

Logatom을 사용한 문서음성변환 시스템

조관선, 이철희

연세대학교 전기·컴퓨터공학과

Text-to-Speech System Using Logatom

Kwansun Cho and Chulhee Lee

Dept. of Electrical and Computer Engineering, Yonsei Univ.

E-mail: chulhee@bubble.yonsei.ac.kr

요약

본 논문에서는 logatom 기반 무제한 한국어 TTS 시스템 구현을 제안한다. 이를 위하여 한국어를 대표할 만한 문서코퍼스를 선택하여 분석하고 이를 바탕으로 합성에 필요한 logatom을 설계한다. 일반적으로 음성코퍼스를 통해 음성세그먼트를 추출하여 접속에 기반한 TTS 시스템에서는 음성세그먼트를 의미있는 단어 또는 어절로부터 추출한다. 하지만 음성세그먼트 추출시 고려되는 사항은 합성단위에 기초한 음소간의 결합형태이므로 본 논문에서는 음성세그먼트 추출을 위하여 무의미한 음소열인 logatom을 설계한다. Logatom은 문장세그먼트의 어절내 위치와 문서코퍼스 분석 결과 얻어진 음소간의 결합형태를 기반으로 설계된다. 제안된 시스템의 합성음질을 평가하기 위하여 CVC 기반 logatom을 사용하여 임의의 문장을 합성해 본 결과 대부분의 음성세그먼트 접속이 자음에서 이루어지고 어절의 위치를 고려한 logatom 설계로 인하여 어절 내에서는 비교적 자연스러운 합성음을 얻을 수 있었다.

1. 서론

텍스트를 음성으로 변환해 주는 TTS 시스템에서 합성음질의 향상을 위하여 다양한 종류의 음성코퍼스의 구축의 필요성이 날로 더해가고 있다 [1]. 일반적으로 TTS 시스템 구현을 위해서는 대량의 음성코퍼스로부터 합성에 사용되는 음성세그먼트를 검색하여 추출하고 저장하여야 하는데 이 작업은 매우 까다롭고 시간이 많이 걸리는 작업이다. 또한 의미있는 단어나 어절로 이루어진 음성코퍼스를 무제한 합성 시스템에 사용할 경우 합성에 필요한 모든 음성세그먼트를 포함하는 음성

코퍼스를 제작하기란 쉽지 않다. 본 논문에서는 이러한 문제를 해결하기 위하여 logatom 기반 음성코퍼스의 사전설계를 제안한다. Logatom이란 음성세그먼트 추출을 위한 무의미한 음소열로 문서코퍼스의 통계적 분석 결과를 바탕으로 결합형태와 어절의 위치를 고려하여 만들어 진다 [2]. 문서코퍼스를 통한 합성단위 총수의 예측은 무제한 합성에 필요한 logatom만으로 이루어진 양질의 코퍼스 제작을 가능하게 한다.

본 논문의 구성은 다음과 같다. 서론에 이어 2장에서는 제안된 코퍼스기반 TTS 시스템의 대략적인 구조를 살펴보고 3장에서는 문서코퍼스의 통계적 분석결과 얻어진 음소간 결합확률을 제시한다. 4장에서는 3장에서 제시한 결합확률과 문장세그먼트의 어절내 위치를 고려하여 CVC 기반 logatom을 설계하고 5장에서는 설계된 logatom을 사용하여 임의의 실험문장들을 합성해 보고 결과를 고찰한다. 마지막으로 6장에서 결론을 맺는다.

2. 코퍼스기반 TTS 시스템

그림 1과 같이 일반적인 코퍼스 기반 문서음성변환 시스템의 대략적인 구조는 처리내용에 따라 크게 코퍼스분석부, 전처리부, 언어처리부, 음성합성부로 나눌 수 있다 [3].

2.1. 코퍼스 분석부

합성음질의 기본이 되는 음성코퍼스를 녹음, 분석하고 무제한 합성을 위한 음성세그먼트와 자연성 향상을 위한 운율파라미터를 추출하여 저장한다. 본 논문에서는 음성코퍼스 녹음을 위하여 설계된 logatom 목록을 사용한다.

2.2. 전처리부

한글 텍스트에서 나올 수 없는 영어, 단위, 문장부호, 숫자, 약어 등과 같은 한글 이외의 문자를 한글로 변환한다.

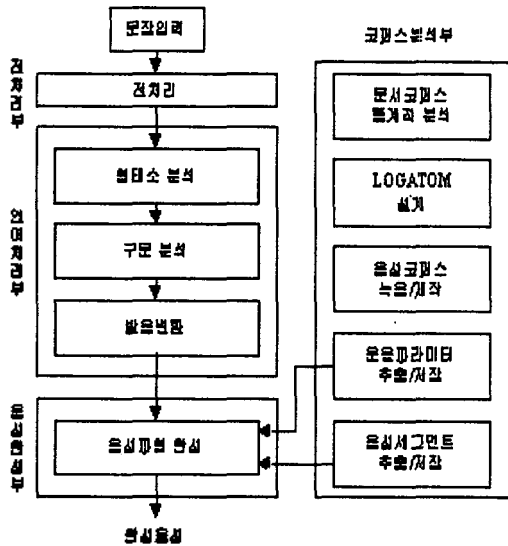


그림 1. 코퍼스 기반 TTS 시스템.

2.3. 언어처리부

언어처리부는 입력된 한글 텍스트 문자의 품사를 분류하는 형태소분석부, 출력음성의 운율효과를 만들어 내기 위한 정보와 한국어 문장에서 어절간의 문법적인 관계정보를 제공하는 구문분석부, 그리고 입력된 문장을 소리나는 대로 변환시켜주는 발음변환부로 나눌 수 있다 [4].

2.4. 음성합성부

실제 합성파형을 만드는 부분으로 음성세그먼트를 접속하고, 합성음의 품질을 높이기 위하여 언어처리부와 코퍼스분석부에서 추출된 운율정보를 바탕으로 피치 조절이나 휴지부 삽입 등의 과정이 이루어 진다.

3. 출현빈도 기반 음소별 결합형태

Logatom 기반 무제한 한국어 TTS 시스템 구현을 위한 logatom으로 구성된 양질의 음성코퍼스 제작을 위해서는 한국어 음소 및 음소열을 통계적으로 분석하여야 한다. 본 논문에서는 한국어를 대표할 만한 문서코퍼스로 연세대학교 언어정보연구개발연구원에서 사전 편찬을 위하여 수년간 구축한 균형말뭉치인 '연세말뭉치 I, II'를 선택하고 분석하였다 [5][6]. 문서코퍼스는 약 A4 용지 11,700장 분량의 텍스트로 이루어져 있다. 출현빈도를 기반한 결합형태는 다음 3.1-3.5절에 나타

낸 것과 같다. 이것은 문서코퍼스 전반에 걸쳐 출현빈도를 조사한 결과로 각각의 음소 및 음절과 가장 빈번하게 결합되는 음소 및 음절을 선택하여 목록을 작성했다. 이 목록은 logatom 설계시 참조될 것이다.

3.1. 중성자음과 가장 빈번하게 결합되는 초성자음의 예 (발생빈도에 의해 정렬)

ㄱ+ㅈ=ㄱㅈ, ㄴ+ㄷ=ㄴㄷ, ㄷ+ㅌ=ㄷㅌ,
 ㄹ+ㄹ=ㄹㄹ, ㄱ+ㄴ=ㄱㄴ, ㅂ+ㅍ=ㅂㅍ,
 ㅇ+ㅎ=ㅇㅎ

3.2. 중성자음과 가장 빈번하게 결합되는 CV형/V형의 예 (발생빈도에 의해 정렬)

ㄷ+ㄱ=ㄷㄱ, ㄴ+ㄴ=ㄴㄴ, ㅌ+ㄹ=ㄹㄹ,
 ㄱ+ㅁ=ㄱㅁ, ㅌ+ㅂ=ㄷㅂ, ㄷ+ㅇ=ㄷㅇ

3.3. 초성자음과 가장 빈번하게 결합되는 중성모음의 예 (발생빈도에 의해 정렬)

ㄱ+ㅏ=가, ㄱ+ㅑ=까, ㄴ+ㅏ=나, ㄷ+ㅏ=다,
 ㅌ+ㅏ=따, ㄹ+ㅏ=로, ㄱ+ㅑ=마, ㅂ+ㅑ=부,
 ㅃ+ㅑ=뿌, ㅅ+ㅑ=서, ㅆ+ㅑ=쏘, ㅇ+ㅣ=이,
 ㅈ+ㅣ=지, ㅉ+ㅣ=찌, ㅊ+ㅣ=치, ㅋ+ㅣ=카,
 ㅌ+ㅣ=타, ㅍ+ㅣ=피, ㅎ+ㅣ=하

3.4. 초성자음과 가장 빈번하게 결합되는 중성자음의 예 (발생빈도에 의해 정렬)

ㄴ+ㄱ=ㄴㄱ, ㄱ+ㄱ=ㄱㄱ, ㄴ+ㄴ=ㄴㄴ, ㄴ+ㄷ=ㄴㄷ,
 ㄷ+ㅌ=ㄷㅌ, ㄹ+ㄹ=ㄹㄹ, ㅇ+ㅁ=ㅇㅁ, ㅇ+ㅂ=ㅇㅂ,
 ㄱ+ㅃ=ㄱㅃ, ㄴ+ㅅ=ㄴㅅ, ㄱ+ㅆ=ㄱㅆ, ㄴ+ㅈ=ㄴㅈ,
 ㄱ+ㅉ=ㄱㅉ, ㅇ+ㅊ=ㅇㅊ, ㄴ+ㅋ=ㄴㅋ, ㄴ+ㅌ=ㄴㅌ,
 ㄹ+ㅍ=ㄹㅍ, ㅇ+ㅎ=ㅇㅎ

3.5. 초성자음과 가장 빈번하게 결합되는 CV형/V형의 예 (발생빈도에 의해 정렬)

하+ㄱ=하ㄱ, 바+ㄱ=바ㄱ, 하+ㄴ=하ㄴ, 니+ㄷ=니ㄷ,
 어+ㅌ=어ㅌ, 그+ㄹ=그ㄹ, 하+ㅁ=하ㅁ, 지+ㅂ=지ㅂ,
 기+ㅃ=기ㅃ, 거+ㅅ=거ㅅ, 이+ㅆ=이ㅆ, 가+ㅈ=가ㅈ,
 어+ㅉ=어ㅉ, 모+ㅊ=모ㅊ, 시+ㅋ=시ㅋ, 부+ㅌ=부ㅌ,
 아+ㅍ=아ㅍ, 대+ㅎ=대ㅎ

3.6. CV형/V형 음절과 가장 빈번하게 결합되는 중성자음의 예 (발생빈도에 의해 정렬)

가+ㄴ=간, 까+ㄱ=각, 나+ㄴ=난, 다+ㅇ=당,
 따+ㅇ=땅, 로+ㄱ=록, 마+ㄴ=만, 부+ㄴ=분,
 뿌+ㄴ=뿌, 서+ㅇ=성, 쓰+ㄹ=쓸, 이+ㄴ=인,
 지+ㄴ=진, 찌+ㄹ=찹, 치+ㄴ=친, 카+ㄴ=칸,

타+ㄹ=탈, 피+ㄹ=필, 하+ㄴ=한

4. CVC 기반 logatom 설계

4.1 CVC 기반 문장세그먼트

logatom 설계시 고려한 문장세그먼트의 종류는 표 1과 같다. 표 1에서 h, b, t는 문장세그먼트의 어절내 위치를 나타내는 것으로 h는 어절의 처음, b는 어절의 중간, t는 어절의 끝을 나타낸다. CC형은 종성자음-초성자음을, CV형은 초성자음-중성모음을, CVC1형은 초성자음-중성모음-초성자음을, 그리고 CVC2형은 초성자음-중성모음-중성자음을 나타내는 문장세그먼트이다.

표 1. CVC 기반 문장세그먼트.

음소개수	문장세그먼트	단위수
2개	hCtV	249
	hCbV	318
	bCbV	331
	bCtV	308
	bCbC	95
3개	hCVbC1	3,179
	bCVbC1	4,074
	hCVtC2	808
	hCVbC2	1,048
	bCVbC2	1,123
	bCVtC2	1,006
	총단위수	12,539

본 논문에서는 동일한 형태의 음성세그먼트라 할지라도 어절내 위치의 변화에 따른 음소의 발음 변화를 반영하기 위하여 logatom 설계를 위한 문장세그먼트를 2종류 또는 4종류로 다시 세분화하였다. 표 1에 나타난 총 단위수는 문서코퍼스내 문장세그먼트의 출현빈도를 조사하고 실제 문서에서 사용되는 문장세그먼트의 수를 합한 결과이다. 표 1에서 보는 것처럼 CVC 기반 무제한 한국어 TTS 시스템 구현을 위해서는 총 12,539개의 음성세그먼트가 필요함을 알 수 있다.

4.2 Logatom 설계

문장세그먼트가 합성문장의 어절의 처음, 중간 또는 끝에 위치할 경우를 따로 생각하여 logatom을 설계한다. 어절의 처음에 오는 문장세그먼트 중의 hCVC1형의 경우를 예로 들면 logatom 설계방법은 다음과 같다.

- hCVbC1형 문장세그먼트에서 C1 자리에 오는 초

성자음과 가장 자주 결합되는 중성모음을 조사한다.

- 합성문장에서 hCVC1형 문장세그먼트로 합성할 음절의 중성유무를 조사한다.
- 중성이 없으면 hCVbC1형 문장세그먼트와 처음에 조사한 중성모음을 결합한 2음절을 logatom으로 사용한다.
- 중성이 있으면 처음에 조사 결과인 초성자음 C1과 중성모음으로 이루어진 음절과 가장 자주 결합되는 중성자음을 조사한 후 이를 hCVbC1형 문장세그먼트와 결합하여 logatom으로 사용한다.

5. 실험 및 결과

5.1. 실험내용

본 논문에서 설계된 logatom을 기반으로 실제 무제한 한국어 문서음성변환 시스템을 위한 logatom 기반 음성코퍼스의 녹음, 제작에 앞서 임의의 문장들을 선택하여 합성음질평가 실험을 수행하였다. 실험에 사용한 문장들은 한국어 발음변환법칙과 형태별 음성세그먼트를 고려하여 선택되었고 선택된 실험문장은 다음과 같다.

<실험문장>

1. 발이 크지 않아 일은 쉽게 끝났다.
2. 급한 일이 생겨서 급행열차를 타고 갔다.
3. 사람은 누구나 행복하게 살기를 원합니다.

표 2. 발음변환된 실험문장과 사용된 logatom.

	실험문장	logatom
1	바치 크지 아니 이른 쉽게 곤난 따.	바치, 모치, 크지, 가 지, 아나, 하나, 이륙, 그른, 쉽스, 업까, 적 계, 곤당, 는난, 는난 따, 언따, 언따
2	그판 이리 생겨서 그행열차를 타고 갔다.	그필, 아판, 이로, 그 러, 생하, 정가, 는거 서, 거서, 그필, 아평 한, 열로, 롤치, 정차 륙, 그를, 타가, 하고, 갔다, 언따, 언따
3	사라본 누구나 행 보카게 살기를 원 합니다.	사로, 그라만, 하른, 누 가, 하구나, 하나, 행 하, 정부, 정보카, 시카 가, 하계, 살로, 를가, 는기록, 그를, 원당, 는 한, 정함나, 금나, 는니 다, 니다

표 2는 발음변환된 실험문장과 합성에 필요한 음성 세그먼트 추출을 위하여 사용된 logatom 목록이다.

발성화자는 20대 남성이며 실험을 위한 logatom은 11.025 KHz의 샘플링 주파수와 16비트의 정밀도로 양자화하여 저장되었다. 합성음의 자연성 향상을 위하여 실험문장을 각각 10번씩 발음하고, 발음된 문장을 바탕으로 발성화자 고유의 운율파라미터(휴지부, 지속시간 등)를 추출하여 저장하였다.

5.2. 실험결과

설계된 logatom 목록으로 구성된 음성코퍼스로부터 음성세그먼트를 추출하여 실험문장을 합성한 결과 실험문장의 각 어절별 합성음의 명료성 및 자연성은 어느 정도 확보할 수 있었다. 하지만 본 논문에서는 logatom의 길이를 2음절 또는 3음절로 제한함으로써 합성문장에 각각의 logatom 고유의 음의 높낮이가 그대로 반영되어 합성문장 전체의 발음속도가 느려졌다. 또한 본 논문에서는 어절의 중간에 오는 음절의 초성에 음가없는 'ㅇ'이 올 경우 'ㅇ'을 기준으로 음성세그먼트를 재분할하여 이 과정에서 이루 음절간 천이구간이 유실되어 합성음질이 저하된다. 그림 2는 logatom을 사용하여 합성된 음성파형을 보여준다.



(a) 발이 크지 않아 일은 쉽게 끝났다.



(b) 급한 일이 생겨서 급행열차를 타고 갔다.



(c) 사람은 누구나 행복하게 살기를 원합니다.

그림 2. 합성음성파형

6. 결론

본 논문에서는 무제한 TTS 시스템 구현을 위한 음성코퍼스 녹음, 제작에 앞서 문서코퍼스의 통계적 분석을 바탕으로 무제한 합성에 필요한 합성단위 총수를 예측하고 음소간의 결합확률을 고려한 무의미한 음소열 logatom을 설계함으로써 양질의 음성코퍼스 제작의 가능성을 제시하였다. 또한 logatom을 어절내 위치에 따라 다르게 설계함으로써 어절의 위치변화에 따른 음소간의 발음변화를 반영하여 어절내에서는 비교적 자연스러운 합성음을 얻을 수 있었다. 하지만 본 논문에서는 단순히 문장 내의 띄어쓰기를 바탕으로 어절을 나누고 logatom의 길이를 2음절 또는 3음절로 제한함으로써 인하여 합성문장 전체의 발음속도의 저하를 가져왔다. 이를 해결하기 위해서는 표기상의 띄어쓰기를 기준으로 어절을 분리하여 logatom을 설계하는 것이 아니라 보통사람이 긴 문장을 발음할 때 끊어 읽는 '억양구'를 기준으로 좀 더 긴 logatom을 설계하여야 한다. 결론적으로 보다 고품질의 무제한 문서음성변환 시스템 구현을 위해서는 문서코퍼스를 기반으로 한 정확한 언어처리기술, 특히 구문분석이 이루어져야 한다.

참고문헌

- [1] 김재홍, "고품질 한국어 음성 합성을 위한 문서 음성 변환 시스템," 연세대학교 석사논문, 1998년.
- [2] Thierry Dutoit, *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, Dordrecht, 1997.
- [3] 안수길 외, *한국어 음성합성 및 인식 시스템 구현*, 서울대학교 공학연구소, 1994년.
- [4] 연세대학교 한국어학당, *한국어발음*, 연세대학교 출판부, 1995년.
- [5] 이상섭, "말뭉치: 그 개념과 구현," 사전편찬학연구 5, 6집, pp. 7-28, 1996년.
- [6] 이상섭, "뭉치 언어학의 기본 전제," 사전편찬학연구 5, 6집, pp. 29-60, 1996년.