp. 99-102, June 2007.

[6] Myung-Hun Kim, Chi-Geun Lee, In-Mi So, Sung-Tae Jung, "Design and Implementation of a Bimodal User Recognition System using Face and Audio," Journal of the Korea Society of Computer and Information, vol. 10, no. 5, pp. 353-362, November, 2005.

[7] Jin-Koo Ji, Sung-Il Yun, "Design and Implementation of Speaker Verification System Using Voice," Journal of the Korea Society of Computer and Information, vol. 5, no. 3, pp. 91-98, September 2000.

[8] Sangho Lee, Yung-Hwan Oh, "Tree-based modeling of prosodic phrasing and segmental duration for Korean TTS systems," Speech Communication, vol. 28, no. 4, pp. 283-300, 1999.

[9] Sang-Min Lee, Ho-Joon Lee, "How to Express Emotion: Role of Prosody and Voice Quality Parameters," Journal of the Korea Society of Computer and Information, vol. 19, no. 11, pp. 159-166, November, 2014.

[10] Gi-Jeong Lim, Jung-Chul Lee, "Improvement of Naturalness for a HMM-based Korean TTS using the prosodic boundary information," Journal of the Korea Society of Computer and Information, vol. 17, no. 9, pp. 75-84, September 2012.


## 저 자 소 개


**이 호 준**

2001: 한국과학기술원 전산학과 공학사

2003: 한국과학기술원 전산학과 공학석사

2010: 한국과학기술원 전산학과 공학박사

현 재: 영동대학교 스마트IT학과 조교수

관심분야: 자연언어처리, 정보 추출, 감정 음성 합성, 한국어처리

Email : hjlee@webmail.yd.ac.kr