

그놈 목소리

조수희, 신유경, 양형원, 조영선, 김위백, 윤혜빈, 김영준, 이인구, 홍연정,
김서현, 남호성¹, 윤종성²

- 1) 고려대학교 영어영문학과
- 2) 미디어젠

요약

인간은 목소리만으로 그 음성을 발화한 화자의 특징을 포착하고 식별한다. 이러한 화자 고유의 특징은 최근의 인공지능의 발전에 중심 기술로 부각되고 있다. 음성 인식은 화자 고유의 특징을 최소화하는 작업인 반면, 화자 인식은 화자 고유의 특징을 극대화하는 작업이다. 화자 고유의 특징은 주로 벡터의 형태로 표현되는데 이러한 화자 벡터를 만들기 위해서는 음향 특징이 먼저 추출되어야 한다. 추출된 음향 특징을 입력으로 전통적 통계 방식이나 최근의 딥러닝을 이용하여 화자 벡터를 구하게 된다. 본 연구는 화자 벡터의 기반이 되는 음향 특징의 다양한 추출 방법을 소개하고 각각의 장단점을 비교한다. 나아가 화자 벡터를 구하는 방식으로 통계 방식과 딥러닝 방식을 비교한다. 마지막으로 화자 벡터의 응용으로 음성 복제 기술을 소개한다.