# 개별 음향 정보를 이용한 화자 확인 알고리즘 성능향상 연구*

# The Study for Advancing the Performance of Speaker Verification Algorithm Using Individual Voice Information

이 재 형** · 강 선 미***
**Jeyoung Lee · SunMee Kang**

## ABSTRACT

In this paper, we propose new algorithm of speaker recognition which identifies the speaker using the information obtained by the intensive speech feature analysis such as pitch, intensity, duration, and formant, which are crucial parameters of individual voice, for candidates of high percentage of wrong recognition in the existing speaker recognition algorithm. For testing the power of discrimination of individual parameter, DTW (Dynamic Time Warping) is used. We newly set the range of threshold which affects the power of discrimination in speech verification such that the candidates in the new range of threshold are finally discriminated in the next stage of sound parameter analysis. In the speaker verification test by using voice DB which consists of secret words of 25 males and 25 females of 8 kHz 16 bit, the algorithm we propose shows about 1% of performance improvement to the existing algorithm.

**Keywords: Speech Feature Analysis, Speaker Recognition, Pitch, Intensity, Duration, Formant**

## 1. 서 론

대개의 경우 전화 등의 매체를 통해서 주변 상황이 일정 부분 잡음이 있음에도 불구하고 상대방의 말소리만 듣고도 누구인지를 판별한다. 이는 일정 기간동안 알고 지내온 사람들의 경우 그 목소리의 주요한 특징들에 의해 우리의 귀가 매우 잘 훈련되었기 때문이다. 그러나, 같은 가족들의 목소리 특히, 쌍둥이의 경우나 성대 모사를 통한 모방된 목소리의 경우는 분별이 매우 어려운 것을 종종 경험한다. 일반적으로 사람은 이미 훈련된 패턴과 유사한 패턴이 입력된 경우 거의 대부분 정성적인 분석으로 동일인으로 결론을 짓지만, 기계의 경우는 현재 수준으로는 대개 정량적인 평가로만 결과를 얻어내므로 아직은 사람에 비하여 낮은 인

* 본 논문은 한국과학재단 목적기초연구(R01-1999-000-00229-0) 지원으로 수행되었습니다.
** 보이스미디어텍 (주)
*** 서경대학교 컴퓨터과학과

식률을 보인다. 그러나 정량적인 분석에 정성적인 값을 추가한다면, 기계는 인간이 감지하지 못하는 미세한 변화도 변별할 수 있게 된다.[1]

그러므로 본 논문에서는 과연 인간은 음성의 어떤 정보를 이용하여 구별하는지 그 방법을 예측하고, 그 방법을 컴퓨터와 같은 기계화된 수단에 적용 할 수 있는 알고리즘을 보완 개선하기 위한 음성의 특징을 분석해 보고자 한다. 이를 위해서 다음과 같은 분석 작업을 집중한다. 먼저 화자인식에 사용할 수 있는 음향학적인 특징추출의 실험을 실시한다. 이는 언어학이나 음성학에서 주로 사용한 음향적 파라미터로 사용하는 Pitch, Intensity, Formant, Duration 등의 파라미터를 이용하여 화자인식의 가능성을 제시하여 준다. 두 번째로 화자 인식 특징추출의 적용 및 보완 방법을 제시한다. 기존의 특징추출 알고리즘에서 화자 인식률의 한계를 좀 더 본질적으로 해결해 보기 위한 방법으로 음향학적인 파라미터를 적용하여 화자 인식률의 향상을 꾀한다. 세 번째로는 기존의 화자 인식을 보완하기 위한 DTW (Dynamic Time Warping)를 이용한 하이브리드적인 인식 적용 실험을 실시한다. 스펙트럼에서 얻을 수 있는 제한된 음향 파라미터의 적은 데이터량을 감안하여 DTW를 이용한 병렬적인 개별 파라미터 패턴을 비교함으로서 HMM (Hidden Markov Model)에서 나타나는 FA (False Acceptance)의 감소를 얻어낸다. 상기 실험 결과를 통해서 본 논문에서는 8 kHz 16 bit의 남녀 각 25 명의 비밀단어로 구성된 음성 DB로 화자 확인 실험을 한 결과 기존 알고리즘의 결과보다 약 1%의 성능 향상을 얻어낼 수 있었다.

## 2. 음성의 음향 파라미터 분석

### 2.1 MFCC를 이용한 화자 확인 시험 결과 분석 및 개선점

일반적으로 많이 사용되는 MFCC를 특징 값으로 한 HMM을 이용한 기존의 화자 확인 알고리즘 실험 결과는 다음과 같다. 실험에 사용된 음성 DB는 전화선에서의 녹음 환경을 고려하여 8 Khz로 하였으며 양자화 수는 16 비트를 선택하였다. 단어는 남녀 각 25 명씩 50 명의 개인 암호로 구성하였다. 각각 자기 자신의 비밀단어를 15 번 발음하였으며 이 중 5 개를 학습 데이타로, 10 개의 단어는 시험용으로 사용하였다. 그리고 사칭자에 의한 화자 인증 실험을 위하여 타인의 비밀단어 49 개를 각각 3 번씩 발음하여 실시하였다.

먼저, 화자 모델을 만들기 위해 각 화자가 발음한 10 개의 비밀단어를 사용했다. 화자 확인의 FRR (False Reject Rate)를 측정하기 위해 사용된 데이터는 각 화자 당 5 개씩 총 250 개, 그리고 FAR (False Accept Rate)를 측정하기 위해 사용된 데이터는 각 화자 당 3 개씩 총 7350 개를 사용했다. 얻어진 실험 결과를 이용하여 유사도 측정에 사용된 log-likelihood를 frame 수로 normalization하였다. 실험결과는 모델 1의 경우는 word 모델링시 각 화자 당 1 개씩, 총 50 개의 단어를 사용한 경우이고 모델 2의 경우는 각 화자 당 2 개씩, 총 100 개의 단어를 사용했을 경우이다.

표 1. 각 word 모델에 따른 인식 실험 결과 비교 (단위: %)

| 모델 1 | | | | 모델 2 | | | |
|---|---|---|---|---|---|---|---|
| Threshold | FRR | FAR | HTER | Threshold | FRR | FAR | HTER |
| 1.1 | 2.00 | 0.79 | 1.39 | 1.1 | 2.40 | 0.67 | 1.13 |
| 1.4 | 2.00 | 0.64 | 1.32 | 1.4 | 2.80 | 0.44 | 1.58 |

일반적으로 threshold 값의 변화에 따라서 threshold 값을 증가시키면 FRR은 증가하고 FAR은 감소한다. 따라서 이러한 관계를 이용하여 threshold 값을 어떻게 정할지 결정해야 한다(그림 1 참조).

상기 실험 결과를 살펴보면 threshold 값의 설정에서 FRR과 FAR 간의 상호 상반되는 결과를 가져오므로 threshold 값의 조절에 의한 성능향상은 크게 기대할 것이 없다고 본다. 그러므로 성능향상을 위해서는 먼저 기존에 사용한 특징 값 외의 필수적으로 새로운 정보가 요구된다. 즉, 다른 분야에서 연구 분석되는 새로운 파라미터의 적용이 필요하다고 본다. 현재 얻어진 95%의 화자 확인율은 이미 상용화된 다른 생체인식 수단에 비해서는 아직 낮은 편이므로 이의 성능 향상을 위한 구체적인 대안이 요구되어진다.

### 2.2 음향 파라미터를 이용한 화자확인실험[2]

#### 2.2.1 주요 음향 파라미터 추출

언어학적인 입장에서 봤을 때 강세나 억양은 목소리를 변별하는 중요한 특징이 된다. 이러한 특징들과 음향적으로 가장 중요한 관계를 가지는 것으로 성대의 기본주파수, 지속시간, 음파의 강도를 들 수 있다. 본 논문에서는 기본적인 음향 파라미터로 피치, 강세, 포만트를 들 수 있으며 이에 대한 분석을 실시하였다.[3]

피치(pitch)를 추출하는 알고리즘으로는 LPC에 의한 방법, 평균차 함수법, 저 표본화에 의한 LPC 방법, Cepstrum에 의한 방법[4]을 들 수 있다. 본 논문에서는 LPC에 의한 방법과 Cepstrum에 의한 방법을 사용하여 피치 값을 추출하였다.

강도(intensity)는 피치나 포만트 보다는 간단히 구할 수 있다. 일반적으로 전체 에너지값을 구하여 정규화과정을 위한 로그를 취하면 된다. 이 과정은 초기 전처리 과정에서 음성의 유무를 판정하는 끝점검출 과정에서도 가장 많이 사용된다. 그러나 본 논문에서는 음성의 유무가 아닌 음소와 음소사이의 연결음에서의 억양과 운율에 따른 강도의 궤적을 각 프레임별로 구하게 된다. 정규화과정 중에는 마이크 입력에 따른 에러를 최소하기 위하여 스케일링을 이용하여 상대적인 값으로 변환하였다. 그러나 이러한 과정의 유무와 관계없이 마이크볼륨에 대한 조절 과정이 반드시 필요하다.

포만트(formant)의 경우 Burg의 방법을 사용하였다.[5][6] 가장 널리 알려졌으며 알고리즘 구현이 간단하다. 먼저 FFT나 LPC[7]을 이용하여 스펙트럼 포락선을 구해낸 다음 이중 피크치를 나타내는 로브(Lob)값들 중 제 1 포만트부터 3 포만트까지의 후보를 선출한 뒤에 인터폴레이션 방법이나 경로탐색을 이용하여 각 구간의 정점인 부분의 좌표값을 얻어내었다.[8]

음성학이나 언어학에서는 화자 확인 수단으로 분절음의 **지속시간(duration)**을 추출하여 실험하기를 권유하고 있으나 현재의 기술로는 분절음의 지속시간은 컴퓨터나 기계적인 세그

먼트 레이블링만으로는 얻어낼 수 없다. 대부분 자동 세그먼트 프로그램은 간단하게 일차적으로 레이블링을 마치고 마지막으로 사람이 직접 눈으로 확인하여 세그먼트 레이블링하여 분절음을 구분하게 된다. 아직까지는 컴퓨터로 분절음을 100% 구분할 수 없다. 비교적 많이 알려진 웨이브렛과 같은 신호처리 알고리즘에서는 분절음을 구분하는 여러 가지 논문이 나오고 있으나 실시간 처리를 요구하는 시스템의 특성을 고려하여 본 실험에서는 HMM에서의 Viterbi decoding과정에서의 백트래킹을 이용하여 추출하였다. 다시 반복주기 검사를 통하여 모음성분에 따라 분절음의 구간을 알아낸 뒤 지속시간을 산출하였다.

2.2.2 각 파라미터의 변별력 시험 모델

각 파라미터의 성능 실험을 위해서 DTW[8][9]를 인식 모델로 사용하였다. Sakoe와 Chiba의 제약조건을 이용하여 구현하면 DA는

$$\begin{aligned} DA(n,m) = MIN( & DA(n-1, m-2) + 2d(n,m-1) + d(n.m), \\ & DA(n-1, m-1) + 2d(n,m), \\ & DA(n-2, m-1) + 2d(n-1,m) + d(n.m) \quad ) \end{aligned} \tag{1}$$

로 표현되며, accumulated distance DA(n,m)가 구해지면 reference pattern과 test pattern의 총 distance D(N,M)은

$$D(N,M) = \frac{D_A(n,m)}{No} \tag{2}$$

로 결정된다. 여기에서 No는 normalization factor로서 No 값으로는 보통 reference pattern의 frame수인 N을 사용한다.

2.2.3 실험결과 및 분석

본 실험에 사용된 음성 DB는 화자들 간의 음향 파라미터를 기준으로 하기 위하여 전국의 6 개 지역별(서울, 부산, 대구, 대전, 광주, 춘천) 화자들을 대상으로 남녀 30 명에 대하여 수집된 것이다. 이는 음향분석법에서 제시되어지는 분류기준으로 단어를 음운 환경별로 나눈 것이다. 단어의 내용은 화자의 주변정보들로서 주민등록번호와 같은 숫자음 20 개와 자신의 이름과 같은 일상적인 단어와 실험단어로 연세말뭉치 연구에서 조사된 4,200만 어절 가운데 빈도순위 200 위 이내의 50 단어와 숫자음을 조합한 20 단어를 사용하였다.

표 2. 각 음향 파라미터만을 이용한 화자 확인율

| 파 라 메 터 | 화자 확인율(%) |
|---|---|
| 피 치 | 75 |
| 강 세 | 72 |
| 제 1 포만트 | 60 |
| 제 2 포만트 | 74 |
| 제 3 포만트 | 72 |
| 인식 결과 | 79.8 |

2.2.4 음향 파라미터 적용시의 결과 분석

1) 각각의 개별 파라미터만으로는 원하는 화자 확인율을 얻을 수 없다.

음성학이나 언어학분야 심지어 범죄인의 음성을 판단하는 분야에서는 음향적 파라미터와 포만트만으로 화자를 확인 또는 식별을 할 수 있다고 하지만, 실제 실험결과에서 보면 기대치 이하의 낮은 화자 확인율을 얻었다.

2) 피치와 강세를 적용하기에는 파라미터로서의 데이터량이 부족하다.

음성인식에서 사용하는 MFCC나 LPC 등은 FFT의 데이터량 보다는 적지만 음성으로서의 특징을 충분히 압축하여 HMM과 같은 통계모델에 적용할 수 있었다. 그러나 피치나 강세는 특징의 중요도에 비하여 적은 데이터량으로 인해 개별적으로 분석하기는 어렵다. 추출 알고리즘의 한계로 인하여 피치나 강세를 추출하고자 하는 프레임을 더 이상 짧게 할 수는 없다.

3) 포만트의 특징 비교는 1차 실험에서의 MFCC와 같이 스펙트럼을 고려한 실험이므로 중복된 과정이다.

4) 부가적인 화자확인 시스템의 파라미터

파라미터 각각을 개별적으로 적용하여 사용할 수는 없으나 보조적인 수단이나 MFCC의 특징추출과 같이 상호보완적으로 사용하여 기존의 화자 확인 시스템의 성능을 향상시키는 파라미터로 이용이 가능하다.

## 3. 화자 확인 성능 향상을 위한 제안된 하이브리드 알고리즘

상기 소개한 개별 실험 결과를 토대로 음성 및 음향 파라미터를 같이 적용하는 새로운 알고리즘을 제안하였다. 그 타당성 검증을 위해서 실험에 사용한 음성 DB는 기존 화자 확인 성능 시험에 사용한 50 명 DB를 사용하였다.

3.1 제안된 하이브리드 알고리즘

기본적인 생각은 다음과 같다. 먼저 기존 알고리즘의 성능한계는 threshold 값에 따라서 어느 정도의 제한된다는 점이다. 그래서 threshold를 값으로 정하는 것이 아니고 구간으로 정해서 그 구간에 들어오는 인식 후보들에 대한 음향학적 파리미터로 정밀 분석을 함으로서 인식률의 향상을 기대한다는 것이다. 그러므로 1 단계 실험에서 얻어진 threshold의 범위를 구한 후 전체 FR과 FA를 고려하여 화자인증을 위하여 FRR과 FAR이 0%인 각각의 이상적인 한계 threshold를 결정하고 이에 따른 완충지역을 설정한다. 인식하고자 하는 화자의 threshold값이 완충지역을 벗어나 상한선 이상이나 하한선 이하로 나타나면 2 단계 과정을 거칠 필요 없이 accept나 reject로 판별되어진다. 그러나 만약 완충지역 이내에 존재한다면 2 단계 과정으로 들어가서 음향 파라미터를 추출한 후 DTW의 유사도를 이용하여 최종 판별을 하게 된다.

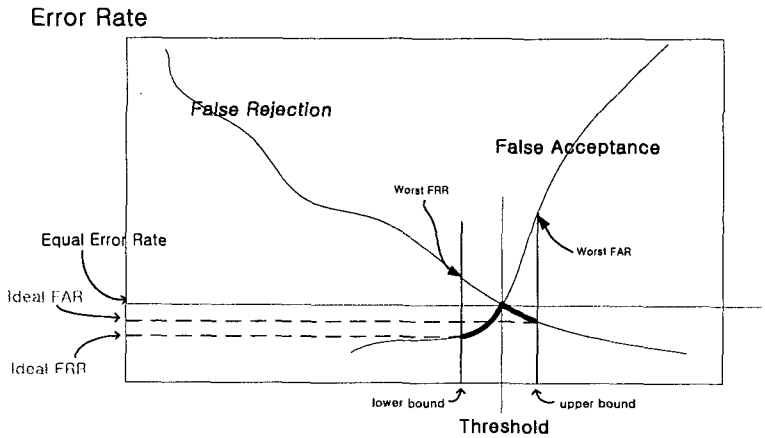


그림 1. Threshold 경계 범위

단계 2에서는 녹음 된 화자의 5 번 발음한 음성 데이터를 웰드 모델과의 거리와 표준 편차를 구한 결과 각각의 화자별로 저장한 뒤 다음 테스트 하고자 하는 화자의 10번 발음을 입력받아 이를 웰드모델과의 거리를 구한 뒤 레퍼런스 패턴과의 거리와 비교하여 표준편차이내이면 본인임을 확인(Accept)하고 표준편차 이상이면 거부(Reject)한다. 식 (3)과 식 (4)를 사용한다.

$$D(T:R_{all}) = | D((T:W) - ( \sum_{i=1}^{5} D(R_i:W) / 5) | \quad (3)$$

D ($R_i$ : W) : 음성 DB의 i 번째 레퍼런스와 웰드 모델 W와의 거리

D (T : W) : 화자확인을 위한 테스트 패턴과 웰드 모델 W와의 거리

IF $D(T:R_{all}) < SD( D(R_i:W) )$ 이면 Accpet

IF $D(T:R_{all}) > SD( D(R_i:W) )$ 이면 Reject (4)

SD(D(Ri:W)): 웰드모델과 I개의 레퍼런스 모델과의 거리 평균에 대한 표준편차

본 논문에서 제안한 화자 확인 알고리즘의 전체 흐름도는 그림 2와 같다.

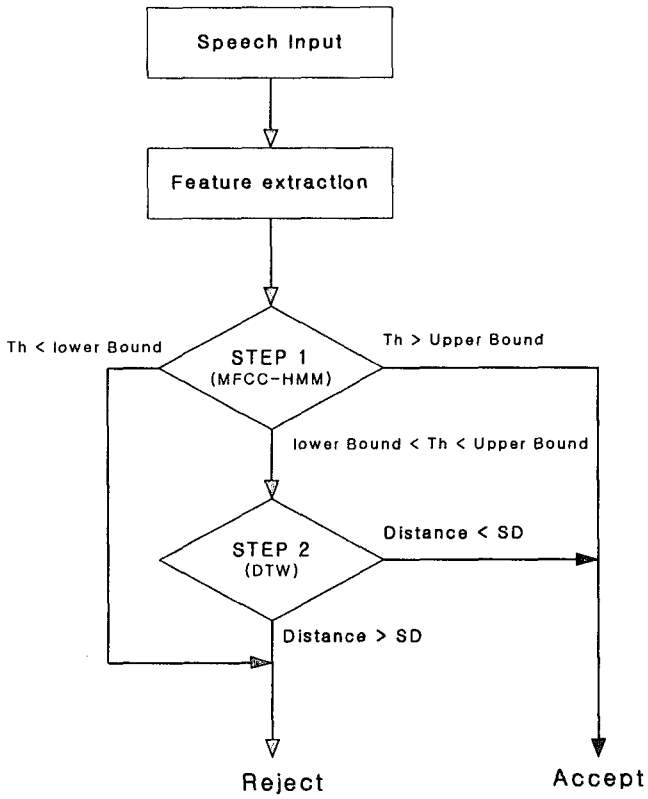


그림 2. 제안한 화자 확인 시스템의 Flow chart

3.2 실험결과 및 분석

FRR이 이상적으로 0%에 가깝게 하는 threshold는 대략 -2.5 이하이다. 표 3에 나타난 시험 결과는 threshold 경계에 따라서 FRR과 FAR을 얻은 것이 아니고, FRR과 FAR 각각에 대해서 얻어낸 threshold의 경계 범위이다.

표 3. Threshold 범위

| Threshold 경계 | FRR(%) 독립 | FAR(%) 독립 | HTER(%) |
|---|---|---|---|
| - 6.0 ~ 5.0 | 0.8% | 0.6% | 0.7 |
| - 2.5 ~ 2.0 | 0.6% | 0.5% | 0.55 |
| - 1.0 ~ 0.5 | 0.4% | 0.3% | 0.35 |

표 3에서 나타낸 threshold의 범위를 이용하여 2 단계로 각 음향 파라미터를 적용한 실험 결과는 표 4에서 표 7과 같다.

먼저 피치만을 적용하였을 때는 표 4와 같이 단계 1의 threshold의 경계를 너무 크게 주

면 화자 확인율이 전체적으로 낮아서 단계 1의 고정 threshold에서의 결과보다도 성능이 떨어짐을 알 수 있다. 그러나 threshold의 경계범위를 적게 줄수록 인식 단계 1의 결과에 근접함을 알 수 있다.

표 4. 피치만 적용시의 오류율

| Threshold | FRR(%) | FAR(%) | HTER(%) |
|---|---|---|---|
| - 6.0 ~ 2.0 | 11.0 | 8.0 | 9.5 |
| - 2.5 ~ 1.0 | 5.0 | 4.5 | 5.5 |
| - 1.0 ~ 0.5 | 2.2 | 0.8 | 1.55 |

표 5와 표 6에 나타난 실험 결과를 살펴보면 피치와 강세 적용에 따른 결과는 별 차이가 없다. 음향 파라미터 실험 결과에서는 강세가 의외로 높은 분별력을 보여 주지만, 표 5에 나타낸 결과에는 낮은 변별력을 보여 준다. 이는 본 실험에서 사용되어진 음성 DB는 음향 파라미터 실험에서 사용한 음성 DB에서와 같은 지역적 특색이 나타나지 않은 영향이라고 추정된다. 표 6에서 피치와 강세를 상대적인 값으로 정규화하여 얻어진 값을 2 차의 벡터열로 간주하여 인식 실험을 하였다.

표 5. 강세만을 적용시의 오류율

| Threshold | FRR(%) | FAR(%) | HTER(%) |
|---|---|---|---|
| - 6.0 ~ 2.0 | 11.6 | 9.2 | 10.4 |
| - 2.5 ~ 1.0 | 6.2 | 5.2 | 5.7 |
| - 1.0 ~ 0.5 | 2.0 | 0.75 | 1.35 |

표 6. 피치와 강세를 동시에 적용시의 오류율

| Threshold | FRR(%) | FAR(%) | HTER(%) |
|---|---|---|---|
| - 6.0 ~ 2.0 | 9.2 | 7.5 | 8.35 |
| - 2.5 ~ 1.0 | 4.9 | 5.4 | 5.15 |
| - 1.0 ~ 0.5 | 1.8 | 0.6 | 1.2 |

표 7에는 지속시간 값을 이용한 실험 결과인데, 기존 알고리즘만을 적용한 결과와 동일하게 나왔다. 이에 본 실험에서의 지속시간 추출 과정에서 문제가 있다고 판단되어 지속시간의 결과 값을 직접 수동으로 확인한 결과 세그먼트 레이블링의 에러에 기인한 것으로 확인했다. 이는 본 실험에서 지속시간을 추출하는 세그먼트 레이블링 기법의 적용 오류로 표 7과 같이 지속시간이 화자확인의 파라미터로는 부 적절하다고는 볼 수 없으므로 차후 실험에서는 지속시간을 정확하게 추출할 수 있는 방법들을 모색해야 한다고 본다.

표 7. 지속시간 적용시의 오류율

| Threshold | FRR(%) | FAR(%) | HTER(%) |
|---|---|---|---|
| - 6.0 ~ 2.0 | 30 | 28 | 29 |
| - 2.5 ~ 1.0 | 22 | 21 | 21.5 |

표 8은 각 실험에서의 화자 확인율 결과를 %로 표시한 것이다. 그 결과 threshold의 범위를 -1.0에서 0.5로 선택하였을 때 가장 높은 화자 확인율을 얻음을 알 수 있다. 표 8은 그림 1의 Worst FRR과 FAR을 피하기 위한 실험 결과이다. 0.0을 기준으로 threshold 값이 클수록 본인임을 나타내며 작을수록 사칭자를 나타내게 된다. 표에서 양수(+)쪽의 값의 범위를 작게 둔 이유는 이 시스템은 보안을 위한 화자 확인 실험이므로 FAR을 좀더 낮추기 위해서이다.

표 8. 단계별 화자 확인율

| Threshold 경계범위 | Only STEP 1 | STEP 1 + STEP 2 | | | |
|---|---|---|---|---|---|
| | Only MFCC | Only Pitch | Only Intensity | Pitch + Intensity | Duration |
| - 10.0 ~ 5.0 | 62% | 52.0% | 55.0% | 61% | 43% |
| - 6.0 ~ 5.0 | 82% | 79% | 76% | 77.5% | 49% |
| - 3.0 ~ 2.0 | 89% | 85% | 84% | 84.2% | 52% |
| - 1.0 ~ 0.5 | 93.5% | 96.0% | 96.2% | 96.8% | 62% |

Threshold의 범위를 -10.0에서 5.0으로 선택하였을 때 기존의 화자 확인 방법(MFCC-HMM 화자인증 시스템)보다도 훨씬 낮아진다는 것을 알 수 있다. 그림 1에서 나타낸 이상적인 FAR과 FRR을 얻기 위하여 threshold의 범위를 좁혀가면서 실험한 결과 마지막의 -1.0~0.5 사이로 두었을 때 threshold를 0.0으로 두었을 때 95%의 화자 확인율을 얻어낸 것보다는 낮아졌지만 제안되어진 2 번째 과정을 거친 후에는 약 1%~2% 정도의 화자 확인율 상승 효과를 볼 수 있었다. 피치나 강세와는 달리 지속시간은 화자 확인율을 전체를 떨어뜨리는 효과를 보였으며 향후 연구 과제에 있어서는 정확한 지속시간 추출에 관한 보완이 요구되면 Saoke방식의 DTW가 아닌 다른 방식의 DTW와 함께 벡터열에 대한 다양한 유사도 측정기법을 사용하면 좀더 나은 결과를 얻을 수 있다고 기대한다.

## 4. 실험 결과 분석 및 결론

본 논문에서 사용한 기존의 화자 인식 알고리즘은 음성인식에서 사용하는 특징추출방법인 MFCC를 사용하였고, 패턴 비교분석 방법으로 음성인식에서 그 효과가 검증된 모델인 HMM을 적용하였다. 본 논문에서 사용한 인증기 모델의 성능 평가를 위해서 남녀 각각 25 명의

음성 데이터를 녹음하여 50 개의 개인적인 암호인 단어를 대상으로 실험한 결과 약 95%의 화자 확인율을 얻었다. 그러나 이러한 성능은 지문인식이나 홍채인식과 같은 다른 종류의 생체인식기술과 비교했을 때 다소 낮은 인식률을 나타낸다.

인식률을 높이기 위해 기존에 사용한 발성기관이나 청각기관을 모델로 하여 얻어진 특징 벡터들이 아닌 음향학적인 측면과 발음기관의 특징을 구체화 할 수 있는 특징 값들을 고려한 실험을 시도하였다. 일반적으로 발성학이나 언어학을 연구하는 분야에서 음성 분석에 사용되는 파라미터로 피치나 강세, 지속시간, 포만트 등이 있으며 이러한 값만을 이용하여 DTW를 이용하여 실험 한 결과 약 80%의 화자 확인율을 얻었다. 이 결과는 앞서 HMM모델로 실험한 결과와 비교했을 때 낮은 수치를 나타내지만, 각 음향 파라미터 개별적인 정보만으로도 화자확인을 위한 파라미터로서 사용할 수 있다는 것을 나타낸다.

이러한 실험 결과를 바탕으로 기존 방법에 음향적 파라미터 값을 첨부한 하이브리드적인 알고리즘을 제안하게 되었다. 현재 화자인식 분야에서는 HMM과 GMM (Gaussian Mixture Model)을 연계하여 실험한 논문이 다수 발표되고 있으며 대부분 이러한 실험들은 먼저 HMM을 이용하여 화자가 발음한 단어를 확인하고 화자가 정확한 단어를 발음하였다면, 다시 2 차로 GMM을 이용하여 화자 확인을 한다. 본 논문에서는 1 차적으로 HMM에 의한 음성인식 작업 및 화자확인 작업을 진행하고, 2 차적으로 적용하는 모델은 음향학적 특징을 나타내는 파라미터들을 이용하여 DTW로 패턴비교를 하여 화자를 확인하는 방법을 제안하였다. 1차 단계에서는 MFCC 특징을 HMM을 이용하여 비교함으로써 화자가 발음하는 단어뿐 만 아니라 그 단어를 말하는 화자의 습관 등을 비교하며 이는 MFCC에 포함되어져 있는 포만트 곡선에 따른 각 화자의 성도기관의 상이점을 구분하게 된다. 이러한 1 단계 실험에서 걸러진 결과를 이용하여 2 단계 실험에서는 화자마다 다른 성대 특징과 지속 시간을 이용하여 얻어낸 결과를 이용하여 최종적인 화자 인증 알고리즘을 구성하였다. 그 결과 약 1-2%의 화자 확인율의 향상을 얻었으며, 각 단계에서 실험한 화자 인증 시스템의 취약점을 보완 할 수 있다는 결론을 얻었다.

향후 연구 방향은 2 단계에서의 화자 확인율을 높이기 위하여 DTW에서의 유사도 측정 방법을 다양하게 적용시켜야 하며 duration의 추출 방법의 개선을 이용하여 2 단계 실험의 화자 확인율을 끌어올려 1 단계에서의 threshold의 범위를 넓게 설정하여 다양한 조건에서의 화자 확인 실험을 하고자 한다.

## 참 고 문 헌

[1] Furui, S. 1994. "An overview of speaker recognition technology." *ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, 1-9.

[2] 이재형, 양병곤, 고도흥, 강선미. 2000. "우리말에서의 피치·포만트정보를 이용한 화자확인용 특징비교." *한국음성과학회 제9회 학술발표논문집*, 131-136.

[3] 강선미. 2002. "화자확인을 위한 화자 고유의 음성특징추출과 적응 모델에 관한 연구." *3차년도 중간 보고서*, 과학재단 특정기초과제.

[4] 박경범. 2000. *개정판 음성의 분석 및 합성과 그 응용*, 31~63. 도서출판 그린.

[5] Vaseghi, Saeed V. 1996. *Advanced Signal Processing and Digital Noise Reduction.* pp. 186-200. Queen's University of Belfast, UK.
[6] Press, William H., Brian P. Flannery, Saul A. Teukolsky & Willam T. Vetterling. 1988. *Numerical Recipes in C.* pp. 450~452. Cambridge University Press.
[7] 한진수. 2001. *음성신호처리.* pp 12, pp32~36.
[8] Huang, Xuedong, Alex Acero & Hsiao Wuen Hon. 2001. *Spoken Language Processing.* pp 322~323. PH PTR.
[9] Booth, I., M. Barlow & B. Watson. 1993. "Enhancements to DTW and VQ Decision Algorithms for Speaker Recognition." *Speech Communication*, 13(3), 427-433.

접수일자: 2002. 7. 31.
게재결정: 2002. 9. 13.

▲ 이재형
서울특별시 성북구 안암동 5가 고려대학교 산학관 6층 (우: 136-701)
보이스미디어텍 (주)
Tel: +82-2-923-8830 Fax: +82-2-923-8830
E-mail: jhlee@voicemediatech.com

▲ 강선미
서울특별시 성북구 정릉동 16-1 (우: 136-704)
서경대학교 컴퓨터과학과
Tel: +82-940-7291 Fax: +82-2-919-5075
E-mail: smkang@skuniv.ac.kr