

하이브리드 시스템을 이용한 화자인식에 관한 연구

⁰강현규*, 남호정*, 김형수*, 양성일*, 권영현**

*한양대학교 제어계측공학과, **한양대학교 물리학과

A Study on the Speaker Recognition using the Hybrid System

⁰Hyun Kyu Kang*, Hojung Nam*, Hyung Soo Kim*,
Sung-il Yang*, Young Hun Kwon**

*Dept. of Control & Inst. Eng., Hanyang Univ.,

**Dept of Physics, Hanyang Univ.

E-mail : khkang@newton.hanyang.ac.kr

요약

본 논문은 MLP, HMM, DTW가 결합된 하이브리드 시스템을 기반으로 한 화자인식시스템에 대한 연구이다. 이미 시스템에 등록되어 있는 화자의 인식과 비등록 화자의 등록된 화자 키워드 도용에 대한 시스템 거부. 그리고 등록된 화자의 다른 등록 화자에 대한 키워드 도용에 따른 시스템 거부에 관하여 검토하였다.

키워드는 화자의 이름을 사용하였고, 시스템에 사용된 특징 파라미터는 15차 캡스트럼, 10차 LSF, 10개 대역으로 나뉘어진 주파수 대역별 에너지값(10차 FB)을 사용하였다. 이 세 가지의 특징 파라미터들과 세 개의 인식기를 조합, 모두 아홉 개의 인식 결과값을 누적하여 인식 여부를 결정하도록 하였다. 개별적 시스템에서 발생하는 오인식을 Hybrid model을 이용하여 처리한 본 시스템에서는 실험대상 화자에 대하여 100%의 시스템 신뢰도를 얻었다.

1. 서론

정보화 사회의 기반이 되는 통신망을 통한 대규모 데이터 베이스의 접근과 정보 처리요구가 날로 증가하고 있다. 정보의 보안 문제가 심각해지고, 특정한 자료에

대한 접근을 막는 보안시스템의 필요성이 부각되고 있다. 따라서, 사용자의 신분에 대한 확인이 필수적이고, 과거부터 사용되어 왔던 카드, 도장, 신분증, 비밀번호 등은 위조와 도난 등 타인이 도용할 수 있는 문제에 대한 대책이 거의 없는 상태이다. 특히, 원격시스템으로부터 통신망을 통해 접근을 할 경우에 더욱 사용자의 확인이 어려워지는 단점이 있다. 그러나 화자인식을 이용한 화자확인용 화자의 음성이 가지고 있는 정보들에 기반을 두고 사용자를 확인하는 기술로 타인의 정보를 노리는 비등록화자(imposter)에 대한 처리와 보안 문제 등에 효율적이며, 음성을 사용하여 확인을 하므로 별도의 조작없이 시스템에 접근할 수 있어서 응용 분야가 대단히 넓고, 사용이 편리하다는 장점이 있다.

본 논문에서는 현재까지 사용되고 있는 화자확인 기술들중, HMM(Hidden Markov Model)과 MLP(Multi Layer Perceptron), 그리고 DTW(Dynamic Time Warping)를 개별 시스템으로 사용하였을 때 생길 수 있는 오인식 문제를 보완하기 위하여 세가지 시스템을 동시에 사용하는 하이브리드 시스템을 이용한다. 그 결과, 실험 대상 화자에 대해 100%의 시스템 신뢰도를 얻을 수가 있었다.

2. 화자인식 시스템

본 논문에서 사용한 화자인식시스템은 하이브리드 모델을 이용하며, 그 개략적인 흐름은 그림2.1과 같다.

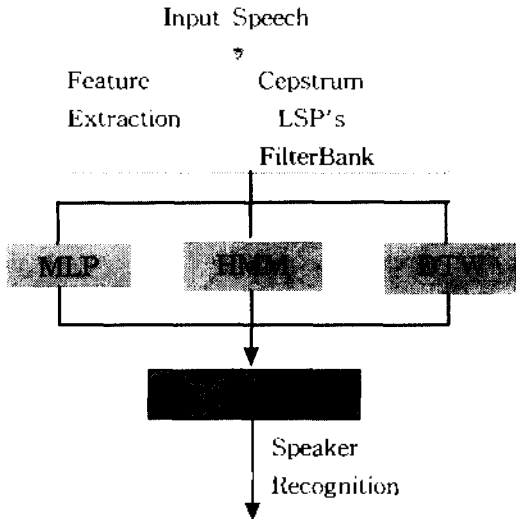


그림 2.1 Speaker Recognition Block Diagram

각각의 시스템의 입력으로 쓰이는 특징 파라미터는 15차 캡스트럼(CEP), 10차 LSF, 10차 FB를 이용한다.

이 세가지 특징 파라미터는 모두 개별 시스템의 입력으로 사용되며 각 시스템은 모두 세가지의 결과를 내게 된다. 이 결과들을 종합하여 얻어진 인식결과를 최종 인식값으로 결정한다.

3. 인식 실험 및 결과

3.1 데이터 구성

본 논문에서 사용한 음성데이터는 일반 실험실에서 20대의 성인 남성화자 40명이 발성한 음성으로 구성하였다. 이들 중 20명을 등록화자, 나머지 20명을 비등록화자의 음성으로 결정하였다.

▶ 녹음 환경 : 일반 실험실 환경

▶ A/D 변환 : 11kHz sampling, 8bit linear PCM

▶ 음성 DB 내용 : 3음절에 해당하는 이름(키워드)

▶ 전체 발음 횟수 : 화자당 본인 이름 25회,

타인 이름 5회(20명)

실험은 등록화자가 자신의 키워드를 사용하여 시스템에 접근한 경우(실험1), 비등록화자가 자신의 키워드를 사용하여 시스템에 접근한 경우(실험2), 비등록화자가 다른 등록화자의 키워드를 사용하거나 등록화자가 다른

등록화자의 키워드를 사용하여 시스템에 접근한 경우(실험3)에 대해서 각각 시행하였다.

3.2 실험1 : 허용(Access)

실험1은 20명의 등록된 화자가 자신의 키워드를 발음한 경우이다. 개별 시스템과 하이브리드 시스템을 사용한 인식결과는 다음과 같다.

●MLP를 이용한 경우

MLP	# of error	Error rate (%)
CEP	13/400	3.25
LSF	16/400	4.0
FB	1/400	0.25
Total	20/1200	1.67

●HMM을 이용한 경우

HMM	# of error	Error rate (%)
CEP	7/400	1.75
LSF	8/400	2.0
FB	6/400	1.5
Total	21/1200	1.75

●DTW를 이용한 경우

DTW	# of error	Error rate (%)
CEP	1/400	0.25
LSF	1/400	0.25
FB	1/400	0.25
Total	3/1200	0.25

●하이브리드 모델을 이용한 경우

Hybrid System	# of error	Error rate(%)
	0/400	0.0

3.3 실험2 : 거부(Deny 1)

실험2는 20명의 비등록화자가 등록되지 않은 키워드를 발음한 경우이다. 개별 시스템과 하이브리드 시스템을 사용한 인식결과는 다음과 같다.

●MLP를 이용한 경우

MLP	# of error	Error rate (%)
CEP	0/500	0.0
LSF	0/500	0.0
FB	0/500	0.0
Total	0/1500	0.0

●HMM을 이용한 경우

HMM	# of error	Error rate (%)
CEP	0/500	0.0
LSF	0/500	0.0
FB	0/500	0.0
Total	0/1500	0.0

●DTW를 이용한 경우

DTW	# of error	Error rate (%)
CEP	0/500	0.0
LSF	0/500	0.0
FB	0/500	0.0
Total	0/1500	0.0

●하이브리드 모델을 이용한 경우

Hibrid System	# of error	Error rate(%)
	0/500	0.0

3.4 실험3 : 거부(Deny2)

실험3은 20명의 비등록화자가 시스템에 등록되어 있는 등록화자의 키워드를 도용하여 발음한 경우와 등록화자가 다른 등록화자의 키워드를 도용한 경우이다. 개별 시스템과 하이브리드 시스템을 사용한 인식결과는 다음과 같다.

●MLP를 이용한 경우

MLP	# of error	Error rate (%)
CEP	2/100	2.0
LSF	1/100	1.0
FB	1/100	1.0
Total	4/300	1.33

●HMM을 이용한 경우

HMM	# of error	Error rate (%)
CEP	3/100	3.0
LSF	3/100	3.0
FB	4/100	4.0
Total	10/300	3.33

●DTW를 이용한 경우

DTW	# of error	Error rate (%)
CEP	7/100	7.0
LSF	5/100	5.0
FB	7/100	7.0
Total	19/300	6.33

●하이브리드 모델을 이용한 경우

Hibrid System	# of error	Error rate(%)
	0/100	0.0

4. 결론

본 논문에서는 기존의 화자인식 기술을 혼합하여 화자를 식별해 내는 화자인식을 위한 하이브리드 시스템을 제안하였다. 이 시스템은 개별적 시스템을 사용하였을 때 발생하는 오인식을 효과적으로 개선시킬수 있어서, 개별적 시스템을 사용하여 구현한 것보다 신뢰도 면에서 우수함을 알 수 있었다.

5. 참고 문헌

- [1] J.M.Naik and Lubensky, "A Hybrid Model HMM-MLP Speaker Verification Algorithm for Telephone speech," in *Proc. ICASSP'94*, Vol 1 pp.153-156, 1994
- [2] T.Kohonen, "The Self-Organization Map," *Proc. of IEEE* Vol. 78, No.9, Sep., 1990.
- [3] B. A. Dautrich, L. R. Rabiner and T. B. Martin, "On the Effect of Varing Filter Bank Parameters of Isolated Word Recognition," *IEEE Vol. ASSP-31*, No.4 Aug., 1989.
- [4] H. Nam, Y. Kwon, I. Paek, K. S. Lee, S. Yang, "A Study on Speech Recognition of Korean

- Digits Using TDNN," *Proc. Conference on Acoustics, Korean Institute of Acoustics*, Vol. 15, pp 87-90, Nov. 1996.
- [5] D. K. Burton, J. E. Shore and J. T. Buk, " Isolated-Word Speech Recognition Using Multisection Vector Quantization Codebooks," *IEEE Trans. On ASSP*, Vol. 33, No.4, Aug., 1985.
- [6] Soong, F. K., and B. H. Juang,"Line Spectrun Pair and Speech Compression," *Proc. of the IEEE international Conference on Acoustics, Speech, and Signal Processing, San Diego, Calif.*, Vol. 1, pp 1.10.1-4, 1984.
- [7] S. Furiui, " Speaker-Independent Isolated Word Recognition Using Dynimic features of speech Spectrum," *IEEE ASSP-34, Vol.34, No. 1*,pp52-59, 1986.
- [8] E. Rosenberg, et al."Connected Word Talker Verification Using Whole Word Hidden Markov Medels," in *Proc. ICASSP' 91*,pp. 381-384. 1991.
- [9] L. R. Rabiner, R. W. Schafer, *Digital Processing of speech Signals*, Prentice Hall, 1978.
- [10] L. R. Rabiner, B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.