

화자 검증 시스템을 위한 PCA 기반 MFDWC 특징 파라미터

A PCA-based MFDWC Feature Parameter for Speaker Verification System

함성준*, 정호열*, 정현열*

(Seong-Jun Hahm*, Ho-Youl Jung*, Hyun-Yeol Chung*)

*영남대학교 정보통신공학과

(접수일자: 2005년 12월 1일; 채택일자: 2005년 12월 23일)

본 논문에서는 화자검증 시스템의 성능향상을 위해서 주성분 분석 (PCA) 기반 Mel-Frequency Discrete Wavelet Coefficients (MFDWC) 추출방법을 제안한다. 제안된 방법에서는 멜척도 (Mel-scale)를 근사화한 각 레벨 (level)의 각 노드 (node) 에너지를 계산하기 위해 기존의 평균치 대신 주성분 분석을 이용한 첫 번째 eigenvector를 이용한다. 이 eigenvector의 제곱의 합은 1로서 일반적인 가중 함수 (weighting function)의 조건을 만족하고, 또한 각 화자마다 서로 다른 값을 갖게 되므로, 화자의 특징을 더 잘 나타내는 MFDWC를 추출할 수 있다. 화자검증은 Gaussian Mixture Model (GMM) 기반의 백그라운드 모델과 화자 모델과의 점수를 비교하는 이진 결정 (binary decision) 방법을 이용하여 Universal 백그라운드 모델 (UBM)과 각 화자 모델의 값을 프레임단위로 비교하여 대상 화자의 수락/거부 여부를 결정하는 방법을 채택하였다. 특징 파라미터에 따른 화자 검증 성능변화를 확인하기 위하여 제안된 화자중속 가중함수를 이용한 MFDWC를 특징 파라미터로 이용한 경우와 Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Cepstral Coefficients (LPCC), 기존의 MFDWC를 특징 파라미터로 이용한 경우에 대하여 성능비교실험을 수행한 결과 각각 0.80%, 5.14%, 6.69%의 향상된 성능을 나타내어 제안한 방법의 유효성을 확인할 수 있었다.

핵심용어: 화자 검증, 주성분 분석, 웨이블릿 변환, Universal 백그라운드 모델

투고분야: 음성처리 분야 (2.5)

A principal component analysis (PCA)-based Mel-Frequency Discrete Wavelet Coefficients (MFDWC) feature parameters for speaker verification system is presented in this paper. In this method, we used the 1st-eigenvector obtained from PCA to calculate the energy of each node of level that was approximated by mel-scale. This eigenvector satisfies the constraint of general weighting function that the squared sum of each component of weighting function is unity and is considered to represent speaker's characteristic closely because the 1st-eigenvector of each speaker is fairly different from the others. For verification, we used Universal Background Model (UBM) approach that compares claimed speaker's model with UBM on frame-level. We performed experiments to test the effectiveness of PCA-based parameter and found that our proposed parameters could obtain improved average performance of 0.80% compared to MFCC, 5.14% to LPCC and 6.69 to existing MFDWC.

Keywords: Speaker Verification, Principal Component Analysis, Wavelet Transform, Universal Background Model

ASK subject classification: Speech Signal Processing (2.5)

I. 서론

Mel-Frequency Cepstral Coefficients (MFCC)는 음성 인식 및 화자 인식에 가장 널리 쓰이는 특징파라미터로서 MFCC 파라미터를 이용한 화자인식에 관한 전반적인 소개는 Gish, H.와 Schmidt, M.[1]의 연구에서 찾아볼 수 있다.

그러나 MFCC에서 멜척도 로그 필터뱅크 (Mel-scaled log filter bank) 에너지에 Discrete Cosine Transform (DCT)를 취하는 MFCC는 몇 가지 단점을 지닌다[2]. 즉, DCT가 모든 주파수 대역에 걸쳐 수행되므로, 잡음에 의한 주파수 대역의 왜곡이 모든 MFCC에 영향을 미칠 가능성이 크다. 이러한 문제를 최소화하기 위해서 subband 인식 방법이 음성인식에서 일부 사용되어 왔다 [3]. 이와 더불어 보다 향상된 음성인식을 위한 특징 추출방법으로서 웨이블릿 변환 (Wavelet Transform; WT)을 이용한 파라미터 추출방법 연구[4]가 수행되고 있으며 이러한 연구의 결과는 웨이블릿 변환을 사용하여 subband 기반 인식기의 단점을 극복할 수 있는 가능성을 보이고 있다. 그 이유는 웨이블릿 변환을 이용한 파라미터가 DCT에 비해 시간과 주파수 분해능이 좋아 위에서 언급한 문제를 해결할 수 있는 가능성이 크기 때문이다[5].

그러나 웨이블릿 변환을 이용한 기존의 MFDWC는 MFCC를 이용한 시스템보다 낮은 성능을 나타냄을 저자들에게 의한 이전 연구[6]에서 확인할 수 있었다. 이는 시간과 주파수축을 따라 분해된 각 노드의 샘플들의 평균을 취함에 따라 시간-주파수 정보를 가진 샘플의 시간 정보를 활용하지 못하므로 기존의 주파수 기반 MFCC에 비해 낮은 성능을 나타낸다고 생각된다.

화자 검증 시스템의 성능은 신호특성의 변이에 매우 크게 영향을 받으므로 화자 간의 변이를 보상하기 위해 정규화 방법이 많이 이용되고 있다[7, 8, 9]. 문맥 독립 화자 인식 방법으로서 화자 특성 변이의 표현과 화자 검증률 면에서 좋은 결과를 나타내고 있는 GMM에 의한 접근방법이 유효하며[10], 이러한 화자 간의 변이를 보상하기 위한 정규화 방법 중 가용한 모든 화자의 데이터를 이용한 하나의 백그라운드 모델을 이용하는 Universal 백그라운드 모델 (UBM)을 이용한 방법이 성능 면에서나 검증 속도 면에서 유효하다고 알려져 있다 [11].

따라서 본 논문에서는 기존의 MFDWC를 추출하는 과

정 중 admissible 웨이블릿 패킷 트리 (Wavelet Packet Tree)의 각 노드의 에너지를 계산하기 위한 방법으로서 노드내의 시간에 따른 데이터 분포를 최대한 활용할 수 있는 주성분 분석[12]을 이용한 화자종속 가중합수를 이용하여 각 화자의 특징을 더 잘 표현하는 특징을 추출하는 방법을 제안하고, UBM을 이용한 프레임 단위의 점수를 이용하여 화자 검증 시스템의 성능을 향상시키고자 한다. 2장에서는 제안된 방법과 관련된 이론인 주성분 분석과 웨이블릿 변환에 대해 간략하고 기존의 MFDWC 추출 방법에 대해 알아본 후, 3장에서는 UBM 기반 화자 검증 방법, 4장에서는 제안된 주성분 분석 기반 MFDWC 특징추출방법에 대해 설명한다. 5장에서는 시스템의 구성 및 실험 결과에 대해 기술하고, 마지막으로 6장에서 결론을 맺는다.

II. 주성분 분석과 웨이블릿 변환 개요 및 MFDWC

2.1. 주성분 분석

주성분 분석은 신호처리, 통계, 신경망 등의 분야에서 널리 쓰이고 있다. 주성분 분석의 기본 아이디어는 주어진 데이터의 분포를 가장 잘 나타내는 선형 변환된 n 개의 성분 p_1, p_2, \dots, p_n 을 구하는 것이다.

주성분 분석의 기본 목적은 주어진 데이터의 차원을 줄이는 것이다. 주성분 분석에 의한 데이터 분포의 표현은 평균 상승 (분산)의 관점에서 최적 선형 차원 감소임이 증명될 수 있다. 그러한 차원 감소는 중요한 이점을 제공한다. 주성분 분석후의 처리 과정에서 계산량을 줄일 수 있고, 처음 n 개의 성분에는 잡음에 의해 발생된 데이터가 포함되지 않을 가능성이 크기 때문에 잡음제거 효과를 얻을 수 있다. 또한 2차원과 같은 매우 낮은 차원의 부 영역으로의 사상 (projection)은 데이터를 시각적으로 표현하는데 매우 유용하다.

2.2. 웨이블릿 변환

웨이블릿 변환은 고주파 영역에서 단구간 윈도우를 사용하는 것이 가능하고, 이러한 특성은 시간 영역과 주파수 영역 모두 분해능이 좋은 신호 표현을 가능하게 한다 [13]. 웨이블릿은 대역통과 필터로 간주될 수 있다. 웨이블릿 변환에서 scaling 함수는 주어진 신호 표현을 위한

무한개의 필터 사용을 피하기 위해 이용된다. 웨이블릿을 이용한 신호 해석과 scaling 함수를 이용하여 주어진 scale까지 신호 표현이 가능하다. 적용을 위한 최적 트리의 선택은 웨이블릿 패킷 접근 방법을 통해 수행될 수 있다.

2.3. MFDWC

MFDWC는 화자 검증에 위한 새로운 특징파라미터로서 웨이블릿 패킷 변환 (Wavelet Packet Transform; WPT)을 적용하여 추출할 수 있다. 인간의 청각능력은 음의 크기에 대하여 근사적으로 대수적인 특성을 나타내며 주파수 분해능은 1kHz이하의 낮은 주파수영역에서는 선형적이고 그 이상의 주파수영역에서는 대수적인 멜척도 (Mel-scale) 특성을 가진다. 이러한 멜척도에 의한 주파수 분석은 음성인식에 유효한 것으로 알려져 있다[14]. MFDWC에서 웨이블릿 패킷 트리의 구조는 MFCC에서 이용되는 멜척도와 유사한 주파수를 갖는 새로운 필터뱅크 형태의 admissible 웨이블릿 패킷 트리로서 Farooq와 Datta[4]에 의해 제안되었다. 여기서 admissible 트리는 한 노드가 분할되지 않거나 분할되어 2개의 자식 노드를 갖는 트리 구조를 말한다. 근사화된 멜척도 필터뱅크의 주파수를 그림 1에 나타내었다. 이진트리를 기반으로 하여 구성되었으므로 기존의 멜척도와는 약간의 차이를 보인다.

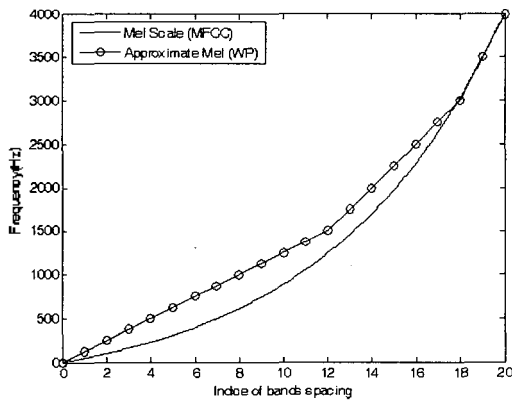


그림 1. 기존의 mel-scale과 근사화된 mel-scale 필터뱅크 주파수
Figure 1. Partitioning of the frequency axis of Mel scale using MFCC and its approximation using MFDWC.

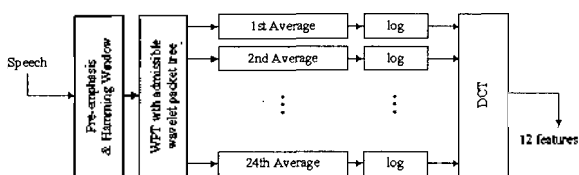


그림 2. 기존의 MFDWC 추출과정
Figure 2. Procedure for extracting MFDWC.

웨이블릿 패킷 변환은 시간축과 주파수축 상에서 분할된 웨이블릿 패킷을 이용하기 위해 적용된다. MFDWC에서 웨이블릿 패킷 변환을 이용한 주파수축 분할은 주어진 scale에서 같은 크기의 두 대역으로 주파수를 나누는 conjugate mirror 필터를 반복적으로 적용하여 얻을 수 있다[13]. 웨이블릿 패킷 트리에서 각 scale은 depth j와 subspace p를 이용하여 나타낸다. 각 부모 (parent) 노드 (j,p)에서 두 직교 기저 (orthogonal base)는 저역 통과 필터와 고역통과 필터로 정의된다[4].

MFDWC는 그림 1의 멜척도를 근사화한 admissible 웨이블릿 패킷 트리의 각 노드 (j,p)로부터 추출된다. 웨이블릿 패킷 트리의 각 노드 (j,p)의 평균 로그 에너지에 DCT를 취함으로써 최종 MFDWC를 얻을 수 있다. 이 과정은 그림 2에 나타내었다.

III. UBM 기반 화자 검증 방법

기존의 화자 검증 시스템에서는 화자 등록 단계에서 미리 결정된 문턱치를 이용한다[15]. 그러나 이 경우, 새로운 화자를 등록할 때마다 새로운 문턱치를 결정해야 하므로 실시간 시스템 구성에 있어 다음과 같은 문제점을 가진다. 첫째, 이전의 등록화자에 대한 음성 정보가 가용해야 하므로 시스템은 모든 화자의 음성정보를 항상 가지고 있어야 한다. 둘째, 각 화자의 모델과 언급한 모든 화자의 음성을 테스트하여 최적화된 문턱치를 계산해야 한다. 이는 실시간 시스템의 관리와 사용에 있어 모두 불편함을 주는 요소로 작용한다. 이에 본 논문에서는 기존의 UBM (가용한 모든 화자의 음성 정보를 이용하여 만든 하나의 모델)과 화자 모델과의 프레임 유사도만을 비교하는 이진 결정 (binary decision) 방법을 이용하여 프레임 단위로 누적하여 검증을 요구한 화자의 점수를 계산하였다. UBM과의 유사도가 더 클 경우 -1, 화자 모델과의 유사도가 더 클 경우 +1의 프레임 점수를 준다. 이에 대한 수식은 아래와 같다.

$$S(X) = \sum_{i=1}^T S_i(X_i) \tag{1}$$

$$S_i(X_i) = \begin{cases} +1 & p_i(X_i | \lambda_0) \geq p_i(X_i | \lambda_{UBM}) \\ -1 & p_i(X_i | \lambda_{UBM}) > p_i(X_i | \lambda_0) \end{cases} \tag{2}$$

여기서 $S_i(X_i)$ 는 검증요구 화자의 점수를 나타내며 T

는 입력음성의 총 프레임 수, λ_0 는 검증을 요구한 화자 모델, λ_{UBM} 은 UBM, $S_i(X_i)$ 는 프레임 단위의 점수를 나타낸다. 이 점수를 누적하여 총 점수 $S(X)$ 가 0보다 크면 승인하고 작으면 거절한다. 가용한 프레임의 수 $\times 2$ (화자모델과 회귀계수모델)가 화자가 얻을 수 있는 최대 점수가 되고, 이때의 문턱치는 화자와 가용한 프레임 수에 관계없이 0으로 고정된다. 따라서 새로운 화자를 등록할 때 문턱치를 다시 계산할 필요가 없고, 다른 모든 화자의 음성 정보를 시스템이 가지고 있어야 하는 문제도 해결 가능하다. 또한 기존의 시스템에서는 특정 프레임의 로그유사도가 화자의 승인여부 결정에 큰 영향을 미치는 경우가 있었지만, 이 방법에서는 프레임의 가중치를 균등하게 하므로 그러한 역효과를 막을 수 있다.

IV. 제안된 PCA 기반 MFDWC

기존의 MFDWC 추출에서는 각 노드의 에너지를 평균치를 통해 계산한다. 이 때, 평균치를 구하기 위한 가중함수는 각 노드의 샘플수의 역수로서 모든 노드와 화자에 대해서 동일하다. 이 평균치는 웨이블릿 변환 이용한 노드내의 시간축 정보를 이용하지 못해 성능저하의 원인이 될 수 있다. 따라서 이러한 노드 내 시간정보를 이용하기 위한 각 레벨의 각 노드 에너지를 계산하기 위해 주성분 분석을 적용하여 구해진 첫 번째 eigenvector를 화자종속 가중함수로 이용한다. 이 eigenvector의 제곱의 합은 1로서 가중 함수 (weighting function)의 조건을 만족하고, 또한 각 화자마다 다른 값을 갖게 되므로, 화자의 특징을 더 잘 나타내는 MFDWC를 추출할 수 있다. 제안된 화자종속 가중함수를 추출하는 과정은 다음과 같다.

1. 입력음성 전처리 수행
2. 근사화된 Mel-scale WPT에 따른 wavelet 변환 수행
3. 각 노드의 샘플 정보 수집

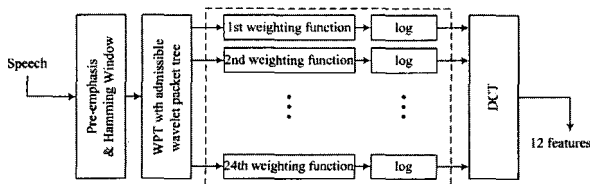


그림 3. 화자종속 MFDWC 특징추출과정
Figure 3. Procedure for extracting speaker dependent MFDWC.

4. 각 노드별 PCA 적용 후, 첫 번째 eigenvector를 구함
5. 구해진 eigenvector의 각 성분을 제공하여 화자종속 가중함수를 얻음

임의의 노드 에너지를 E 라 하고, 노드의 샘플값을 C , 화자종속 가중함수를 W 라고 하면, 위의 5단계를 거쳐 얻은 화자종속 가중함수를 이용하여 각 노드에너지 E 는 식 (3)을 이용하여 계산된다.

$$E_n^k = \sum_i (C_n^k(i)^2 \times W_n^k(i)) \quad (3)$$

여기서 n 은 level, k 는 노드, i 는 노드내의 샘플을 나타낸다.

화자종속 MFDWC 전체 특징 추출 과정을 그림 3에 나타낸다. 위의 가중함수와 노드 에너지를 구하기 위한 과정은 그림 3의 점선 내 블록에 해당한다. 특징 추출과정에서 노드 에너지를 구하기 위한 과정을 제외하고 기존의 MFDWC 추출 과정과 동일하다.

V. 시스템 구성 및 실험 결과

5.1. 데이터베이스 (DB) 및 전처리

GMM에 기반한 UBM 기반 화자 검증 방법을 이용하여 각 파라미터에 대해 화자검증실험을 실시하였다. 실험에는 KLE452 DB와 ETRI445 DB, ETRI 화자검증용 DB를 이용하였다. 이때, 화자검증실험은 ETRI445 DB의 20명, KLE452 DB 38명, ETRI 화자검증용 DB 100명의 화자를 이용한 158명에 대해서 수행되었다. 각 DB의 화자는 모두 남성을 이용하였다. UBM은 각 화자모델을 학습하는데 이용한 모든 화자의 데이터를 이용하여 학습하였다.

GMM 모델 학습을 위해서는 프레임 에너지 문턱치 2000을 넘는 1000, 1500, 2000 프레임을 LPCC, MFCC, MFDWC, 제안된 MFDWC 각각의 방법으로 추출하여 특징파라미터로 이용하였다. MFDWC를 추출하기 위한 웨이블릿 변환을 위해서는 바기역 필터의 특성을 갖고, 저역통과 필터의 차수가 9, 고역통과 필터의 차수는 7인 Daubechies 9/7 필터를 이용하였다. 테스트를 위해서는 학습에 사용되지 않은 단어 또는 문장에서 300 프레임을 추출하여 이용하였다. 각 화자모델의 혼합수는 2, 4로 가변, UBM의 혼합수는 16으로 고정하였다. 전체

표 1. 특징 추출을 위한 전처리단계의 조건

Table 1. Analysis conditions for extracting feature parameter.

Feature Extraction Method	LPCC	MFCC	MFDWC
Sampling Rate	16kHz	16kHz	16kHz
Pre-emphasis coefficient	0.97	0.97	0.97
Hamming Window	yes	yes	yes
Frame length (samples)	512	512	512
Frame Shift (samples)	256	256	256
Cepstrum vector dimension	12	12	12
Cepstral Mean Normalization	yes	yes	-

리단의 분석조건은 표 1과 같다. 표 1에서 제안된 MFDWC와 기존의 MFDWC는 같은 조건이므로 MFDWC로 나타내었다.

5.2. 시스템의 구성

시스템은 화자모델 생성 및 화자 검증의 두 단계로 구성하였다.

그림 4는 화자 검증 시스템의 구성을 나타낸다. 검증 요구화자의 GMM 모델유사도와 World 백그라운드 모델의 유사도를 비교하여 0보다 크면 승인하고 작으면 거절하는 것을 구성도에 나타내었다.

표 2와 3은 각 특징파라미터의 승인율과 거절률을 나타내고, 표 4는 성능을 나타내고 있다. 승인율 (CA)과 거절률 (CR)은 다음의 식으로 계산한 결과이다.

$$CA(\%) = 100 - FA \tag{4}$$

$$CR(\%) = 100 - FR$$

표 2와 3에서 승인율의 경우 제안된 방법이 전반적으로 가장 좋은 결과를 보였고, 거절률의 경우 MFDWC와 LPCC가 거의 비슷한 성능을 나타내었다. 전체적인 비교를 위해 식 (5)를 이용하여 성능을 나타내었을 때,

$$Perf(\%) = 100 - (FA + FR) \tag{5}$$

각 파라미터의 평균성능은 그림 5와 같이 제안된 방법의 경우 94.99%, MFCC 94.19%, LPCC 89.85%,

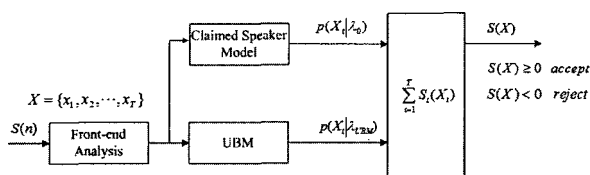


그림 4. 화자 검증시스템의 구성도

Figure 4. Block diagram of speaker verification system.

표 2. 각 특징추출방법에 따른 승인율 (%)

Table 2. Correct Acceptance Rate of each feature extraction method (%).

특징 파라미터	학습 프레임					
	1000		1500		2000	
	2	4	2	4	2	4
LPCC	85.44	78.48	91.14	92.41	95.57	98.10
MFCC	89.87	85.44	97.47	97.47	98.73	98.73
MFDWC	79.75	77.85	89.24	93.67	93.67	97.47
제안된방법	91.77	92.41	95.57	96.84	98.73	99.37

표 3. 각 특징추출방법에 따른 거절률 (%)

Table 3. Correct Rejection Rate of each feature extraction method (%).

특징 파라미터	학습 프레임					
	1000		1500		2000	
	2	4	2	4	2	4
LPCC	99.72	99.94	99.51	99.80	99.36	99.65
MFCC	99.74	99.93	99.33	99.70	99.16	99.55
MFDWC	99.82	99.94	99.59	99.76	99.47	99.56
제안된방법	99.54	99.83	98.94	99.36	98.65	98.94

표 4. 각 특징추출방법에 따른 성능 (%)

Table 4. Performance of each feature extraction method (%).

특징 파라미터	학습 프레임					
	1000		1500		2000	
	2	4	2	4	2	4
LPCC	85.16	78.42	90.65	92.21	94.93	97.75
MFCC	89.61	85.37	96.80	97.17	97.90	98.29
MFDWC	79.57	77.79	88.83	93.43	93.14	97.02
제안된방법	91.31	92.23	94.51	96.20	97.39	98.30

MFDWC 88.30%의 순으로 나타났다.

그림 6은 ETRI445 데이터베이스의 'cjd' 화자의 로그 유사도와 정규 스코어를 사용한 경우를 각각 비교한 결과를 나타낸다.

로그 유사도를 이용한 경우 특정 프레임의 유사도에 따라 변이가 커서 화자검증 여부에 큰 영향을 주는 것을 확인할 수 있다. 반면, 정규화된 (+1, -1) 점수를 이용할 경우 특정 프레임의 유사도를 효과적으로 정규화 함을 그림 6으로부터 알 수 있다.

그림 7은 ETRI445 데이터베이스의 'asw', 'aym' 각 화자의 가중함수를 나타낸다. 각 노드의 가중함수가 화자마다 서로 다름을 확인할 수 있다.

화자종속 가중함수를 이용할 경우의 장점으로서 주성분 분석을 적용한 후의 첫 번째 eigenvector는 특정 화자의 데이터 분포의 분산을 가장 잘 표현하는 주성분 (PC)을 구하기 위한 독립 변환을 수행하는 vector이고,

이 eigenvector를 이용하면 특정 화자에서는 데이터 분포의 분산을 가장 많이 포함하는 주성분을 구하게 되지만, 다른 화자의 경우는 데이터 분포의 분산을 적절하게 표현하지 못하는 주성분이 될 가능성이 크다[16]. 여기서 검증 성능 향상의 이유를 찾을 수 있다. 검증 실험 결과로부터도 제안된 파라미터가 화자인식에 유효함을 확인할 수 있었다.

VI. 결론

본 논문에서는 화자검증 시스템의 성능향상을 위해서 주성분 분석 (PCA) 기반 Mel-Frequency Discrete Wavelet Coefficient (MFDWC) 추출방법을 제안하였다. 제안된 방법에서는 멜척도 근사화한 각 레벨의 각 노드 에너지를 계산하기 위해 기존의 평균치 대신 주성분 분석을 이용한 첫 번째 eigenvector를 이용한다. 화자검증은 Gaussian Mixture Model (GMM) 기반의 백그라운드 모델과 화자 모델과의 점수를 비교하는 이진 결정 방법을 이용하여 Universal Background Model (UBM)과 각 화자 모델의 값을 프레임단위로 비교하여 대상화자의 수락/거부 여부를 결정하는 방법을 채택하였다.

특징파라미터에 따른 화자검증 성능을 확인하기 위하여 제안된 방법과 MFDWC, MFCC, LPCC를 특징 파라미터로 이용한 경우에 대하여 각각 성능비교실험을 수행하였다. 성능평가 실험결과 전체 시스템의 성능 면에서 MFCC에 비해 0.80%, LPCC에 비해서는 5.14%, 기존의 MFDWC보다는 6.69%의 향상된 성능을 나타내어 제안된 화자중속 가중합수를 이용한 특징추출이 화자검증에 유효함을 알 수 있었다. 향후 독립 성분 분석 (ICA)을 이용한 특징 파라미터에 관한 연구와 Hidden Markov Network (HMNet) 시스템과 GMM 시스템을 결합한 문맥 종속/독립 화자 검증 시스템에 대해서 연구를 진행할 예정이다.

감사의 글

본 연구는 정보통신부 및 정보통신 연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었습니다.

참고 문헌

- Gish, H. and Schmidt, M., "Text-Independent Speaker recognition", IEEE Sig. Proc. Magazine, 18-32, 1994.
- Gowdy, J.N. Tufekci, Z., "Mel-Scaled discrete Wavelet Coefficients for Speech Recognition," ICASSP, 3, 1351-1354, 2000.
- Boulevard, H. and Dupont, S., "A New ASR Approach Based on Independent Processing and Recombination of Partial Frequency Bands", ICSLP, 1996.
- Farooq, O. and Datta, S., "Mel Filter-Like Admissible Wavelet Packet Structure for Speech Recognition", IEEE Signal Processing Letters, 8 (7), 196-198, 2001.
- Goswami, J. & all., "Fundamental of Wavelets", Wiley, 1999.
- 함성준, 박하중, 김민정, 김주곤, 정호열, 정현열, "Wavelet 변환을 이용한 GMM 기반 문맥독립 화자검증 시스템의 성능평가", 한국음향학회 학술발표대회 논문집 24 1(s), 한국음향학회, 297-300, 2005.
- A. E. Rosenberg, J. Delong, C.-H. Lee, B.-H. Juang and F. K. Soong, "The use of cohort normalized scores for speaker verification", in Proc. Inter. Conf. on Spoken Language Processing (ICSLP'92), 599-602, 1992.
- A. L. Higgins, L. Bahler and J. Porter, "Speaker Verification using Randomized Phrase Prompting", Digital Signal Processing, 1, 89-106, 1991.
- T. Matsui and S. Furui, "Concatenated Phoneme Models for Text Variable Speaker Recognition", in Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'93), 391-392, 1993
- S. Furui, "An overview of speaker recognition technology," in Acoustic speech and speaker recognition(C.-H Lee, F. K. Soong, and K. K. Paliwal, eds.), Ch. 2, 31-56, Kluwer Acad. Publ., 1996.
- D. A. Reynolds, "Comparison of Background Normalization Methods for Text-Independent Speaker Verification," in Proc. Eurospeech97, 963-967, 1997.
- Jolliffe, I.T. *Principal Component Analysis*, (Springer Verlag), 1986.
- G. Strang, T. Nguyen, *Wavelets and Filter Banks*, (Wellesley-Cambridge Press, 1997).
- S. B. Davis, P. Mermelstein, "Comparison of Parametric Representations of Monosyllabic Word Recognition in Continuously Spoken Sentences," IEEE Trans. Acoustics, Speech, and Signal Processing, 28, 357-366, 1980.
- Dat Tran, Michael Wagner, "A Proposed Likelihood Transformation for Speaker Verification", in proc. IEEE International Conference on, 2, 1069-1072, 2000.
- C. Tadj, A. Benlahouar, "Speaker Characterization Using Principal Component Analysis and Wavelet Transform for Speaker Verification", in Proceedings of Eurospeech, 2981-2984, 2003.

저자 약력

• 함성준 (Seong-Jun Hahm)



2003년 2월: 국민대학교 전자공학과 (공학사)
2004년 3월~현재: 영남대학교 대학원 정보통신공학과
석사과정
* 주관심분야: 화자인식, 음성인식, 디지털 신호처리

• 정호열 (Ho-Youl Jung)



1988년 2월: 아주대학교 전자공학과 (공학사)
1990년 2월: 아주대학교 전자공학과 (공학석사)
1993년 2월: 아주대학교 전자공학과 (박사수료)
1998년: (프)리옹국립응용과학원(INSA de Lyon)
전자공학전공(공학박사)
1998년 4월~1998년 12월: (프)CREATIS 박사후 과정
1999년 3월~현재: 영남대학교 전자정보공학부 교수
* 주관심분야: 음성, 영상 신호처리, 인공지능, 디지털
위터피팅

• 정현열 (Hyun-Yeol Chung)



1975년: 영남대학교 전자공학과 (공학사)
1989년: 일본 동북대학교 정보공학과 (공학박사)
1989년 3월~현재: 영남대학교 전자정보공학부 교수
1992년 7월~1993년 7월: 미국 CMU Robotics 연구소
객원연구원
1994년 12월~1995년 2월: 일본 토요하시기술과학대학
외국인 연구자
2000년 6월~2000년 8월: 미국 Qualcomm Inc. 수석
엔지니어
* 주관심분야: 음성인식, 화자인식, 음성합성 및 DSP 응용분야