

가변합성단위를 사용한 문서 음성 변환 시스템

Text-to-Speech System Using Variable Synthesis Units

조 관 선, 이 철 희

연세대학교 공과대학 전자공학과

Kwansun Cho and Chulhee Lee

Dept. of Electronic Engineering, Yonsei University

요 약

본 논문에서는 자연스러운 음성을 합성하기 위해 가변합성단위를 사용한 합성시스템을 제안한다. 음소나 diphone과 같은 작은 단위를 사용하는 기존의 시스템은 음성세그먼트 연결시 접속점이 많아지는 단점이 있다. 반면에 단어나 복합음소와 같이 큰 단위를 사용할 경우 접속점의 수가 감소하여 음질이 향상되지만 단위수 증가로 무제한 합성이 어려워진다. 이러한 문제를 해결하기 위하여 본 논문에서는 접속점의 수를 줄이고 적절한 크기의 메모리로 향상된 음질을 얻기 위한 방법으로 어절 및 CVC와 같은 큰 단위와 반음절과 같은 작은 단위를 선택적으로 사용하여 음성을 합성한다. 실험은 특정문장을 대상으로 각각 반음절, CVC로 합성한 음성과 이들을 어절과 혼합하여 합성한 음성을 비교하였으며 그 결과 가변단위를 사용하여 합성한 음성이 비교적 자연스러움을 알 수 있었다.

I. 서 론

음성합성 시스템에서 중요한 요소 중 하나는 합성 단위이다. 기존의 음성합성 시스템에서 널리 사용되는 합성단위로는 음소, 반음절, 음절, diphone, 복합음소 등이 있다 [1]. 음소나 반음절을 합성단위로 사용할 경우 적은 양의 메모리로 무제한 합성 시스템을 만들 수 있지만 음소 상호간의 조음현상을 반영해 줄 수 없는 단점때문에 합성음질의 저하를 피할 수 없다 [2]. 이와 반대로 길이가 긴 음절이나 복합음소를 사용할 경우에는 합성단위내에 음소 상호간의 천이구간이 비교적 많이 포함되어 합성음질은 향상되나 합성단위의 총수가 급격하게 증가하여 음소나 반음절에 비해 훨씬 많은 메모리량을 필요로 한

다 [3]. 따라서 고정된 합성단위만을 사용하여 고품질의 무제한 합성 시스템을 구성하는 것은 한계가 있다.

본 논문에서는 위에서 언급한 합성단위별 장점을 살려 가변합성단위를 제안한다. 가변합성단위를 사용한 합성시스템은 어절, CVC, 반음절을 선택적으로 사용하여 문장을 합성한다. 이러한 시스템은 어절이나 CVC와 같은 긴단위 사용에서 오는 상호조음효과를 합성문장에 보다 많이 반영하여 합성음질을 향상시킬 수 있고, 반음절과 같이 작은 단위를 사용함으로써 무제한 어휘 합성 시스템 구현이 가능하게 된다. 합성단위의 선택적 사용을 위해 본 시스템에서는 음성 코퍼스 분석을 통해 사전에 단위목록(unit inventory)을 구성한다. 단위목록을 구성하는 음성세그먼트들은 코퍼스의 문맥정보와 운율정보에 근거하여 선택한다.

본 논문 II장에서는 가변합성단위를 사용한 시스템의 전체 구성을 살펴보고 III장에서는 음성 코퍼스 분석을 통해 구성되는 단위목록설계에 관하여 기술한다. 그리고 IV장에서는 실험 및 결과를 분석하고 마지막으로 V장에서는 결론을 맺는다.

II. 가변합성단위 기반 음성합성시스템

제안된 시스템의 구성은 그림 1과 같이 전체 시스템은 코퍼스 분석부, 언어처리부, 음성합성부로 나눌 수 있다.

2.1 코퍼스 분석부

단위목록(unit inventory)을 생성하는 부분으로 준비된 대량의 음성데이터베이스(corpus)로부터 가변단위 합성에 필요한 어절, CVC, 반음절의 음성세그먼트들을 각각 분할하여 추출한다.

2.2 언어처리부

구문분석기와 발음변환기로 구성되어 있다. 구문 분석기에서는 입력문장안에 있는 어절들 사이의 관계를 분석하여 자연성 향상을 위한 운율정보인 억양, 휴지부의 길이등의 정보를 생성한다. 발음변환기에서는 입력된 문장을 자음동화나 된소리 현상과 같은 한국어 발음규칙에 따라 실제적으로 발음될 문장으로 변환시킨다.

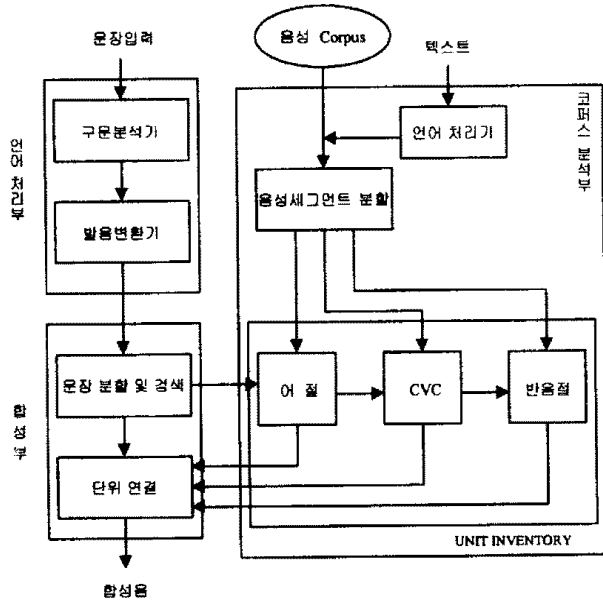


그림 1: 가변합성단위 기반 음성합성시스템

2.3 음성합성부

음성합성의 과정은 그림 2와 같다. 언어처리된 입력문장이 음성합성부로 들어오면 우선 문장을 어절별로 분할하여 각각의 어절을 어절단위 음성세그먼트들로 구성된 단위목록 전체를 검색하고 동일한 어절이 존재하면 추출하여 합성에 사용한다. 만약 원하는 어절이 어절목록에 없으면 다시 CVC단위로 재분할하여 어절보다 작은 CVC 음성세그먼트로 구성된 단위목록을 검색하여 원하는 CVC를 추출하여 이용한다. 마지막으로, 추출하지 못한 CVC를 반음절로 다시 분할하여 반음절 음성세그먼트로 구성된 단위목록을 검색한 후 추출한다. 결과적으로 단위목록을 통해 합성에 필요한 모든 음성세그먼트를 얻을 수 있다. 단위목록은 어절, CVC, 반음절순으로 검색된다. 이렇게 추출된 서로 길이가 다른 음성세그먼트들을 연결하여 합성파형을 만든다. 음성세그먼트를 연결할 때는 언어처리부에서 추출해낸 운율정보와 문맥정보를 바탕으로 휴지부를 삽입하고 합성음의

크기를 조절한다.

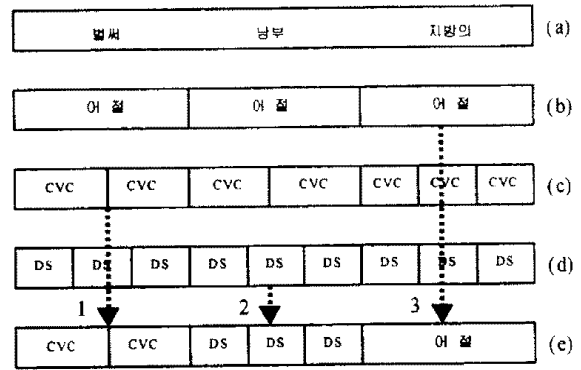


그림 2: 음성합성부

III. 단위목록(Unit Inventory)의 설계

본 논문에서 제안된 시스템에서 단위목록은 합성 음질 향상을 위해 가장 중요한 부분으로 합성에 사용되는 모든 음성 세그먼트들을 여기에서 추출한다. 이러한 단위목록은 앞에서 언급한 것과 같이 대량의 음성 데이터베이스(corpus)를 기반으로 분할된 어절, CVC, 반음절 음성세그먼트들로 구성된다. 어절목록은 전체 문서 코퍼스 검색을 통해 어절별 출현빈도를 통계적으로 조사하여 자주 사용되는 어절들을 바탕으로 만들어 진다. 음성 코퍼스가 사전 설계를 바탕으로 제작된다면 출현빈도가 높은 어절들을 어절 목록에 많이 포함하게 되어 고음질의 음성합성이 가능할 것이다. CVC목록은 코퍼스 검색을 통해 존재하는 문장세그먼트를 조사하고 이를 바탕으로 CVC 음성세그먼트들이 만들어 진다. 마지막으로 무제한 어휘합성이 가능하도록 음절의 종성의 유무를 고려하여 반음절목록을 만든다.

IV. 실험 및 결과

4.1 실험 환경

본 실험에서 합성에 필요한 대량의 음성 데이터베이스(corpus)를 구성하기 위하여 40대 남성 아나운서의 음성을 사용하였다. 전체 크기는 약 100Mbytes 이고 코퍼스에 관한 사전 설계없이 정기뉴스의 내용을 DAT를 통해 녹음했다. 16bit의 정확도로 양자화했으며 샘플링 주파수는 11.025 KHz이다.

4.2 실험 방법

합성문장을 결정하고 문장을 어절별, CVC별, 반음절별로 각각 분할하여 전체 문서 코퍼스를 검색하였

다. 이를 바탕으로 음성 코퍼스에서 음성세그먼트를 분할 선택하여 각 단위별 단위목록을 사전에 구성하였다. 실험문장은 본 논문에서 제안된 가변합성단위 시스템에 적합한 문장을 선택했다. 합성문장은 다음과 같다.

“벌써 사흘째 남부 지방의 호우가 계속되고 있습니다.”

합성음질을 평가하기 위해 가변합성단위를 사용하여 합성한 문장과 CVC 또는 반음절만으로 합성한 문장을 비교하였으며 가변합성단위를 사용한 합성문장의 어절의 개수를 인위적으로 변화시켜 합성하였다.

4.3 합성단위별 음질의 비교

합성단위의 길이는 음성세그먼트 연결시 접속점의 개수와 관계가 있다. 접속점의 개수가 적을수록 음질은 향상된다. 이것을 비교하기 위해 기존의 합성단위 CVC 또는 반음절만을 사용하여 합성한 음성파형과 본 논문에서 제안한 가변단위를 사용한 것을 비교하였다. 그림 3은 합성된 음성의 파형들이다.

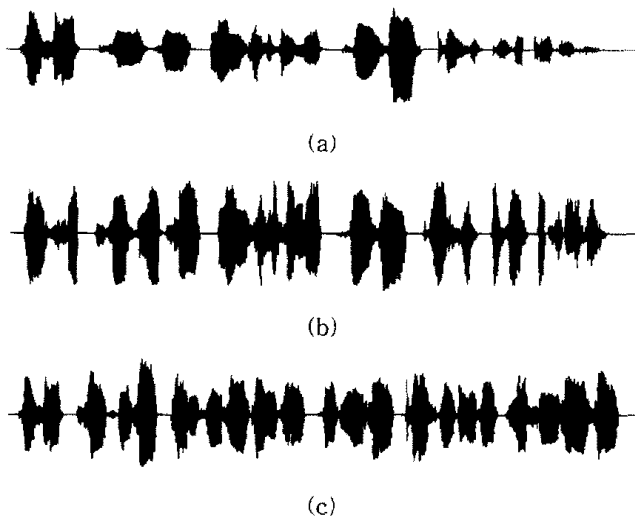


그림 3: 단위별 합성파형
(a) 가변단위 (b) CVC단위 (c) 반음절단위

4.4 어절 개수에 따른 합성음질의 비교

만약 단위목록안에 합성문장을 구성하는 어절과 동일한 어절이 많이 포함되면 조음결합에 의한 음운변동을 좀 더 많이 포함할 수 있어서 합성음질이 향상된다. 음질의 변화를 살펴보기 위해 실험문장의 어절의 개수를 인위적으로 줄여가며 합성음질을 비

교하였다. 그림 4의 (a),(b)는 각각 문장내의 목적 어절이 4개, 2개 존재할 경우의 합성과형이다.

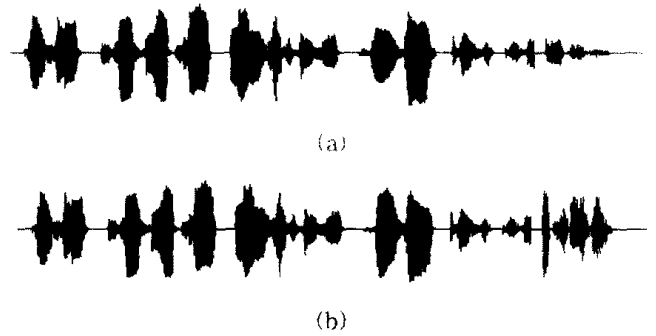


그림 4: 어절 개수에 따른 합성과형
(a) 4개 (b) 2개

4.5 실험 결과

실험문장을 본 논문에서 제안한 가변합성단위 시스템으로 합성해 본 결과 기존의 다른 시스템에 비해 합성음질이 향상되었다. 가변합성단위에 기반한 합성방법은 CVC, 반음절 그리고 길이가 긴 어절을 선택적으로 사용하기 때문에 음성세그먼트 연결시 음절저하가 발생하는 접속점의 수를 감소시켜 음질을 향상시킨다. 실험문장을 합성단위별로 각각 합성하여 얻은 접속점의 수는 표 1과 같다. 참고로 여기서 가변합성단위로 합성한 문장은 단위목록에서 어절 5개, CVC 3개, 반음절 4개의 음성세그먼트들을 사용하여 합성되었다. 어절개수에 따른 접속점의 수의 변화는 표 2와 같다.

	반음절	CVC	가변길이
접속점의 수	21	19	4

표 1: 합성단위별 접속점의 수

	5개	4개	2개
접속점의 수	4	7	14

표 2: 어절 개수에 따른 접속점의 수

V. 결론

본 논문에서는 기존의 고정된 합성단위가 아닌 가변합성단위를 사용하는 합성 시스템을 제안하였다. 합성시 어절, CVC, 반음절단위의 선택적 사용을 위해 코퍼스로부터 문맥 및 운율정보를 기반으로 사전에 합성단위별로 단위목록을 구성하였다. 본 논문의 제안된 합성 시스템은 어절단위를 사용함으로써 음

성세그먼트 연결이 일어나는 접속점 수를 감소시켜 비교적 자연스러운 합성음을 얻을 수 있었다. 또한 코퍼스를 기반으로 복합음소와 같이 길이가 긴 합성 단위를 사용하여 문장을 합성할 경우 존재하지 않는 음성 세그먼트들을 반음절단위로 대신함으로써 무제한 어휘 합성의 가능성을 제시하였다. 하지만 본 논문에서는 음성 코퍼스로 단기간에 공중파 뉴스의 내용을 사전 설계없이 녹음하여 사용하였기 때문에 실생활에서 자주 쓰이는 어절들을 합성문장에 많이 반영할 수 없었다. 따라서 가변합성단위를 사용하여 보다 자연스러운 음성을 합성하기 위해서는 단위목록 생성에 기초가 되는 음성 코퍼스 제작을 위한 사전설계와 정확한 음성세그먼트의 분할 추출이 무엇보다도 중요한 과제이다. 이것이 만족된다면 고품질의 무제한 어휘 합성 시스템 구현이 가능할 것으로 기대된다.

참고문헌

- [1] T. Dutoit, "An Introduction to Text-to-Speech Synthesis, Kluwer Academic Publishers, Dordrecht, 1997.
- [2] 김 재홍, 이 철희, "고품질 한국어 음성합성 시스템을 위한 합성단위의 선택," 한국음향학회 학술대회 발표 논문집, pp. 269-272, 1998년 11월.
- [3] 여 종락, "반음소: 새로운 음성합성 및 인식단위," 음성통신 및 신호처리 워크샵 논문집, pp. 208-212, 1993.