디지털 과학수사를 위한 쌍둥이 음성 화자인식 연구

소 병 민*, 김 경 화**, 김 민 석*, 양 일 호*, 김 명 재*, 유 하 진*
*서울시립대학교 컴퓨터과학부
**대검찰청 음성분석실

A Research on Twin Speaker Recognition for Digital Forensic Investigation

Byung-Min So*, Kyung Wha Kim**, Min-Seok Kim*, IL-Ho Yang*,
Myung-Jae Kim*, Ha-Jin Yu*
*School of Computer Science, University of Seoul
**Supreme Prosecutor's Office
sbm1210@naver.com, savoix@spo.go.kr, ms@uos.ac.kr, heisco@hanmail.net,
arthmody@naver.com, hjyu@uos.ac.kr

Abstract

In this paper, we investigated the performance of speaker recognition on twins. When we hear twin's speech, it is more similar to each other than the other people's speech. In forensics, it has to be solved to investigate. Therefore, our research aims to identify twin's speech using speaker recognition system based on Gaussian mixture models. As the result, we can identify twin's speech with good accuracy even though twin's speech is very close to each other.

1. 서론

디지털 과학수사에서 협박 전화와 같이 녹취한 음성 자료를 분석하여 용의자를 판별해야 할 때 화자인식 기술이 이용된다. 이 때, 용의자가 쌍둥이일 경우 어느 정도의 정확도를 보일 수 있는지 확인하는 것이 본 연구의 목적이다. 이를 위해 쌍둥이 화자의 음성 DB를 수집하고 화자식별 실험을 수행하였다.

2. 쌍둥이 음성 DB 수집

본 연구에서 사용한 쌍둥이 음성 DB는 2008년도 대검찰청 정책연구 용역과제로 수집된 것이다. 이 DB는 일란성 쌍둥이 33쌍(남자 16쌍, 여자 17쌍), 이란성 쌍둥이 8쌍(남자 3쌍, 여자 5쌍)의 화자가 두 종류의 문단에 대해 각각 3회 반복 발성한 음성과 인터뷰로 구성되어 있다. 한 문단의 발성은 약 1분 20초 길이의 분량이고 다른 한 문단의 발성은 약 30초 길이의 분량이다. 그리고 인터뷰는 1분 30초에서 3분 사이의 자유롭게 발화한 내용으로 구성되어있다.

3. 화자식별 실험 설계 및 결과

본 연구에서는 전체 DB중 남녀 일란성 쌍둥이 30쌍(남자 15쌍, 여자 15쌍)의 16kHz, 16bit로 샘플링된 문단 발화 음성을 사용하였다. 총 3회 반복 발성한 음성 데이터 중 1회 분량은 학습 데이터(10초 분량x8~11개 발성x1회=총80~110초 분량)로, 2회 분량은 테스트 데이터(10초 분량x8~11개 발성x2회=총160~220초 분량)로 사용하였다.

화자 특징은 20차 MFCC(mel-frequency cepstral coefficient)s 특징을 사용하고 채널 보상으로 CMS(cepstral mean subtraction)를 사용하였다. 가우시안 혼합 모델[1]을 이용하여 화자 모델을 학습하였다. 국립국어원에서 배포한 '서울말 낭독체 발화 말뭉치' 20대 남녀 40명의 임의 발성 100개씩을 이용하여 혼합수 1024개의 UBM(universal background model)[2]을 학습하고 각 화자별로 MAP(maximum a posteriori)[3] 적응하였다.

다음과 같이 2종류의 실험을 진행하였다.

3.1 실험 1

쌍둥이 중 한 명의 목소리를 식별했을 때 여러 사람 중 자신의 형제 또는 자매로 식별할 확률을 알아보기 위한 실험을 수행하였다. 두 명의 쌍둥이 화자 중 한 명을 A, 다른 한명을 B라 할 때, 30쌍의 쌍둥이 중 A

들만의 음성으로 화자 모델을 학습하고 B들만의 음성으로 테스트를 하였다. 또한 그 반대의 경우도 실험하였다. 그리고 A들만으로 학습하고 A들로 테스트를 하였으며 그 반대의 경우도 실험하였다. 이 실험에 대한 결과는 표 1과 같다. 실험 결과, 쌍둥이 중 한 사람의 음성은 평균 73.0%의 높은 확률로 형제의 음성으로 식별 되었다. 이는 다른 사람들의 음성에 비해 쌍둥이 간의 음성이 상대적으로 더 유사하기 때문으로 판단된다.

표 1. 쌍둥이 화자인식 실험1에 대한 결과 (단위:%)

학습	테스트	식별률
A	B	70.9
B	A	75.1
평균 식별률		73.0
A	A	97.6
B	B	97.5
평균 식별률		97.5

3.2 실험 2

다음으로 쌍둥이 형제, 자매 중에서 특정인의 목소리를 어느 정도 식별할 수 있는지 알아보기 위한 실험을 수행하였다. 이를 위해 쌍둥이 1쌍의 음성을 화자별로 모델 학습하고 식별 실험을 하였다. 각 쌍둥이에 대해 실험 후 평균 식별률을 계산하였다. 이 실험에 대한 결과는 표 2와 같다. 실험 결과, 쌍둥이 중 한 명의 특정인을 식별할 확률은 평균 98.7%였다. 이를 통해, 화자 식별 기술을 이용하여 쌍둥이 가운데 한 명의 특정인을 식별해낼 수 있음을 확인하였다.

표 2. 쌍둥이 화자인식 실험2에 대한 결과 (단위:%)

화자	식별률	화자	식별률	화자	식별률
f2	100	f17	100	m7	97.8
f3	97.8	f18	100	m8	96.1
f4	100	f20	100	m9	98.2
f5	100	f21	97.9	m12	97.8
f6	98	f22	100	m14	100
f7	100	m1	98.2	m15	100
f8	95.5	m2	97.8	m16	95.8
f9	98.1	m4	98	m17	100
f12	100	m5	95.9	m18	100
f13	100	m6	100	m19	100
평균 식별률			98.7		

4. 결론

본 연구에서는 음성을 이용한 디지털 과학수사에서 쌍둥이 화자에 대한 화자 인식 성능을 파악하고자 하였다. 두 종류의 실험을 통하여, 가우시안 혼합 모델을 이용한 화자 식별 시스템이 쌍둥이 화자의 음성을 어떻게 식별하는지 확인하였다. 실험 결과, 쌍둥이 화자의 음성은 다른 사람에 비해 본인의 형제, 자매의 음성과 유사하긴 하지만, 쌍둥이 가운데 누구의 음성인지 식별해낼 수 있었다.

Acknowledgement

본 논문은 2010년도 대검찰청 연구용역의 지원으로 수행된 연구입니다. (과제명: 휴대폰 및 인터넷 전화음성의 화자식별률 제고 방안 연구)

참고문헌

[1] Douglas A. Reynolds and Richard C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech Audio Processing*, Vol.3, No.1, pp.72-83, 1995.

[2] Douglas A. Reynolds, Thomas F. Quatieri and Robert B. Dunn. "Speaker Verification Using Adapted Gaussian Mixture Models," Digital Signal Processing., Vol.10, pp.19-41, Jan. 2000.

[3] J.-L. Gauvain and C.-H. Lee. "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *IEEE Trans. Speech Audio Proc.*, Vol.2, pp.291-298, Apr. 1994.