

## 음성합성을 위한 분절음 길이에측 모델링

대구대학교 영어교육과

정 현 성

음성합성을 위한 운율연구는 실험음성학과 음성공학 분야에서 활발히 이루어져 왔고, 가시적인 성과도 거두어 왔다. 최근 운율의 자연성을 향상시키기 위한 노력은 corpus-based unit selection에서와 같이, 대단위 음성자료에 의존하게 되면서 급격한 자연성의 향상을 가져온 것이 사실이지만, 여전히 새로 주어진 자료에 대해서 부자연스러운 운율을 극복해야 한다는 과제를 안고 있다. 이 논문은 길이, 억양, 에너지 등의 운율요소 가운데, 우선적으로 분절음의 길이에 초점을 두고, 대규모 음성자료를 바탕으로 한 자료기반형 길이에측모델을 지양하고 언어학적 지식과 언어자질을 이용한 지식기반형 길이에측모델을 제시하고자 한다. 모델링을 위한 방법으로는, 기본적으로 자료기반형 모델링 기법이지만 언어자질을 이용한 길이에측이 가능한 Classification and Regression Tree(CART) 모델링과 전통적인 지식기반형 모델링인 sequential rule systems을 확대시킨 Sums-of-Products 모델링을 사용한다. 실험을 위해 실제로 방송된 뉴스원고를 1인의 표준어 화자가 자연스럽게 낭독한 670개 문장을 음성처리를 통해 음성자료로 구축하였다. 이 자료에는 총 42,103개 분절음이 포함되어 있고, 그 중 80%는 학습자료로 20%는 실험자료로 활용하였다. 사용된 언어학적 자질은 목적분절음의 자질과, 주변음의 자질, 음절구조, 어구 내 위치 등이다. CART 모델을 사용한 결과, 학습자료를 통해 훈련된 모델링을 실험자료에 적용하였을 때 가장 좋은 결과는 상관계수가 0.77, 평균제곱근 오차는 25.11msec로 여타 연구결과와 비슷하거나 나은 결과를 보였다. CART 모델링을 통해 언어자질 가운데, 악센트구, 단어경계, 유성음 등이 음절과 분절음의 장음화에 큰 영향을 미치고, 비음과 격음 마찰음 등이 음절이나 분절음의 길이를 짧게 하는 것을 알 수 있었다. CART 모델링의 결과를 참고로 하여 Sums-of-Products 모델을 적용해 본 결과, 최상 모델의 상관계수가 0.68, 평균제곱근 오차는 32.13msec를 보였다. 이 결과는 같은 자료를 이용한 CART 모델이나, 같은 방법을 사용한 영어나 일본어 길이에측모델의 성능에 미치지 못하는 결과이다. 이것은 이 모델 자체의 문제점일 수도 있겠지만, 자질들 간의 상호작용을 제대로 규명하지 못한 원인이 크다고 할 수 있다. 이러한 점을 보완하여 모델을 다시 세워볼 수 있을 것이다. 이 연구의 결과는 직접 음성합성에 적용할 수 있고, 이 모델을 통해 나타난 언어자질의 영향은 corpus-based unit selection의 방법개선과 그를 위한 언어자질 설정에도 기여할 수 있을 것이다.