한국음향학회지 제20권 제5호 pp. 48~53 (2001)

가중된 예측 오차 파라미터를 사용한 화자 확인 성능 개선

Speaker Verification Performance Improvement Using Weighted Residual Cepstrum

위 진 우*, 강 철 호*
(Jin Woo Wee*, Chul Ho Kang*)

*광운대학교 전자통신공학과

(접수일자: 2001년 5월 17일; 채택일자: 2001년 6월 11일)

LPC 분석 기반 화자 확인에서 잔여성분 (residue) 예측은 보통 무시되고, LPCC (LPC-cepstrum)만이 특징 파라미터로 사용된다. 본 연구에서는 잔여성분으로부터 추출된 예측 파라미터인 잔여 켑스트럼 (residual cepstrum)을 LPCC와 함께 여러 환경에서 구축된 데이터 베이스에서 화자 특징 파라미터로 사용하였다. 또한, 잔여 켑스트럼에 포함되어있는 화자 고유성분인 피치 (pitch)성분에 큰 가중치 (weighting)를 줌으로써 화자간 변이 (inter-speaker variation)가 커지도록 하는 가중치 함수를 제안한다. 실험 결과, LPCC만을 특징 파라미터로 사용하였을 경우보다 잔여 켑스트럼 (RCEP)과 LPCC를 동시에 사용했을 경우 약 6%가량의 인식 오류율이 향상되었으며, 제안한 가중치 함수를 적용한 잔여 켑스트럼 (RCEP)과 LPCC를 동시에 사용했을 경우 인식 오류율이 가중치를 주지 않은 경우보다 약 2.45%가량 개선되었다.

핵심용어: LPCC, RCEP, 가중치 함수, HMM
투고분야: 음성처리 분야 (2.5)

In speaker verification based on LPC analysis the prediction residues are ignored and LPCC (LPC cepstrum) are only used to compose feature vectors. In this study, LPCC and RCEP (residual cepstrum) extracted from residues are used as feature parameters in the various environmental speaker verification. We propose the weighting function which can enlarge inter-speaker variation by weighting pitch, speaker inherent vector, included in residual cepstrum. Simulation results show that the average speaker verification rate is improved in the rate of 6% with RCEP and LPCC at the same time and is improved in the rate of 2.45% with the proposed weighted RCEP and LPCC at the same time compared with no weighting.

Keywords: *LPCC, RCEP, Weighting function, HMM*
ASK subject classification: *Speech signal processing (2.5)*

I. 서 론

선형 예측 분석 기법은 가장 유용한 음성 분석 기법 중 하나이다. 이 기법은 음성 파라미터의 정확한 예측을 가능하게 해주며 계산량 측면에서도 매우 우수하다. LPC 분석의 부산물로써 예측 잔여, 혹은 예측 오차 $e(n)$이 발생한다. 이론상 만일 all-pole 모델이 완벽하다면, $e(n)$은 매우 작을 것이다. 그러나, 이 단순한 모델은 비음이나 마찰음에는 적합하지 않으며, 성도 전달 함수에서 극점 (pole)과 영점 (zero) 모두를 필요로 한다. 예측 오차 $e(n)$은 LPC 계수에는 포함되어 있지 않은 피치나 비음에서 기인한 영점의 정보를 가지고 있다. 음성 합성에서 $e(n)$은 all-pole 모델의 이상적인 여기 (excitation) 성분으로 여겨지지만 음성 인식이나 화자 인식에서는 보통 무시되고 단지 LPC 계수나 LPC켑스트럼 (LPCC), 반

책임저자: 위진우 (jwwee@unitel.co.kr)
139-701 서울시 노원구 월계동 447-1
광운대학교 전자통신공학과
(전화: 940-5136; 팩스: 940-5467)

사(PARCOR) 계수만이 특징 파라미터로 사용되어진다 [1-3].

현재 음성 인식이나 화자 확인을 위해 가장 널리 사용되고 있는 알고리즘으로는 음성의 추측통계 (stochastic)적인 성질을 통계학적으로 잘 모델링한 HMM기법이 사용되고 있다. HMM기법은 화자 내 변이를 잘 흡수하며 화자의 특성을 확률적으로 극대화시켜 표현해 줌으로써 화자 확인에 유용한 알고리즘이다. 그러나, HMM은 많은 양의 음성 정보를 요구하는 단점뿐 아니라 학습 환경에 의존적인 결점을 가지고 있다[4].

본 연구에서는 화자 확인을 위한 특징 파라미터로써 LPC켑스트럼 (LPCC) 및 예측 잔여로부터 추출되는 잔여 켑스트럼 (RCEP: residual cepstrum)을 사용한다. 화자 확인 학습 및 인식 모듈로는 DHMM (discrete Hidden Markov Model)을 이용하였고, 실험 결과, LPCC와 LPCC의 단점을 보완해 주는 잔여 켑스트럼을 동시에 사용하였을 경우, 잡음 환경하에서도 좋은 인식율을 나타냈다.

본 논문의 구성은 다음과 같다. II 장에서는 기존의 잔여 켑스트럼을 구하는 알고리즘에 대해 간략히 설명한 뒤, III 장에서는 RCEP에 포함되어 있는 화자 고유성분인 피치(pitch)특성을 강조하는 가중치 기법에 대하여 기술한다. 또한, IV 장에서는 실험 환경 및 기존의 방식과 제안한 방식의 성능 평가 결과를 보여주고 마지막으로 V 장에서는 결론 및 고찰에 대하여 기술한다.

II. 기존의 알고리즘

LPC기반 음성 신호 분석에 포함되어 있지 않는 예측 오차 $e(n)$을 푸리에 변환 (fourier transform)한 후 큐프런시 (quefrency)영역에서 잔여 켑스트럼 (RCEP)을 구하는 일반적인 알고리즘은 다음과 같다[5].

1. 예측 오차 $e(n)$을 구한다.

$$e(n)=(s(n)-\tilde{s}(n))=\frac{1}{G}[s(n)-\sum_{k=1}^{p}a_k s(n-k)]$$

여기서 $s(n)$은 음성 샘플이고, G는 여기 (excitation) 신호의 이득, a_k 는 LPC계수, p는 LPC분석 차수이다.

2. FFT (fast fourier transform)로 스펙트럼 진폭을 구한다.

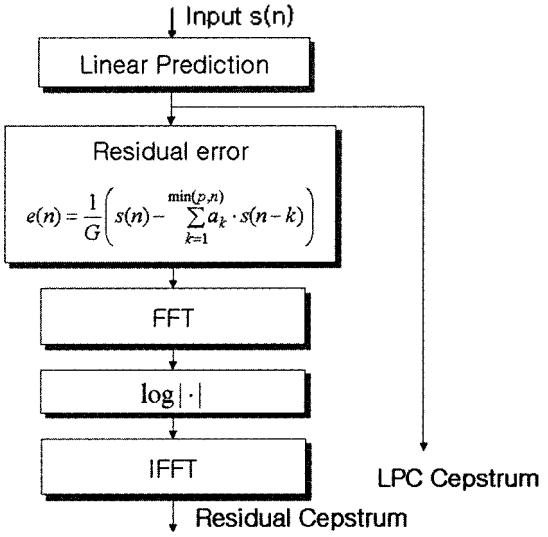

그림 1. 푸리에 변환에 의한 잔여 켑스트럼 추출 모식도
Fig. 1. The block diagram of residual cepstrum extraction by Fourier transform.

$$|E(k)|=\left|\sum_{n=0}^{N-1}e(n)\cdot e^{-j2\pi nk/N}\right|$$

3. RCEP $c(n)=\frac{1}{N}\sum_{k=0}^{N-1}\ln|E(k)|\cdot e^{j2\pi nk/N}$ 을 구한다.

그림 1은 푸리에 변환에 의한 잔여 켑스트럼 추출 모식도를 나타내었다.

또한, LPC 기반 켑스트럼 (LPCC)과 더불어 FFT 기반 켑스트럼인 멜 켑스트럼 (MFCC)은 음성 신호의 대표적 특징 파라미터이다. LPC기반 방법은 all-pole 모델에 기인한 스펙트럼 분석을 수행하며, 음성 특징 파라미터를 빠르고 정확하게 추정한다. 반면, FFT기반 켑스트럼 계수는 복소 켑스트럼이 Fourier 스펙트럼의 진폭만으로 계산될 수 있도록 최소 위상열로써 음성 신호를 가정함으로써 얻어진다. FFT 기반 파라미터의 장점은 잡음에 강하며 비균일 (bark, mel)스케일로 주파수를 묶을 수 있는 것이다. 잔여 켑스트럼 (RCEP)은 LPC 기반 파라미터에 포함되어있지 않은 정보를 가지고 있으므로 대표적 FFT 기반 켑스트럼 추출 방식인 멜 켑스트럼을 구하는 방식으로 다음과 같이 추출한다[6].

1. 예측 오차 $e(n)$을 구한다.

$$e(n)=s(n)-\tilde{s}(n)=s(n)-\sum_{k=1}^{p}a_k s(n-k)$$

2. $e(n)$을 자기상관 (auto correlation)한다.

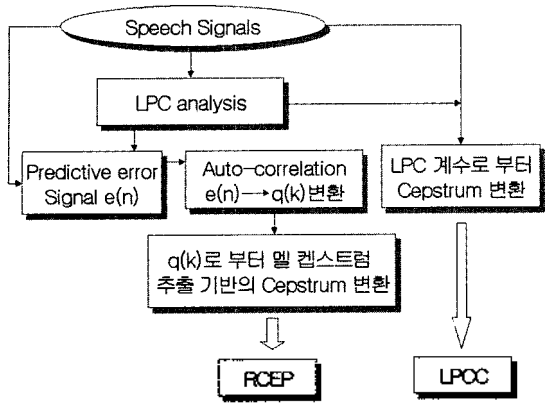

그림 2. 멜 켑스트럼 추출 기반 잔여 켑스트럼 추출 모식도
Fig. 2. The block diagram of Residual Cepstrum based on MFCC analysis method.

$$q(k) = \sum_{n=1}^{N-k} e(n)e(n+k) \quad , \quad 0 \le k \le N-1$$

여기서 N은 윈도우 크기이다.

3. $q(k)$로부터 FFT 기반 켑스트럼을 구한다.

이 절차는 멜켑스트럼 (MFCC)을 구하는 방식과 유사한데, 한가지 다른 점은 입력으로 음성 신호를 사용하는 것이 아니라 $q(k)$를 사용하는 것이다. 주요 절차는 다음과 같다.

(a) $q(k)$에 영점을 첨부해 준다.
(b) FFT로 $q(k)$의 스펙트럼 진폭을 얻는다.
(c) 20필터 뱅크를 만들기 위해 멜스케일에 따라 스펙트럼 진폭을 그룹화 한다.
(d) $q(k)$의 로그 진폭 스펙트럼을 계산한다.
(e) 역코사인 변환후 첫 번째 파라미터로 정규화한다.

그림 2에 멜 켑스트럼 추출 기반의 잔여 켑스트럼 추출 모식도를 나타내었다.

본 논문에서는 그림 2와 같이 멜 켑스트럼 추출 기반의 20차 잔여 켑스트럼을 20차 LPCC와 함께 화자 파라미터로 사용한다.

III. 제안하는 가중치 방법

그림 3은 "안녕하세요" 발음에 대한 전구간 예측 오차 신호 $e(n)$이며 그림 4와 5는 II장에서 설명된 푸리에 변환 방식으로 구한 2명의 화자에 대한 128차의 전구간 평균 잔여 켑스트럼 (RCEP)의 값이다. 그림 4와 5에서 보듯이 예측 오차 $e(n)$으로 부터 구한 잔여 켑스트럼 값에는 화자의 고유성분인 피치 (pitch)성분이 나타남을 알 수 있다. 이러한 화자의 특성을 극대화 해줌으로써 화자간 변이를 더욱 크게 해 줄 수 있다.

많은 화자 확인 시스템에서 피치 주기를 개인 정보로 사용하여 인식율의 향상을 가져 왔다[7]. 그러나, 실제 피치는 잡음 환경에서 구하기 어려울 뿐 아니라 사칭에 대해 취약하다고 잘 알려져 있다[8].

따라서, 잔여 켑스트럼에서 얻어지는 피치 주기만을 파라미터로 사용하기 보다는 피치 성분을 고려하여 화자별로 적절한 가중치를 주는 것이 타당하다.

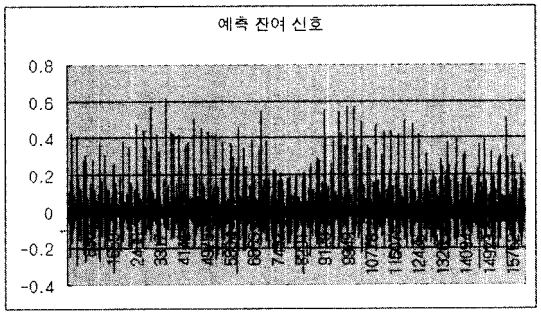

그림 3. 연결 단어 "안녕하세요"의 예측 잔여 신호
Fig. 3. Prediction residue signal of the word, "an-nyoung-ha-se-yo".

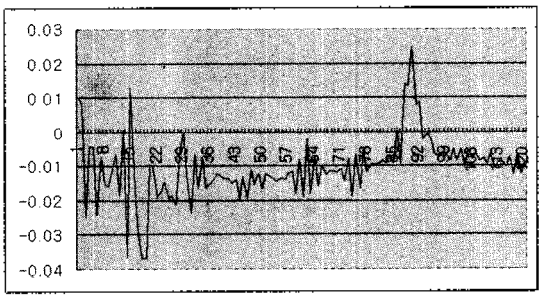

그림 4. 128차 FFT 길이의 푸리에 변환에 의한 전구간 평균 잔여 켑스트럼 (화자 A)
Fig. 4. Total frame average residual cepstrum by Fourier transform (speaker A).

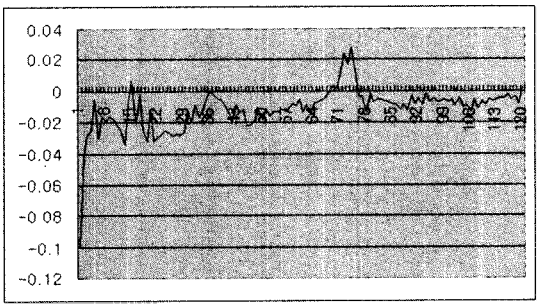

그림 5. 푸리에 변환에 의한 전구간 평균 잔여 켑스트럼 (화자 B)
Fig. 5. Total frame average residual cepstrum by Fourier transform (speaker B).

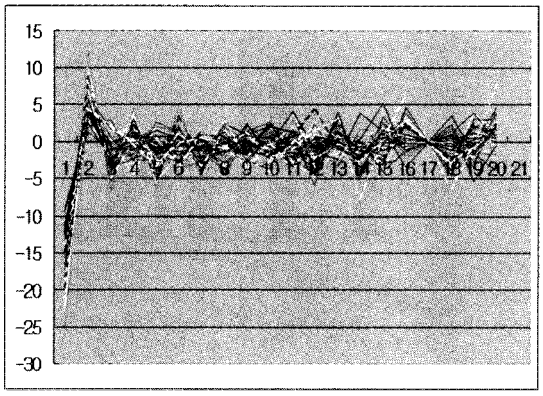

그림 6. 전구간 잔여 켑스트럼
Fig. 6. The total frame RCEP.

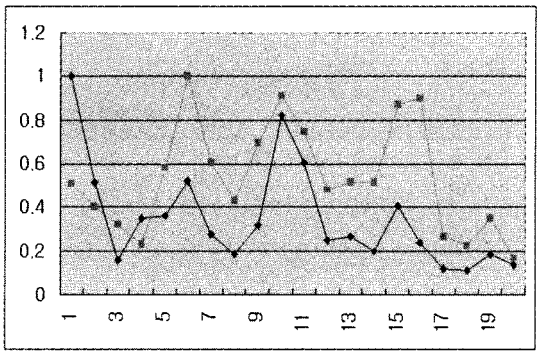

그림 7. 제안한 화자별 가중치 함수
Fig. 7. The proposed personal weighting function.

본 논문에서는 피치주기 부분의 잔여 켑스트럼 값이 변이가 큰 것에 착안하여 가중치로써 화자별 학습 데이터로 구한 잔여 켑스트럼의 화자내 변이 (intra-speaker variance)의 분산값을 다음과 같이 가중함수로 제안한다.

$$average \quad M(k) = \left(\sum_{t=1}^{t=T} R_t(k) \right) / T$$

T: 총프레임수, k Rcep차수 R: residual cepstrum

$$variance \quad v(k) = \sum_{n=1}^{n=N} (M(k) - M_n(k))^2 / N^2$$

N : 학습 발음수

$$weighting \quad w(k) = v(k) / \underset{k}{Max}\ v(k)$$

(단, $1 \le k \le p$)

그림 6은 한 화자에 대한 3회 발음한 학습 데이터의 전구간 잔여 켑스트럼 값이다. 그림 6의 20차 Mel-scale RCEP에는 그림 4나 5에서 보았던 두드러진 특징은 찾을 수 없으나 제안하는 방식으로 가중치를 준다면 화자 고유 특성인 피치 주기를 적절히 강조하는 것임은 분명하다. 그림 7은 제안한 방식으로 구한 화자 A와 화자 B의 화자별 가중함수이다.

잔여 켑스트럼 (RCEP)에만 포함되어 있는 개인별 특성, 즉, 피치 (pitch) 성분이 그림 7의 가중함수의 개인별 차이를 가져온다고 할 수 있다. 이러한 화자 고유의 가중치 함수의 값을 잔여 켑스트럼에 곱하게 되면 유사도 측정시 화자간 거리를 더욱 큰 차이가 나게 할 수 있다.

그림 8은 제안하는 화자 확인 시스템의 학습 모식도이다. 음성 신호로부터 각각 20차의 LPCC와 잔여 켑스트럼을 구한 후, 제안한 방식의 개인별 가중치함수를 잔여 켑스트럼에 적용한 후 개인 코드북을 생성하고 화자 모델을 생성한다.

그림 9는 제안한 화자 확인 시스템의 인식 모식도이다.

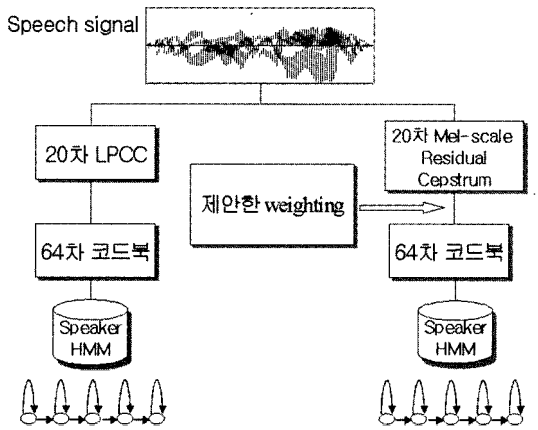

그림 8. 제안한 화자 확인시스템 학습 모식도
Fig. 8. The training block diagram of proposed speaker verification system.

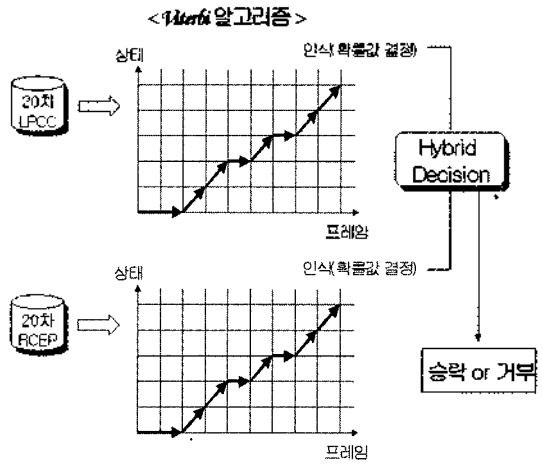

그림 9. 제안한 화자 확인시스템 인식 모식도
Fig. 9. The recognition block diagram of proposed speaker verification system.

표 1. LPCC를 사용한 화자 인식 오류율
Table 1. False speaker verification rate using LPCC as parameters.

	오류 거부율 (FR)	오류 인식율 (FA)
남 자	15.3 %	12.5 %
여 자	18.7 %	14.7 %
총 계	17.0 %	13.6 %

표 2. RCEP를 사용한 화자 인식 오류율
Table 2. False speaker verification rate using RCEP as parameters.

	오류 거부율	오류 인식율
남 자	27.8 %	23.6 %
여 자	31.5 %	27.5 %
총 계	29.65 %	25.55 %

표 3. LPCC와 RCEP를 동시에 사용한 화자 인식 오류율
Table 3. False Speaker verification rate using LPCC as well as RCEP as parameters at the same time.

	오류 거부율	오류 인식율
남 자	10.2 %	8.6 %
여 자	11.9 %	10.9 %
총 계	11.05 %	9.75 %

표 4. 제안한 웨이팅을 사용한 화자 인식 오류율
Table 4. False speaker verification rate using weighted RCEP and LPCC as parameters at the same time.

	오류 거부율	오류 인식율
남 자	8.4 %	6.8 %
여 자	8.6 %	8 %
총 계	8.5 %	7.4 %

음성 신호로부터 추출된 20차의 LPCC와 가중된 20차의 잔여 켑스트럼을 개인 코드북과 유클리드 거리 측정을 통하여 얻은 관측열로 미리 만들어진 화자 모델과의 우도비 (likelihood ratio)를 비터비 (Viterbi)알고리즘에 의하여 계산한 후, 각각의 정규화된 확률값을 더하여 학습시 만들어진 사전 문턱치와 비교하여 승낙 혹은 거부를 결정한다.

IV. 실험 환경 및 결과

본 논문에 사용된 단어는 남녀 각각 50명에 의해 발음된 한 단어 ("안녕하세요")로 다양한 환경 (시끄러운 강의실, 조용한 연구실, 길거리 환경 등)에서 구성되었다. 3번의 발성으로 HMM 학습 알고리즘에 의해 각각의 화자 모델을 만들었고, 각 화자에 대해 100회의 본인에 대한 인증 실험과 4900회의 사칭 실험을 행하였다.

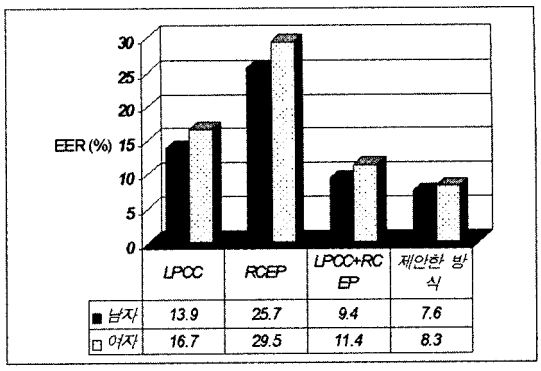

그림 10. 특징 파라미터별 화자 인식 오류율 비교
Fig. 10. The comparison of false speaker verification rate using various parameters.

실험에 사용된 음성 데이터의 샘플링 주파수는 11.025 kHz 이고 240포인트 (Point)의 음성 샘플을 한프레임 (Frame)으로 하여 80포인트씩 shift하여 2/3씩 중첩되도록 하고 해밍창 (Hamming Windows)을 취한후 각각 20차의 LPC-켑스트럼과 잔여 켑스트럼을 구하였다. 이렇게 얻어진 20차 LPC-켑스트럼과 잔여 켑스트럼을 이용하여, k-means 알고리즘으로 64개의 코드워드로 구성된 개인별 코드북 벡터를 생성하였다. 또한, HMM상태수 (State number)는 이산 HMM 5개로 하고, left-to-right 모델을 적용하였다. 인식 실험에서는 로그함수 (log function)로 계산 속도를 줄인 변형 비터비 알고리즘 (Alternative Viterbi algorithm)이 사용되었다.

표 1에서부터 표 4까지 LPCC, 잔여 켑스트럼, LPCC와 잔여 켑스트럼을 동시에 사용한 경우 및 제안한 가중치를 준 경우의 남자, 여자 오류 거부율과 오류 인식율을 나타내었다.

잔여 켑스트럼만을 특징 파라미터로 사용한 경우 (표 2), 인식율이 우수하지는 않았지만 시간에 따른 화자의 변이나 환경에 따른 변화에 우수한 것으로 나타났다. 또한, 제안한 가중치를 준 잔여 켑스트럼과 LPCC를 사용하였을 경우의 인식율 (표 4)이 가중치를 주지 않은 경우 (표 3)보다 평균 2.45 % 가량 우수하였다.

그림 10은 특징 파라미터별 평균 화자 인식 오류율을 비교하였다.

V. 결론 및 고찰

본 논문에서는 All-pole 모델에서 무시되는 예측 오차 $e(n)$을 LPC 기반 파라미터에 포함되지 않으나 중요한 정보로 보고 FFT 기반 파라미터 추출 방법의 대표적 방법인 멜 켑스트럼 (MFCC)추출 방식으로 잔여 켑스트럼 (residual cepstrum)을 추출하였다. 또한, 잔여 켑스트럼에 포함되어 있는 개인 고유 성분인 피치 성분에 적절한 가중치를 줌으로써 화자간 분별력을 높여주는 기법을 제안하였다. 실험 결과, LPCC와 잔여 켑스트럼을 동시에 사용하였을 경우의 인식 오류율이 LPCC만을 사용하였을 경우보다 4~6%가량의 성능 개선됨을 확인하였고, 제안한 방식의 가중치를 주었을 경우가 가중치를 주지 않은 경우보다 약 2.45%가량 인식율이 향상됨을 확인하였다.

참 고 문 헌

1. J. Makhoul, "Linear Prediction: A Tutorial Review", *Proc. IEEE*, Vol. **63**, No. 4, pp. 561-580, 1975.
2. S. Furui, "Cepstral analysis technique for automatic speaker verification", *IEEE Trans.* ASSP-29, no. 2, pp. 254-272, 1981.
3. G. R. Doddington, "Speaker recognition-Identifying people by their voices", *Proceedings of the IEEE*, vol. **27**, no. 11, pp. 1651-1664, Nov. 1985.
4. Lawrence Rabiner, Biing-Hwang Juang, *Fundamentals of Speech Recognition*, Prentice-Hall International Inc, pp. 321-389, 1993.
5. P. Thevenax, H. Hugli, "Usefulness of the LPC-residue in text-independent speaker verification," *Speech Communication*, Vol. **17**, pp. 145-157, 1995.
6. J. He, L Liu, G Palm, "On the use of residual cepstrum in speech recognition", *Speech Communications*, Vol. **17**(1-2), 1995.
7. W. Hess, "Pitch Determination of Speech Signals: Algorithm and Devices", Berlin: Springer, 1983.
8. Jayant M. Naik, "Speaker Verification: A Turorial", *IEEE Communication Magazine*, January 1990.

저자 약력

● 위 진 우 (Jin-Woo Wee)

1989년 2월: 광운대학교 전자통신공학과 학사
1991년 8월: 광운대학교 전자통신공학과 석사
1995년 2월: 광운대학교 전자통신공학과 박사수료
1997년 2월~현재: 인천기능대학 정보통신과 교수
※ 주관심분야: 음성신호처리, 멀티미디어, 영상매체 등

● 강 철 호 (Chul-Ho Kang)
한국음향학회지 제19권 제4호 pp. 23-30 참조