Open Quotient 와 Fundamental Frequency 정보를 이용한 성별 분류

김혜진*, 윤영우*, 윤호섭*, 이재연*
*한국전자통신연구원
e-mail :marisan@etri.re.kr

Gender Classification Using Open Quotient and Fundamental Frequency

Hye-Jin Kim, Young-woo Yoon, Ho-sub Yoon, Jae-Yeon Lee
Intelligent Robot Team
Electronics Telecommunications Research Institue

요 약

본 논문은 Open Quotient 정보와 Fundamental Frequency 정보를 이용한 성별 분류에 관한 것이다. 기존의 대표적인 성별 분류 특징정보로 Fundamental Frequency 가 있으나, Fundamental Frequency 정보로는 이용하여 분류하는 데에 중점을 두어왔으나 이 정보는 노인 혹은 어린이에 대해서는 성별 분류 특징이 어렵다는 단점이 있었다. 한편 본 논문에서 제안하는 방법은 Open Quotient 와 Fundamental Frequency 의 연령대에 따른 차별 정보를 이용하여 학습시켜 성별분류를 보다 나은 성능으로 분류할 수 있다.

1. 서론

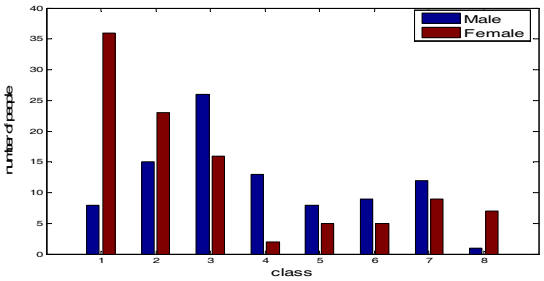
사용자 정보 수집의 요구는 정보화 시대로 접어들면서 기하급수적으로 늘어나고 있다. 예를 들어, 전략적 마케팅 분야에서는 소비자의 기호를 파악하기 위한 목적으로 개인 정보를 취합하고자 많은 노력을 기울인다. 그러나, 개인정보의 노출을 꺼리는 소비자의 입장과의 상충, 그리고 생물학적인 나이보다 젊게 살려는 경향성에 따라, 실제 나이 및 성별 보다 겉으로 드러난 나이 및 성별에 더 영향을 받는다는 연구가 있었다. 또한, 음성 정보만을 이용하여 소비자에게 서비스를 제공해야 하는 상황에서 소비자의 성향을 파악을 위한 요구 증대가 가속화되고 있다.

본 논문에서는 음성을 기반으로 사용자의 성별을 분류할 수 있는 효과적인 방법을 제안한다.

기존에 알려진 대표적인 성별 분류 특징으로는 Fundamental Frequency[1]가 있다. 그러나 이 특징은 노인이나 어린 아이들에 있어서는 특징값이 성별에 따른 구별이 잘 되지 않는다.

2. 분류 평가를 위한 음성 데이터베이스

음성 데이터는 16kHz, 16bit 로 샘플링으로 195 명을 녹음하였다. 화자당 50 문장을 녹음하였으며 한 문장은 평균 발성으로 2 초-5 초 길이이나 65 세 이상이신 분들은 길게 발음하는 경향이 있어 좀 더 길게 녹음되었다. 음성특징 추출을 위해 20 초의 음성구간을 이용하였다. 데이터베이스의 구성은 남자 92 명, 여자 103 명이며, 연령은 10 대 미만(Class I), 44 명 10 살-19 살(Class II) 38 명, 20 살-29 살(Class III) 42 명, 30 살-39 살(Class IV) 15 명, 40 살-49 살(Class V) 13 명, 50 살-59 살(Class VI) 14 명, 60 살-69 살(Class VII) 21 명 70 살 이상(Class VIII) 8 명으로 구성된다. 이 데이터의 구성은 아래 그림(Fig. 1)과 같다.

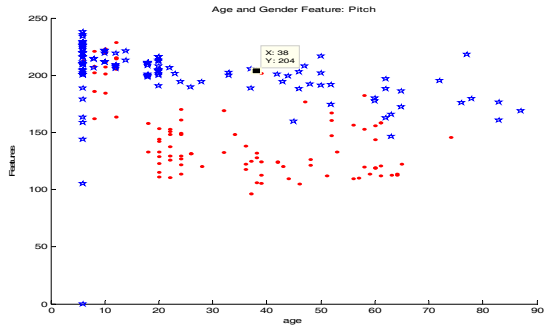


(그림 1) HRI-ETRI SpeechDB 2009

3. Open Quotient 와 Fundamental Frequency 추출

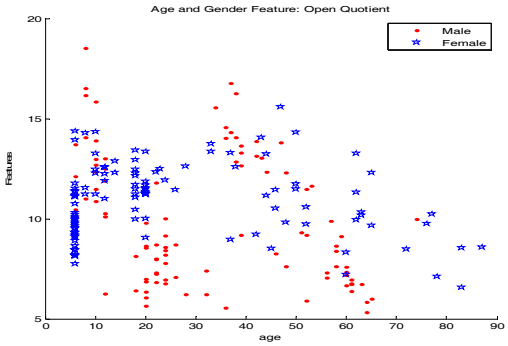
Fundamental Frequncy 는 음성신호 중에서 가장기본이 되는 주파수 즉 시간 축에서 커다랗게 나타나는 peak 들의 주파수를 의미하며 성대의 주기적인 떨림에 의해서 생성된다. 이 특징값은 허파에서 압축되어진 공기가 성대의 진동을 일으키면서 생기는 주기적인 pulse 이므로 성대의 진동 없이 난류를 일으키는 무성음의 경우에는 나타나지 않는다. 일반적으로 남성의 경우 50~250Hz, 여성의 경우 120Hz 에서 500Hz 에 존재한다. 연령대에 따라서는 남성의

경우 노인과 어린이의 특징값이 성인의 특징값보다 높은 대역에 위치하고 여성의 경우 노인과 어린이의 특징값이 성인 여성의 경우보다 낮은 영역에 존재한다. 또한 강, 억양, 감정 등에 따라서 변하는 것으로 알려져 있다. 본 논문에서는 COLEA TOOLBOX[2]을 기반으로 자기상관법(auto-correlation)을 이용하여 Fundamental Frequency 를 얻었다.



(그림 2) HRI-ETRI SpeechDB 2009 에서의 Fundamental Frequency

Open Quotient(성문개방율)은 한 주기내에서 성대가 개방되어 있는 비율을 나타낸 것이다. 보편적인 경우 성문개방율은 대략 0.4~0.7 정도이며 이 값이 커질수록 맑고 작아질수록 거친 음색을 가진다. 성문개방율이 0.8 이하로 낮아지면 폐쇄구간은 길어지고 개방구간은 짧아지게 되며 고주파 성분이 증가하게 된다. 성문개방율의 변화는 음성원음에서 지배적인 기본주파수 성분에는 영향을 주지 않지만 조화주파수(harmonics frequency)의 영향을 증가시킨다.[3] 본 논문에서는 성문개방율을 구하는 다양한 방법 중에 Five-Point Period Perturbation Quotient(PPQ5)를 구현하였다[4].



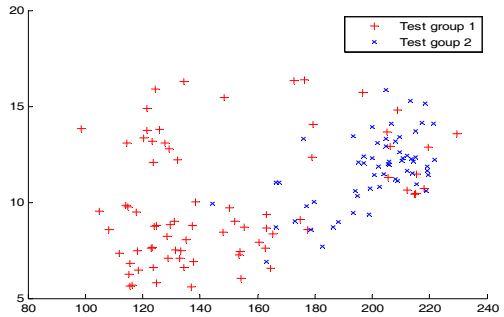
(그림 2) HRI-ETRI SpeechDB 2009 에서의 Open Quotient

4. Open Quotient 와 Fundamental Frequency 를 이용한 성별 분류 결과

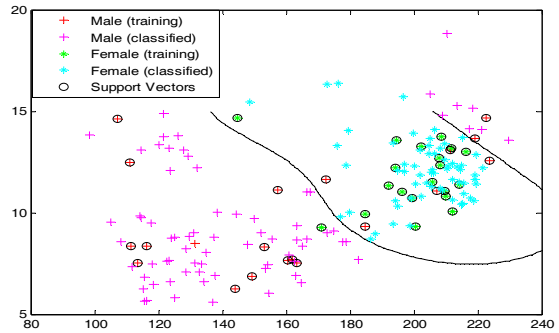
Open Quotient 와 Fundamental Frequency 를 이용한 성별 분류 결과는 k-nearest neighbor hood 방법과 svm 을 이용하여 표 1 과 그림 3, 4 에 잘 나타나 있다. 이 결과로 두 특징값을 이용하여 보다 나은 성능의 성별 분류 결과를 얻을 수 있음을 확인하였다.

<표 1> knn 과 svm 을 이용한성별분류 Error 율

	knn	svm
Funermental Frequency only	0.204	0.204
Open Quotient	0.321	0.292
Both	0.182	0.197



(그림 3) Test Data 에 대한 성별분류 결과



(그림 4)SVM 을 이용한 Test Data 에 대한 성별분류

Acknowledgement

This work was supported by the R&D program of the Korea Ministry of Knowledge and Economy (MKE) and the Korea Evaluation Institute of Industrial Technology (KEIT). [2008-F-037-01, Development of HRI Solutions and Core Chipsets for u-Robot].

참고문헌

[1] Alain de Cheveigne, Hideki Kawahara, "YIN, a fundamental frequency estimator for speech and music," J. Acous. Soc. Am. Vol. 111 No. 4 pp.1917-1930 2002

[2] COLEA: A Maltab Software Tool for Speech Analysis,http://www.utdallas.edu/~loizou/speech/colea.htm

[3] 배영민, 문영준, "성대 주위의 유동 및 음향장에 대한 수치적 모사," 3rd National Congress on Fluids Engineering August 26-28 2004 Jeju, Korea pp.459-462

[4] Crica N. et al., "Acoustic characteristics of children who speak Arabic," Elsevier International Journal of pediatric Otorhinolaryngology, vol.42 pp117-124 1997