

한국어 음성 합성을 위한 운율 및 길이 정보의 추출

양진석, 박광철, 양세라, 김재범, 이정현
인하대학교 전자계산공학과

An Extraction of the Prosody and Duration Information for Speech Synthesis in Korean

Jinseog Yang*, Kwangchul Park, Serah Yang, Jaebeom Kim, Junghyun Lee
Dept. of Computer Science & Engineering, Inha Univ.

요 약

자연스러운 음성 합성을 위해서는 운율 및 장단음 처리가 선행되어야 한다. 본 논문에서는 음성학적 실험과 억양 규칙을 이용하여 한국어 문장으로부터 운율 제어 정보와 모음 길이 정보를 추출한 뒤 음성 합성에 적용함으로써 합성음의 자연성을 향상시키는 방법을 제안한다. 이러한 정보는 문장 분석후 일련의 운율 규칙을 적용하여 반복된 실험을 통해 수치화함으로써 얻을 수 있었다. 실험결과, 운율 및 장단처리를 적용한 본 시스템에서는 자연성이 향상된 결과를 얻을 수 있었다.

1. 서론

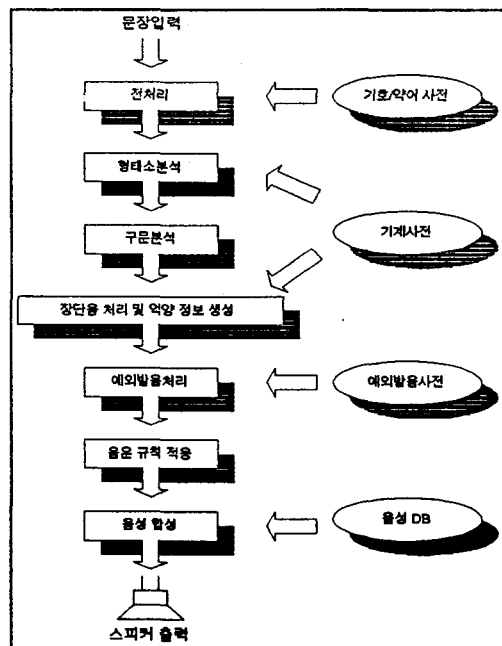
음성합성 시스템을 개발하는 데 있어서 중요하게 고려해야 할 사항은 명료성(intelligibility)과 자연성(naturalness)의 문제이다.

특히, 한국어를 위한 음성합성 시스템에 있어서 명료성은 자음이나 모음에 따라서 단어의 의미가 바뀔 수 있고, 자연성은 컴퓨터로 하여금 보다 인간다운 발성을 가능케 한다는 점에서 각각 중요한 역할을 한다^[1,2,3]. 자연성에 영향을 미치는 요소에는 여러가지가 있지만^[3], 그 중에서 가장 큰 요소로는 운율 제어와 길이 제어를 들 수 있다.

본 논문에서는 일련의 억양 규칙에 의하여 운율 및 길이 정보를 추출하였고 이를 실제 시스템에 적용하여 자연성이 향상된 음성합성 결과를 얻을 수 있었다.

2. 음성 합성기의 구성

한국어 처리를 위한 한국어 음성 합성기는 [그림 1]의 전체 구성도와 같이 크게 7단계로 구성된다.



[그림 1] 합성기의 전체 구성도

2.1 전처리

전처리는 입력된 문장에 있어서 한글이 아닌 숫자, 약어, 기호 등을 모두 한글로 바꿔주는 역할을 한다. 이러한 처리는 한글로 된 문장만 합성이 가능하도록 제한을 두었기 때문이며, 한번 처리된 기호나 약어는 기호/약어 사전에 저장되어 다음에 또 그러한 기호/약어가 나타날 때는 이 사전을 탐색한 후 자동 변환이 이루어진다. 그리고 숫자인 경우는 문장에서 숫자위의 단위에 따라 읽는 방식이 달라지므로 이를 고려해서 적절하게 변환이 이루어진다.

2.2 형태소 분석

형태소란 의미를 가진 최소의 단위를 말하며, 문장에서 형태소들을 분리해 내는 것을 형태소 분석이라 한다. 본 논문의 실험에서는 전편집 단계, 가능한 결합 추정 단계, 1차 필터링과 불규칙 처리 단계, 2차 필터링과 후편집 단계의 4 단계로 구성된 형태소 분석기를 사용하였다¹⁾.

2.3 구문 분석

형태소 분석을 하게 되면 하나의 단어에 대해서도 여러개의 후보 형태소 결과들이 나타나게 된다. 그러므로 형태소 분석 결과에서 나온 여러 후보 형태소들 가운데 문장에 맞는 하나의 형태소를 선택해 주는 구문 분석이 필요하게 된다. 하지만 본 실험에서는 결합 가능성이 가장 높은 형태소 결합을 일련의 규칙으로 만들어 여러 후보 형태소들 가운데 하나를 선택하는 방법을 사용하였다.

2.4 장단음 처리 및 억양 정보 생성

적절히 선택된 형태소들 중 명사인 경우는 이미 장단음 정보를 포함하고 있는 기계사전을 탐색하여 장음에 대해서는 명사 뒤에 장음 정보를 추가한다. 그리고 품사 정보에 따라 음운 단어, 음운구, 억양구 등으로 나누어 적절한 표시를 하게 되는데, 이들은 운율 정보 추출에 있어 매우 중요한 역할을 한다.

2.5 예외발음 처리

음성합성시 자연성을 높이기 위해서는 음절 단위의 문자표기를 직접 음성으로 합성하기 보다는 사람이 실제로 발음하는 표기로 변환해서 합성하는 것이 바람직하다. 그러나 현재 한국어의 음운 규칙으로는 모든 단어들에 대한 올바른

표기가 불가능하므로 이를 해결하기 위해 예외 발음 사전을 두고 음운 규칙을 적용하기 전에 먼저 이들 단어들을 알맞은 표기로 바꾼다.

2.6 음운 규칙의 적용

예외발음이 아닌 단어들은 한국어 표준 발음 규칙에 해당하므로 이들에 대해서는 문교부에서 정한 규칙에 의해 변환하여 표기한다. 적용되는 음운 규칙으로는 적용 순서대로 ㄴ 첨가, 구개음화, 음운축약, ㅎ 변환, 자음동화, 경음화, 말음법칙, 연음법칙이 있다.

2.7 합성기

음성합성을 하는 방법에는 조음(articulatory) 모델 방식, 포만트 합성(formant synthesis) 방식 등이 있다. 조음 모델은 구현이 매우 복잡하고 계산량도 많은 단점이 있으며^{1),2)}, 포만트 합성 모델은 자음을 합성해 내는데 있어 명료성이 떨어진다는 단점이 있다. 그래서 본 논문의 실험에서는 반음절 단위로 녹음된 음성 데이터를 결합하는 방법을 사용하였다.

3. 운율 및 길이 정보의 추출

운율은 발화 단위에 내포된 의미정보에 따라 결정되는데 구와 절의 위치, 기능, 상호결합관계 등의 복합적인 요인에 영향을 받게 된다. 그래서 본 논문에서는 이러한 사항들을 고려하여 운율 제어 정보와 길이 제어 정보를 추출한 뒤 수치화하였다.

3.1 억양 규칙

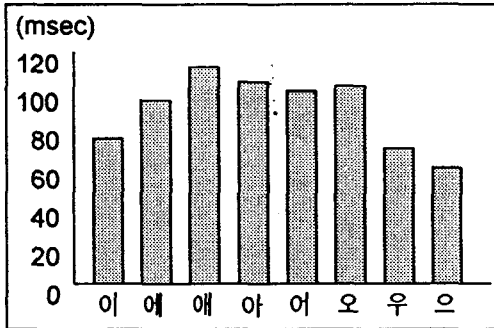
운율 정보를 추출하기 위해서 ¹⁴⁾와 ¹⁶⁾으로부터 얻어진 결과를 종합하여 다음과 같은 억양규칙을 정의하였다.

- ① 어떤 음절에 강세가 오면 그 다음 음절부터 피치가 올라간다.
- ② 문두와 문중에는 억양구의 끝에서 피치가 최고조가 된다.
- ③ 문두와 문중에는 한 억양구가 끝난 후 임시 휴지기를 갖는다.
- ④ 문미 끝의 억양구는 그 문장의 유형에 따라 피치 유형이 결정된다.
- ⑤ 강세는 억양의 시작을 나타낸다.
- ⑥ 하나의 발화 단위 안에서 각 억양구의 최고 피치는 발화 끝으로 갈수록 낮아진다.

⑦ 억양구 안에 있는 음절의 수가 많을수록 각 음절이 차지하는 시간이 짧아진다.

3.2 모음에 따른 고유 지속 시간

모음의 길이는 조음 기관의 조음 시간으로 결정이 된다. 일반적으로 조음 기관의 이동거리가 짧은 경우(고모음 /이/가 해당)는 큰 경우(저모음 /아/가 해당)보다 길이가 짧으며, 움직임이 빠른 조음 기관(혀끝이 해당)으로 발음되는 경우가 움직임이 느린 조음기관(입술이 해당)으로 발음되는 경우보다 짧다. 또한 조음 기관의 긴장이 없는 연음(lax)의 경우는 긴장이 있는 경음(tense)보다 짧는데, 이는 경음을 발음하기 위해서는 조음 기관의 근육의 긴장이 이루어지는 시간이 소요되기 때문이다. 이처럼 모든 음성학적 조건이 같을 때는 조음적인 제약에 의해 말소리의 길이가 달라질 수 있으며, 이것을 말소리의 고유 지속 시간(intrinsic duration)이라고 한다. 그러나 실제 말소리의 길이는 앞뒤의 소리, 발화의 길이, 악센트, 위치, 속도 등의 영향을 받아 나타나게 되는데, 이것의 상대적인 차이를 나타내면 [그림 2]와 같다.



[그림 2] 모음 지속 시간의 상대적 길이

3.3 운율 및 길이 정보의 생성

형태소 분석과 구문 분석 후에 부가된 음운 단어, 음운구, 억양구의 표시에 따라 위의 억양 규칙을 적용하여 음성 사건의 피치를 변경할 정보를 수치로 표시한다. 수치값은 1.0을 기준으로 하여 피치가 낮은음을 만들 때는 1.0보다 큰 값을, 그리고 피치가 높은음을 만들 때는 1.0보다 작은 값을 사용해서 피치를 조절한다. 그리고 길이 정보는 장단음 처리를 통해 얻은 명사의 장단 정보와 억양 규칙에 따른 길이 변화량을 고려하여 1에서 5까지의 정수값으로 표시하여 가장 긴 길이 정보를 1로 하였다. 다음의

[그림 3]은 입력 문장으로부터 운율 및 길이 정보를 추출하여 수치화한 예를 보여준다.

입력 문장 :
그는 웃으면서, 그녀를 향해 이렇게 말했습니다.
문장 분석 후 :
그+는/ 우스면서/ 그녀+를/ 향해 이러케/ 말+했 습니다.#
결과 :
<2,1>그<1,0.97>는<4,0.98>우<2,0.99>스<4,0.98>면 <2,0.97>서<3,0><3,0.98>그<2,0.98>녀<1,0.98>를 <4,0.99>향<2,0.98>해<3,0.97>이<3,0.96>러 <2,0.98>케<3,0.98>말<1,0.99>했<4,0.97>슴 <3,0.99>니<2,1.02>다<3,0>
주)+: 음운 단어 /: 음운구 #: 억양구 <길이 정보, 피치 변경값>

[그림 3] 적용 예

4. 운율 및 길이 정보에 따른 파형의 변환

추출해 낸 정보를 이용한 운율 제어에는 Multirate Interpolation/Decimation 필터를 사용하였으며^[3], 길이의 제어에는 수치 정보와 모음의 고유 지속 시간에 따라 모음 중간의 피치를 삽입 또는 제거하는 방법을 사용하였다. 그리고 피치를 변경하는 수치는 그 범위를 0.90과 1.10 사이로 한정하였는데, 이 범위는 실험에 의해 정해진 값으로 이 범위를 벗어나면 스펙트럼의 왜곡이 심해져서 음색이 변하게 된다. 또한 길이를 제어할 때는 정확한 피치를 찾아낼 수가 없기 때문에 삽입 또는 제거시 불연속점이 발생하게 되는데, 이를 제거하기 위해 불연속점 사이에 몇개의 샘플을 추가하여 보상하는 방법을 사용하였다.

5. 실험

실험은 운율 및 길이 정보를 사용하여 합성한 경우와 정보를 사용하지 않고 단순히 음운 규칙만을 적용하여 합성한 경우를 비교하여 평가하였다.

5.1 실험환경

한국어 처리 및 음성 합성을 위해 실험에서 사용한 시스템으로는 75MHz의 CPU를 장착한 펜티엄을 사용하였으며, D/A 컨버터로는 Sound blaster-16을 사용하였다.

5.2 음성 사전의 구축

음성 사전은 한국 전자 통신 연구소(ETRI)에서 학술용으로 공개한 음성 DB를 사용하였다. 원래 이 DB는 16-비트/16kHz로 연속 녹음되었으나 실험에 사용하기 위해 음질별로 hand-labeling을 하고 PC에서 사용이 가능하도록 화일의 저장 포맷을 변환하였다. 또한 1,244개로 구성된 데이터들 중에서 2음절로 된 데이터들을 뺀 466개의 단음절 데이터만을 선택하여 사용하였으며, KBS 아나운서에 의해 발성된 여성 음성 사전과 전문 발성가에 의해 발성된 남성 음성 사전으로 두가지를 구축하였다.

5.3 실험 결과

합성음의 평가를 위해 청취자 15명을 대상으로 청취 실험을 하였고, 운율과 길이를 제어한 합성음과 제어하지 않은 합성음을 비교 평가하였다. 평가는 10점을 기준으로 자연음을 들려준 후, 합성음을 들려주어 0에서 10사이의 점수를 주도록 하여 평균한 MOS(Mean Opinion Score) 평가를 하였다. 그 결과 6.7의 점수를 얻었으며, 운율 제어를 하지 않은 합성음의 5.8과 비교할 때 0.9 정도 자연성이 향상되었다.

6. 결론 및 향후 과제

실험 결과 자연음에 비해 아주 만족스러운 합성음을 만들어 내지는 못했지만 반응질 단위로 합성했기 때문에 평료성은 만족할 수준이었다. 다만 음절간의 연결부분이 완벽하지 않아 자연성이 다소 떨어지는 문제가 발생하였다. 또한 운율 및 길이 정보를 추출하기 위한 한국어 처리 단계가 완벽한 것은 아니었지만, 그에 비해 합성음의 자연성 향상이라는 측면에서는 많은 효과가 있는 것으로 나타났다. 그러므로 한국어 처리 단계에서 보다 정확한 분석이 이루어져 운율 및 길이 정보가 추출되고, 음절간의 연결부분에 대한 보상이 적절히 이루어진다면 합성음의 자연성이 향상될 것이라고 기대된다.

참고문헌

- [1] Dennis H. Klatt, "Software for a cascade /parallel formant synthesizer," J.Acoust. Soc.Am, Vol. 67 No. 3, pp.971-995, 1980.
- [2] Dennis H. Klatt, "Review of text-to-speech conversion for English," J.Acoust.Soc.Am, Vol. 82 No. 3, pp.737-793, 1987.
- [3] L.R.Rabiner,R.W.Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.
- [4] 구희산, "음성합성의 운율처리를 위한 악센트 연구," 음성-음운형태론연구, 한국문화사, pp.21-34, 1993.
- [5] 여상화, *다단계 필터링을 이용한 형태소 분석기의 설계 및 구현*, 인하대학교 공학석사 학위 논문, 1992.
- [6] 임홍빈, "국어 억양의 기본 성격과 특징," 새국어생활, 제3권 1호, pp.58-89, 1993.
- [7] 지민제, 최운천, 김상훈, "우리말 소리의 길이: 실험음성학적 연구," 제5회 한글 및 한국어 정보처리 학술발표 논문집, pp.119-130, 1993.