제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

운율 정보를 이용한 문장 독립형 화자인식

경연정, 이황수
한국과학기술원 정보 및 통신공학과

Text Independent Speaker Recognition System Using Prosody

*Y.J.Kyung, **H.S.Lee
Dept. of Information & Communication Engineering, KAIST
*kdotori@spectra.kaist.ac.kr, **hslee@sktelecom.re.kr

요 약

본 논문에서는 문장 독립형 화자인식 시스템에 운율정보 사용을 제안한다. 스펙트럴 특징패턴만을 주로 사용하고 있는 기존의 화자인식 시스템은 채널왜곡이나 기타 잡음환경에서 성능이 크게 저하된다. 그러나 화자의 speaking style을 반영하는 운율정보는 주위환경에 강인한 특성을 갖는다. 적합한 코드북 크기와 피치 컨투어 특징 벡터의 길이를 실험 치로 구하여 자동차 소음과 백색 가우시안 소음이 섞인 음성에 대하여 화자인식 실험을 하였다. 실험 결과 소음 환경에서 운율 정보를 이용한 화자 인식 시스템이 스펙트럴 모델보다 인식율이 높음을 보였다.

1. 서론

인간의 가장 자연스러운 의사소통 수단인 음성을 이용하는 화자인식 기술은 사용의 편리함과 보안면에서 매우 유용한 기술로 인정되고 있다. 화자인식은 일반적인 보안 수단인 열쇠, 서명과 같이 도난이나 위조의 염려가 없고 지문인식과 같은 보안방식의 경우처럼 고비용의 장비를 필요로 하지 않는 장점이 있다.

화자인식은 그 분류에서 1) 발성하는 문장의 고정여부에 따라 문장 고정형과 문장 독립형으로 나눌 수 있고 2) 화자의 아이디 제공 여부에 따라 화자확인과 화자식별로 나누어 진다. 즉, 사용자가 주장하는 화자가 맞는 지를 판별하는 것이 화자 확인이고 등록된 사용자 중 한 명으로 판별하는 것이 화자식별 이다.

이러한 화자 인식은 다음 두 가지 사실에 기초하여 화자를 판별하게 된다. 첫째, 모든 사람은 서로 다른 vocal cord와 vocal tract shape을 갖는다. 둘째, 사람들 마다 자신의 특징적인 speaking style을 갖고 있다. 첫번째 특징은 스펙트럴 정보에 나타나게 되며 대부분의 화자인식 시스템은 이러한 특징패턴을 이용한다. 두 번째 정보는 운율 정보에 나타나며 피치 컨투어나 길이, 에너지 정보 등이 이에 속한다. 스펙트럴 정보를 이용하는 기존의 화자인식 시스템은 대체로 좋은 성능을 갖고 있는 것으로 보고되고 있으며[1] 실제로 Sprint's Voice Phone Card와 같이 상용화되어 사용되고 있는 것도 있다[2].

그러나 비록 선행 연구 결과들이 좋은 결과를 갖고 있더라도 아직 풀어야 할 화자인식의

문제점으로 몇 가지가 지적되고 있다[3]. 이는 대부분 '변화'에 대한 것으로 화자인식의 최종 목표는 시간, 주위환경 잡음, 발속도, 말의 크기, 사용환경 변화 등에 강인한 시스템의 개발에 있다고 볼 수 있다.

본 논문에서는 이 중 채널이나 사용환경 변화에 따른 화자인식율 저하 방지를 위해 화자의 speaking style을 반영할 수 있는 운율 정보 사용을 제안한다. 2절에서 스펙트럴 정보를 사용하는 기존 시스템을 보이고 3절에서 피치 관련 운율을 특징패턴으로 하는 화자인식 시스템을 보이도록 한다.

2. 스펙트럴 특징을 사용한 화자인식 시스템

본 논문에서는 벡터 양자화(VQ : Vector Quantization) 방법에 의해 화자인식을 하였다. VQ 방법은 비교적 간단하고 인식율도 좋은 것으로 알려져 있다[]. 그림 1에 VQ 모델을 나타내고 있다.

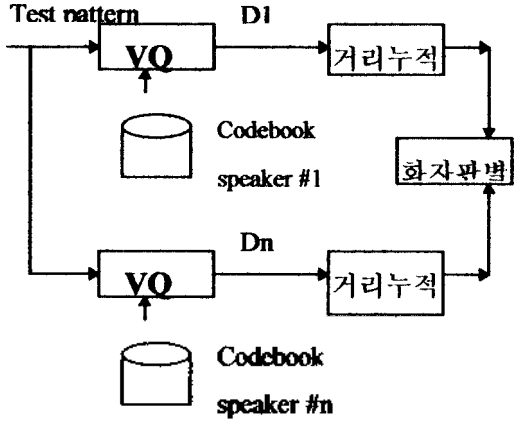


그림 1. VQ를 이용한 화자인식 시스템

VQ 모델은 사용하는 코드북의 크기에 따라 성능의 차이가 있으므로 가장 좋은 코드북 크기를 선정하기 위해 코드북 크기를 변화시켜 실험을 하였다. 실험 결과는 그림 2에 보이고 있다.

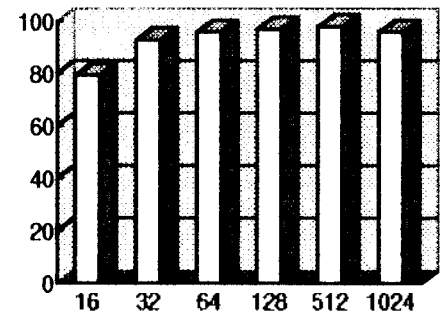


그림 2. Codebook size에 따른 인식율

실험 결과에 따라 코드북 크기는 128로 하였으며 그림 3에 소음환경 하에서의 실험 결과를 보이고 있다. 소음은 자동차 소음과 가우시안 백색잡음 두 가지이며, 그래프에서 C6020dB는 60Km로 주행하는 자동차 소음, 20dB 음성 데이터라는 뜻이고 WGN10dB는 가우시안 백색 잡음, 10dB라는 뜻이다. 각 소음 환경에서의 에러율을 나타내고 있다.

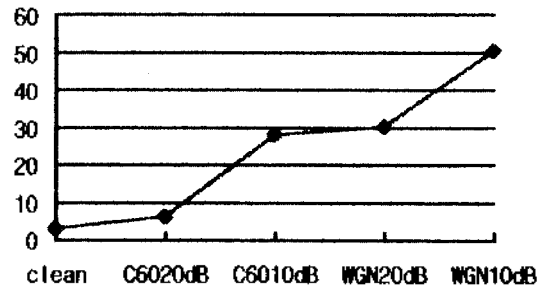


그림 3. 소음환경에서의 화자인식 결과 - 스펙트럴 정보를 이용하여

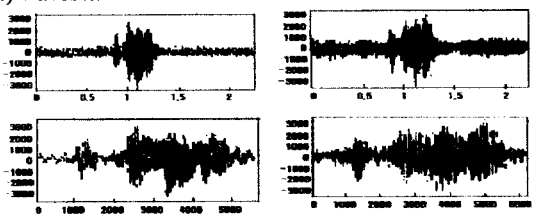
그림 3에서 알 수 있듯이 일반 VQ 모델은 소음이 심해질수록 인식율이 저하되며 가우시안

노이즈에서 성능이 더욱 떨어짐을 알 수 있다.

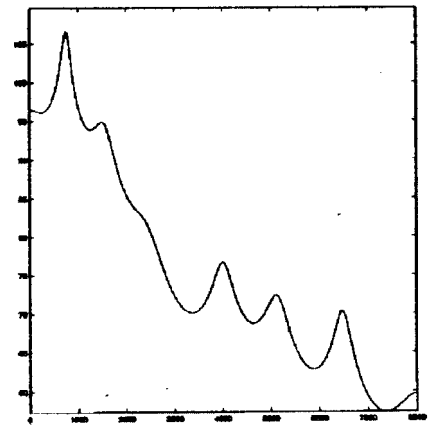
3. 운율 정보를 이용한 화자인식

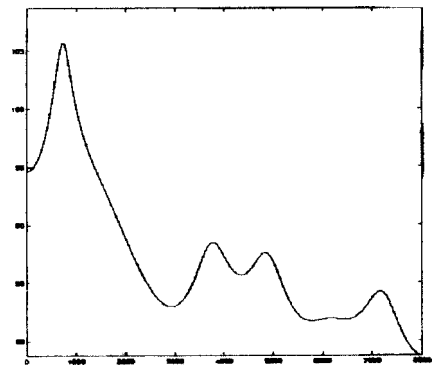
운율 정보를 이용한 화자인식에 대한 연구는 몇 몇가지 보고 되어 있다[4]-[6]. 그러나 대부분 문장 고정형이거나 운율 패턴의 통계적 성질을 이용하는 것이다. 본 논문에서는 문장 독립형의 화자인식 시스템에 피치 컨투어를 사용하여 화자인식을 행하고자 한다. 그림 4에 clean 음성 데이터와 noisy 데이터를 보이고 있다. 먼저 a)에서는 왼쪽에 clean 환경의 음성 데이터와 오른쪽에 소음 환경에서의 음성 데이터를 보이고 있다. b)는 a)의 동일 위치를 분석하여 스펙트럴 인벨로프를 보인 것이다. 1 폴반트 부분과 4 폴반트 부분이 상당히 왜곡되었음을 알 수 있다. c)는 두 파형의 전체 피치컨투어를 보이고 있다. 전반적인 컨투어가 동일하게 나타난다. 그림에서 알 수 있듯이 스펙트럴 정보는 소음에 많은 영향을 받아 왜곡되지만 피치 값은 소음 환경에 큰 영향을 받지 않는다.

a) waveform

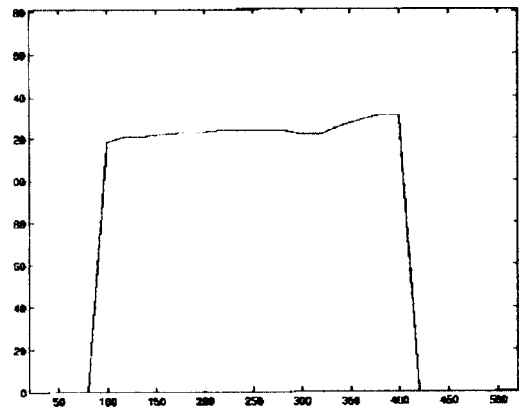


b) spectrum





c) pitch contour



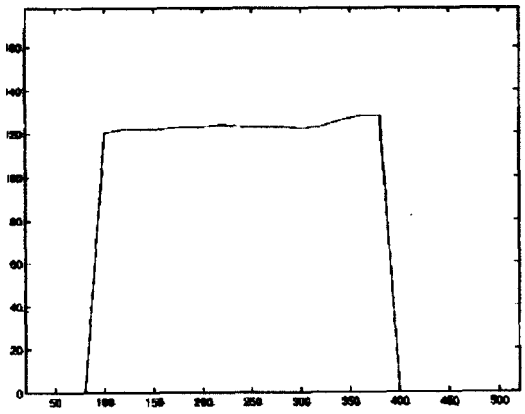


그림 4. Clean speech 와 Noisy speech

사용한 운율정보는 피치 컨투어로 적합한 컨투어 길이를 실험적으로 구하였다. 표 1에 피치 벡터 길이에 따른 에러율을 보이고 있다.

표 1 피치 컨투어 길이에 따른 화자인식 결과

	9	12	16	18	20
32	32.81	31.64	25.78	31.25	32.03
64	22.27	38.28	21.87	24.22	23.83

실험결과에 따라 피치 컨투어의 길이는 16프레임으로 하였다. 이는 약 한 음소 정도의 길이이다.

실험결과 결정된 피치컨투어 길이에 적합한 VQ 코드북 크기를 결정하기 위한 실험을 행하였다. 실험결과는 그림 5에 보이고 있다.

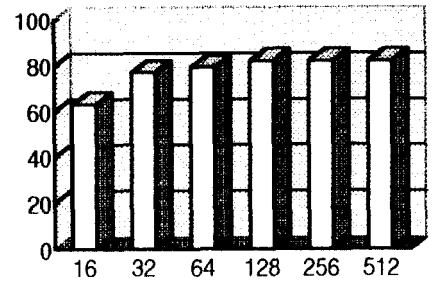


그림 5. 코드북 크기에 따른 화자인식 결과

실험결과 피치 컨투어 VQ 모델의 코드북 크기는 128로 정하였다. 정해진 값으로 인식 시스템을 만들어 스펙트럴 모델과 동일한 환경에서 인식 실험을 하였다. 그림 6에 최종 인식결과를 보이고 있다. 그림 3과의 비교를 위해 각 음성 DB 별로 에러율을 나타내 보이고 있다. 여기서도 C6020dB 는 자동차 60km 주행 시의 소음환경에서 20dB 의 음성신호라는 뜻이다.

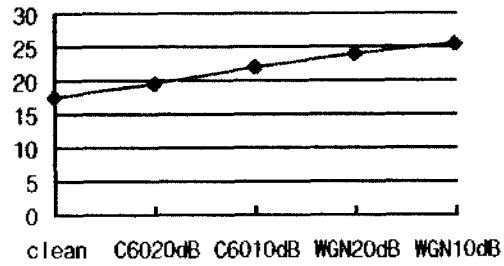


그림 6 소음환경에서의 화자인식 결과 - 운율정보를 이용하여

그림 6과 그림 3을 비교하면 clean 음성에 대해서는 스펙트럴 정보를 이용한 화자인식 시스템이 더욱 우수한 성능을 보임을 알 수 있으나 소음환경에서는 운율정보를 이용하는 것이 더욱 좋은 성능을 갖는 다는 것을 알 수 있다.

4. 결론

본 논문에서는 다양한 환경에 강인한 화자 인식을 위하여 기존의 스펙트럼 정보 뿐 아니라 운율정보를 함께 사용할 것을 제안하고 있다. 실험결과 운율 정보를 이용한 화자 인식은 스펙트럼 특징 벡터에 비해 소음 환경과 같은 주변 환경 변화에 강인한 것으로 나타났다. Clean 상태에서의 인식율 향상을 위해 두 특징 벡터를 함께 사용하는 혼합 모델을 생각할 수 있으며 좋은 인식율을 갖는 것으로 보고되었다[7].

ACKNOWLEDGEMENTS

본 연구는 과학재단의 수탁과제 연구지원에 의해 수행되었습니다. (과제번호 : 95-0100-22-01-3)

[참고문헌]

[1] Kin Yu et.al., "Speaker Recognition Models," Proc. of EUROSPEECH'95, 1995.

[2] H.Gish and M.Schmidt, "Text-independent speaker identification," IEEE Signal Processing Magazine, Vol.ASSP-35, No.2, pp.18-32, 1994.

[3] Rosenberg and Soong, Advances in Speech Processing, pp. 701-737, Marcel Rekker, 1991.

[4] B.S.Atal, "Automatic Speaker Recognition based on Pitch Contours," JASA, pp.1687-1697, 1972.

[5] J.Kraayeveld et al., "Speaker characterization in dutch using prosodic parameters," Proc. of EUROSPEECH'9, pp.427-430, 1991.

[6] B.Yegnanarayana et al., "A speaker verification system using prosodic features," Proc. of ICSLP 94, pp.S31-9.1 - S31-9.4, 1994.

[7] Y.J.Kyung and H.S.Lee, "Text-Independent Speaker Recognition Using Micro-Prosody," Proc. of ICSLP 98, (accepted), 1998.