

# 목소리에서 연령별 특성 파라미터의 추출에 관한 연구

서지호, 손정호, 배명진  
 송실대학교  
[jihoseo@lycos.co.kr](mailto:jihoseo@lycos.co.kr)

## A Study on the Voice Parameter by an Age Group

Jiho Seo, Jungho Sohn, Myungjin Bae  
 Soongsil University

### 요 약

음성신호는 화자, 잡음, 채널, 발성 스타일에 따라서 특성이 변화한다. 사람이 바뀌면 음성 특성이 조금씩 달라지고, 남녀에 따라서도 기본 주파수가 크게 달라진다. 또한 어린이, 노인의 음성은 기본 주파수, 발음 명료도, 발성속도가 달라진다. 음성인식과 음성합성에 있어서도 이러한 특성들이 영향을 미치며, 현실 점에서 그러한 기술들이 한계에 부딪치는 원인이 되기도 한다.

본 논문에서는 인간의 목소리 정보에서 그 사람의 나이를 추정할 수 있는 파라미터를 찾아냄으로써 데이터 정보를 확장(multi-modal)하고 그 사람에 대한 보다 정확한 정보를 얻어 낼 뿐만 아니라 다른 음성기술클 들을 효과적으로 개선하고 적용할 수 있도록 한다.

### I. 서론

최근에는 인간의 의사소통인 음성을 이용하여 여러 가지 일을 수행하고 있다. 이를테면 잘못된 발음을 교정하거나, 목소리를 이용하여 사용자를 확인하는 과정, 음성을 합성하여 다양한 목소리를 재현하는 등 여러 가지 방면에서 이용되고 있다. 음성인식의 경우에는 최근 음성정보만 아니라 지문정보, 동공 정보 등을 모두 취함으로써 보다 정확한 사용자 확인과정을 수행하고 있다. 최근 음성인식기의 상용화가 진행되어 많이 사용되고 있지만 아직도 실제 환경에서 사용시에 인식률은 많이 저하된다. 음성인식률은 대체로 SNR 이 25dB 까지는 변화가 없으나 그 이하가 되면 인식률이 급격히 저하된다. 특히 잡음 환경, 전화 음성에 대한 인식률은 매우 저하된다. 사람이 바뀌면 음성 특성이 조금씩 달라지고, 남녀에 따라서도 기본 주파수가 크게 달라진다. 또한 어린이, 노인의 음성은 기본 주파수, 발음 명료도, 발성속도가 달라진다. 그래서 음성인식기는 환경변화에 강인한 인식기를 구현하고자 하는 것을 목표로 하고 있다.

본 논문에서는 인간의 목소리 정보를 이용하여 그 사람의 나이를 추정함으로써 데이터 정보를 확장(multi-modal)하고 그 사람에 대한 보다 정확한 정보를 얻어내고자 한다.

인간의 음성을 결정짓는 이러한 성도의 특징 중 나이에 따라 차이를 보이는 성질을 파악하기 위해 인간의 기본주파수를 추출하고, 성도특성을 나타내는 공명현상을 분석하며 에너지 변화 특성을 비교해본다. 2 장에서 음성발생의 원리, 성도 특성과 에너지의 음성학적 의미를 알아보고, 실험 근거를 밝힌다. 3 장에서는 실험 및 결과를 보이며, 4 장에서 결론을 내리도록 한다.

### II. 음성특성

#### II-1. 음성발성모델(성도모델)

음성신호의 구조에 대한 연구는 음성정보를 추출하거나 강조할 수가 있다. 따라서 음성신호의 생성에 대한 수학적 모델은 음성을 처리하는데 있어서 매우 중요한 영역이다. 성도(vocal tract)는 성대(vocal cord)와 입술 끝까지를 말한다. 따라서 성도는 인두(식도에서 입을 연결하는 부위)와 입 또는 구강으로 구성된다. 남성의 성도 길이는 평균 17 cm 정도이다. 성도의 단면적은 혀, 입술, 턱 그리고, 0 cm<sup>2</sup>(완전히 닫혔을 때)에서 약 20 cm<sup>2</sup>까지 변화하는 연구개의 위치에 의해 결정된다. 비도(nasal tract)는 연구개에서 시작하여 콧구멍에서 끝난다. 연구개가 낮아질 때 비도는 비음을 생성하기 위해 음향학적으로 성도에 연결된다. 음성생성의 과정을 연구함에 있어서 수학적 모델로서 물리적 시스템을 표현하는 것은 매우 중요하다. [그림 1]은 성문의 구조적 그림을 보여주고 있다. 불턱도는 혀파, 기관지와 호흡기관으로 구성된 하부-성문 시스템을 표현한다. 이 하부 성문시스템은 음성을 생성하는 에너지원이다. 음성은 간단히 공기가 혀파로부터 방출되고 결과적으로 성도에 있는 협착점에 의해 공기가 동요될 때 이 시스템으로부터 방사되는 음향학적 파형이다.

음성생성에 있어서 성도의 공명주파수를 포만트 주파수 또는 포만트라고 한다. 포만트 주파수는 성도의 모양과 면적에 따라 다르고, 소리의 형태는 성도의 모양이 변화함으로써 시간에 따라 변한다.

음성생성에 대한 모델은 음성출력을 음원이 여러 가지를 통과하여 나오는 신호로 가정하고, 음원과 성도의 각 부분을 독립적인 것으로 간주하는 선형예측모델을 제시 하였다. 음원에 대한 모델로 유성음의 음원은 준