피치변경을 이용한 화자인식 시스템

정종순, 배명진*
숭실대학교 정보통신공학과

The Speaker Recognition System using the Pitch Alteration

JongSoon Jung, MyungJin Bae*
Dept. of Telecommunication, Soongsil University
E-mail * : mjbae@saint.soongsil.ac.kr

요약문

화자인식에 사용하는 파라미터는 화자의 특징을 충분히 표현함과 더불어 발성 시마다 변동이 작은 것이 바람직하다. 즉, 파라미터의 화자내의 변이보다 화자간의 변이가 큰 특성을 가져야 화자간의 구분이 용이하다. 또한, 화자간 오류를 최소화하기 위해 화자간 구별이 뚜렷한 특징 파라미터뿐만 아니라 분별력이 뛰어난 인식방법도 필요하다. 최근의 실험결과들을 살펴보면 발성기관에 의한 정적인 특징뿐 아니라, 발성습관에 의한 동적인 특징을 같이 이용함으로써 보다 정확한 인식결과를 얻고 있다. 따라서 본 논문에서는 이러한 문제점을 해결하기 위해 다음과 같이 제안한다.

음성의 특징벡터로 운율정보 사용을 제안한다. 현재 화자인식 시스템에서 일반적으로 많이 사용되고 있는 특징벡터는 스펙트럼 정보를 모델링하고 있는 것으로 비잡음 환경에서 좋은 성능을 보이고 있다. 그러나 잡음 환경 변화에 크게 왜곡되며 인식율이 현저하게 저하되는 문제점이 나타난다. 그러므로 본 논문에서는 음성의 동적 변화를 측정할 수 있는 세그먼트로 분할한 피치열을 변경하여 인식의 특징패턴으로 사용한다. 이는 문장의 운율정보를 보여주는 것으로 소음환경에서 강인한 특성을 보였다.

Abstract

Parameters used in a speaker recognition system are desirable expressing speaker's characteristics filly and have in a speech. That is to say, if inter-speaker than intra-speaker variance has a big characteristic, it is useful to distinguish between speakers. Also, to make minimum error between speakers, it is required the improved recognition technology as well as the distinguishing characteristics. When we see the result of recent simulation performance, we obtain more exact performance by using dynamic characteristics and constant characteristics by a speaking habit. Therefore we suggest it to solve this problem as followings.

The prosodic information is used by a characteristic vector of speech. Characteristic vector generally using in speaker recognition system is a modeling spectrum information and is working for a high performance in non-noise circumstance. However, it is found a problem that characteristic vector is distorted in noise circumstance and it makes a reduction of recognition rate. In this paper, we change pitch line divided by segment which can estimate a dynamic characteristic and it is used as a recognition characteristic. we confirmed that the dynamic characteristic is very robust in noise circumstance with a simulation.

We make a decision of acceptance or rejection by comparing test pattern and recognition rate using the proposed algorithm has more improvement than using spectrum and prosodic information. Especially stational recognition rate can be obtained in noise circumstance through the simulation.

1. 서론

최근 화자인식에 있어 많은 종류의 연구가 성공하고 발전되어 왔으나, 좋은 답을 찾기 위한 문제는 여전히 남아 있다. 즉, 이들 문제 중 대부분은 화자가 발성할 때마다 생기는 변화 그리고 채널과 레코딩 환경의 변화로 인한 변이들이다. 그러므로 모든 시간과 환경에 안정된 특징을 추출하는 것은 상당히 중요하다. 즉 발성습관의 변화에 덜 민감 하고, 목소리 위장과 감기로 인한 목소리 변화에 강한 그 화자만이 가지는 고유한 특징을 추출하는 것은 중요하다.

본 논문에서는 환경에 강한 화자인식 시스템을 위해 두 가지 측면 - 특징 파라미터와 모델 구성 - 을 고려하였다. 우선, 발성자가 가지는 70 - 80 %의 운율 정보를 나타내기 위해 운율 특징을 사용할 것을 제안한다. 운율정보와 발성습관은 일반적인 스펙트럼 특징과 대조적으로 배경 환경 변화에 민감하지 않았다. 스펙트럼 정보는 잡음 환경에서 인식율이 급격히 저하되었으나 운율 정보는 인식율에 급격한 저하 없이 안정되게 나타났다. 따라서 본 논문은 운율정보를 이용하는 방법에 초점을 맞추었다. 그러므로 스펙트럼 특징과 운율 정보를 조합한 모델을 제안한다. 그리고 두 번째로 운율 정보를 벡터 양자화 모델에 이용한 화자인식 시스템을 제안한다. 이들 모델은 기존의 벡터 양자화 모델보다 잡음 환경에 강인한 화자인식 시스템을 보였다.

2. 인식 시스템의 구조

화자인식의 기본 동작의 요소는 그림 2.1에 보였다. 테스트할 화자의 발성음성은 먼저 특징 파라미터의 추출을 위해 분석된다. 추출된 파라미터는 기존에 저장되어 있는 학습용 파라미터 모델과 비교하는데 화자식별의 경우 모든 파라미터 모델간의 비교가 이루어지게 된다. 화자확인의 경우, 테스트 파라미터와 학습 파라미터의 비교는 단지 요구되는 신원에 대응하는 모델에만 이루어지게 된다. 비교에 의해 얻어지는 점수를 가지고서 화자식별과 화자확인에 대한 결정이 내려진다.

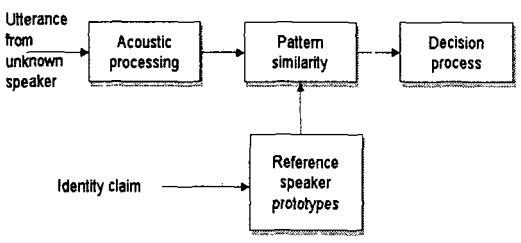

그림 2.1 일반적인 사용자 인식 시스템의 구조

3. 본 연구의 동기

최근 화자인식은 많은 발전과 연구가 이루어지고 있지만 아직 풀어야할 많은 문제를 가지고 있다. 이들 문제중 대부분은 화자내 변화, 채널 변화 그리고 녹음 환경의 변화로부터 발생된다. 배경잡음이 있을 경우 언어구사 스타일이나 운율 특징은 변화하지 않으나 스펙트럼 특징은 변화한다. 스펙트럼 특징은 비잡음 상태에서는 운율특징보다 좋은 성능을 보였으나, 잡음 환경에서는 많은 성능저하 현상이 나타났다.

본 논문에서는 운율특징이 잡음환경에 강인하다는 것을 실험을 통해 관찰했다. 이것을 그림 2.2에 보였고 그림 (a)는 "아"에 대한 음성 파형을 보여주고 있다. 위 부분의 것은 비잡음 상태에서의 음성신호이고, 아래 부분의 음성신호는 잡음 상태의 것이다. 신호에 대한 잡음(SNR)은 약 10dB이다. 그림 (b)는 비잡음 상태와 잡음 상태에서의 LPC 포락선을 보여주고 있다. 그림 (c)는 잡음가 비잡음 상태의 피치 궤적을 나타내고 있다. 이 그림에서 보듯, 운율 특징은 잡음 환경에 강인하다는 것을 알 수 있다. 스펙트럼 포락선은 잡음 환경에서 많은 부분 왜곡됨을 보였고, 피치 궤적 또한 변화가 있음을 보였다. 그러므로 본 논문에서는 운율정보를 화자인식에 적용하였다.

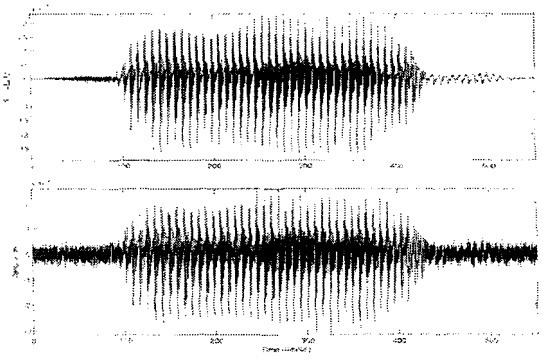

그림 2.2 (a). 음성 파형: 발성음 "아"

그림2.2(b). 스펙트럼 포락선

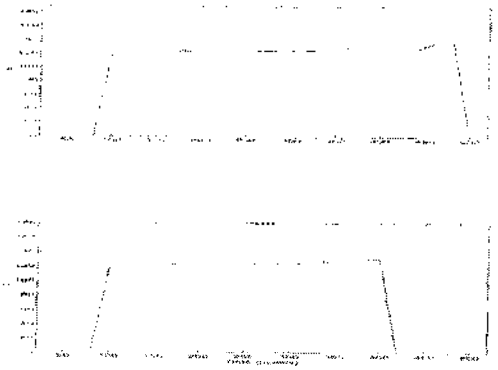

그림2.2(c). 피치 궤적

3. 제안한 피치변경을 VQ에 이용한 화자인식 시스템

3-1 운율정보를 이용한 코드북

운율 정보를 코드북에 적용하기 위해서 앞 절에서 언급한 피치알고리즘과 피치변경 방법을 선택했다. 운율정보를 코드북에 이용하기 위해서 본 논문에서는 다음과 같은 과정을 거쳤다. 우선, 발성음으로부터 피치 연속식을 추출한다. 그리고 추출된 피치를 필요에 따라 피치변경을 하고 이 피치변경 연속식과 피치 연속식을 이용하여 코드북을 설계한다. 이 코드북 블럭도는 그림 3.1에 보였다.

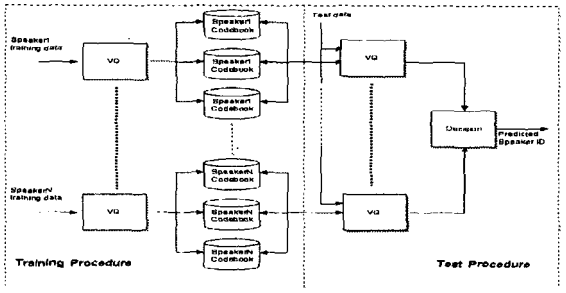

그림 3.1 운율정보를 적용한 코드북 블럭도

V/UV의 스펙트럼 상에서의 피치 변경법을 이용하여 원래의 음성 스펙트럼을 각각 70%, 80%로 압축하고 110%, 120% 신장시킨 피치 연속식으로 코드북을 구성 하였다. 한 화자당 4개의 코드북을 구성했다. 즉, 발성음으로부터 피치검출 알고리즘에 의해 검출된 피치 연속식과 이것을 피치변경 알고리즘에 의해 스펙트럼 상에서 신장, 압축된 4개의 연속식을 사용하였다. 연속식들은 P 차 벡터 단위로 세그멘트한 다음 Q 차으로 겹치면서 다음 세그멘트를 찾는다. 이것을 그림3.2에 보였다

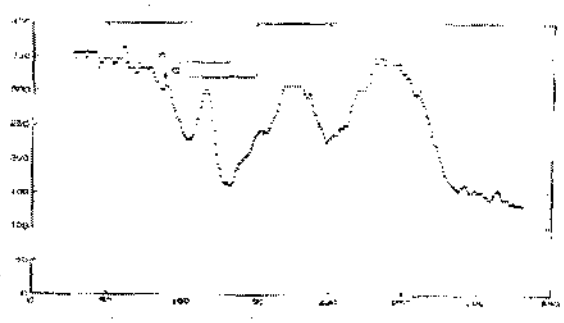

그림3.2. 운율정보를 이용한 핵심적인 개념

테스트 발성음이 입력될 때, 화자의 운율특징 VQ 코드북과 테스트 발성음에 피치 연속식과 피치변경 연속식을 비교한다. 이 운율 VQ모델은 그림3.3에 보였다. 여기서 적당한 크기의 P차를 구하는 것은 중요하다. P값이 너무 크게 되면, 전체 문장과 비교한다. P값이 너무 작으면, 운율 정보가 필요 없다는 것을 의미한다. 이 거리측도는 다음 식으로 정의된다.

$$d(t_i, c_g) = \sum_{j=1}^{P} (t_{ij} - c_{gj})^2$$

$$D_{avg} = \frac{1}{N} \sum_{i=1}^{N} (t_i - c_g) \qquad (3.1)$$

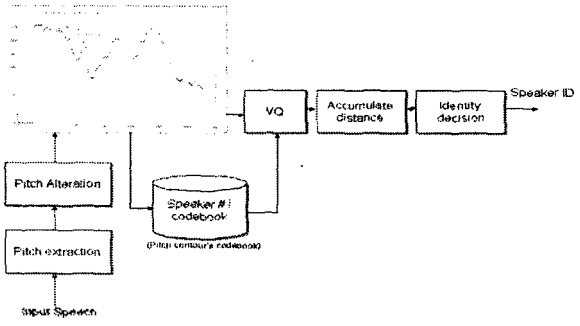

그림 3.3 운율정보를 이용한 VQ 모델의 블럭도

4. 실험 및 결과

데이터베이스 환경은 다음과 같다

- 발성음 : 256개(16*16)
- 참여인원 : 16명(남자:8명, 여자:8명)
- 문장수 : 16개(표 3-3. 참조)
- 잡음환경: 4가지(WGND40, WGND30, WGND20, WGND10)
- 잡음환경 시 문장수 : 1024개(256*4)

3가지 실험결과로 구성되어 있다. 첫 번째, P의 최

적의 차수를 찾기 위한 것이다. 만약 P = 9 라면, 한 음소길이의 대한 운율정보를 고려해야 한다는 것을 의미한다. 만약 P = N 라면, 전체문장의 운율정보를 고려해야 한다는 것을 의미한다.

표 4.1. 세그멘트 피치궤적의 차수 의존도에 대한 에러율

P-디멘전 / 코드북 사이즈	9	12	16	18	20
32	32.8	31.6	25.8	31.3	32.0
64	24.3	28.3	21.9	24.2	23.8
128	20,1	24.2	16.7	19.8	20.0

P의 값은 9 - 20까지 변화한다. 20 프레임은 대략한 음절 또는 반음절이 된다. 표 4.1에 이 결과를 나타냈다. 표에서 보듯 최적의 차수는 16이며, 이것은 한 음절 길이이다. 즉 한음절에 대한 운율정보를 의미한다. 두 번째 실험은 코드북 크기에 따른 제안한 VQ 모델의 인식실험을 했는데, 이 실험에서 VQ모델의 차수는 16으로 고정시켰다. 실험 결과는 그림4.1에 보였고, 제안한 모델 역시 코드북 크기는 128로 해도 별무리가 없음을 알 수 있다.

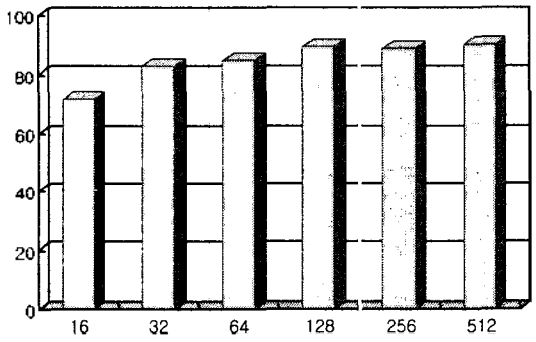

그림 4.1 코드북 크기에 따른 제안한 VQ모델의 인식실험 결과

세 번째, 화이트 가우시안 잡음 환경에서 운율 정보 이용한 VQ모델의 인식실험을 수행했다. 실험결과는 그림 4.2에 보였는데, 그림에서 보듯 가우시안 잡음에 민감하지 않음을 알 수 있다.

화이트 가우시안 잡음에서 스펙트럼 특징과 비교한 결과를 그림 4.3에 보였다. 결과에서 보듯, 스펙트럼 특징은 잡음에 상당히 민감함을 알 수 있으며 상대적으로 본 논문에서 제안한 운율 특징은 잡음 환경에서 강함을 나타냈다.

우리는 다음과 같은 사실을 발견했다. 비록 운율 특징이 비잡음 상태에서는 스펙트럼 특징처럼 좋은 성능을 보이지는 않았으나, 화이트 가우시안 잡음 환경에서는 스펙트럼 특징보다 많이 강인함을 보였다.

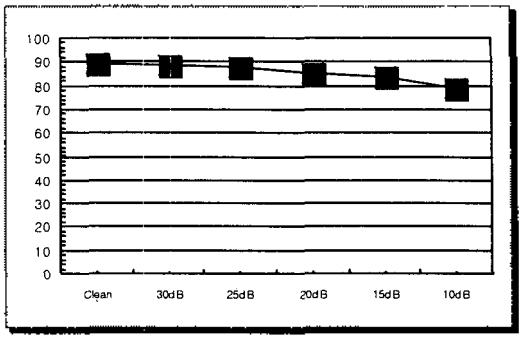

그림 4.2 화이트 가우시안 잡음에서 제안한 VQ 모델의 인식실험 결과

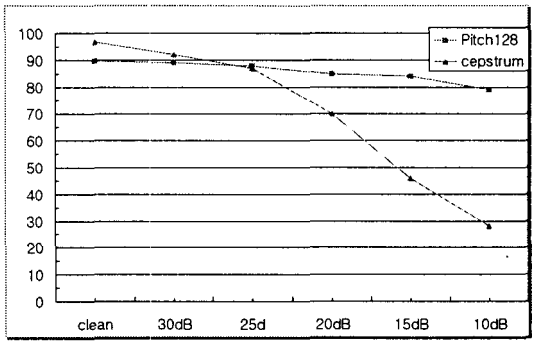

그림 4.3 화이트 가우시안 잡음 환경에서의 인식율 결과

5. 결론

본 논문에서는 음성의 특징벡터로 운율정보 사용을 제안했다 현재 화자인식 시스템에서 일반적으로 많이 사용되고 있는 특징벡터는 스펙트럼 정보를 모델링하고 있는 것으로 비잡음 환경에서 좋은 성능을 보이고 있다. 그러나 잡음 환경변화에 크게 왜곡되며 인식율이 현저하게 저하되는 문제점이 나타난다. 그러므로 본 논문에서는 음성의 동적 변화를 측정할 수 있는 세그먼트로 분할한 피치열을 변경하여 인식의 특징패턴으로 사용한다. 이는 문장의 운율정보를 보여주는 것으로 소음환경에서 강인함을 실험을 통해 보였다.

6. 참고 문헌

[1]. LR. Rabiner, Juang., "Fundamentals of speech recognition", 1993., Prentice-Hall.

[2]. LR. Rabiner, R.W. Schafer., "Digital processing of speech signals", 1978., Prentice-Hall.

[3] P.Latace, R. DeMori., "Speech recognition and understanding recent advances trends and applications", 1990., NATO.