

# 화자 확인에서 SPRT를 위한 새로운 테스트 데이터 생성

## A New Test Data Generation for SPRT in Speaker Verification

서 창 우\*, 이 기 용\*  
(Chang-Woo Seo\*, Ki-Yong Lee\*)

\* 숭실대학교 정보통신전자공학부

(접수일자: 2002년 11월 12일; 수정일자: 2002년 12월 30일; 채택일자: 2003년 1월 6일).

본 논문에서 제안하는 방법은 화자 확인 (speaker verification)에서 시퀀스 확률비 테스트 (SPRT: sequential probability ratio test)를 위한 시작 프레임의 샘플 시프트를 이용해서 새로운 테스트 데이터를 생성하는 방법이다. SPRT는 테스트 계산량을 줄일 수 있는 효과적인 알고리즘이다. 그러나 테스트의 결정과정에서 SPRT 방법은 입력신호가 확률밀도함수로부터 독립적이고 균일하게 분포되어 있다는 가정하에 수행할 수 있으며, 또한 발성길이 짧은 데이터에는 적용하기에 적절하지 못하다. 제안한 방법은 시작 프레임의 샘플 시프트를 통한 새로운 테스트 데이터를 생성하는 방법이기 때문에 테스트 데이터의 길이에 상관없이 SPRT를 수행할 수 있다. 또한 SPRT 방법에서 고려해야 하는 데이터의 상관성은 주성분 분석 (principal component analysis)을 이용함으로써 효과적으로 제거하였다. 실험 결과 제안한 방법은 기존의 방법보다 샘플 시프트를 위한 데이터의 계산량은 약간 증가하였지만, 등가 오류율 (EER: equal error rate)에서 평균 0.7% 이상 좋은 성능 결과를 보였다.

**핵심용어:** 화자확인, 시퀀스 확률비 테스트, 은닉마코프모델, 주성분 분석, 유사도 비

**주요분야:** 음성처리 분야 (2.5)

This paper proposes the method to generate new test data using the sample shift of the start frame for SPRT (sequential probability ratio test) in speaker verification. The SPRT method is a effective algorithm that can reduce the test computational complexity. However, in making the decision procedure, SPRT can be executed on the assumption that the input samples are usually to be i.i.d. (Independent and Identically Distributed) samples from a probability density function (pdf), also it's not suitable method to apply for the short utterance. The proposed method can achieve SPRT regardless of the utterance length of the test data because it is method to generate the new test data through the sample shift of start frame. Also, the correlation property of data to be considered in the SPRT method can be effectively removed by employing the principal component analysis. Experimental results show that the proposed method increased the computational complexity of data for sample shift a little, but it has a good performance result more than a conventional method above the average 0.7% in EER (equal error rate).

**Keywords:** Speaker verification, SPRT, HMM, PCA, Likelihood ratio

**ASK subject classification:** Speech signal processing (2.5)

## I. 서론

화자인식에서 신호처리의 목적은 신뢰할 수 있는 정보를 추출하여 사용하기에 편리한 형태로 음성신호를 변환하는 것이다. 현재 가장 널리 사용되고 있는 분석은 선형 예측 cepstrum 계수 (LPCC: linear prediction cepstral

coefficients)와 멜주파수 cepstrum 계수 (MFCC: mel frequency cepstral coefficients) 방법 등이 있다. 그러나 이러한 일반적인 특징벡터는 화자인식 시스템에서 신호의 상관성 때문에 성능감쇠가 일어난다[1,2]. 따라서 높은 인식률을 얻기 위해서는 많은 파라미터 개수를 필요로 한다. 그러나 파라미터의 증가는 많은 계산량을 필요로 하기 때문에 실시간 구현을 보다 어렵게 만든다. 주성분 분석 (principal component analysis)은 원 신호의 특징벡터를 선형변환을 통하여 특징벡터의 상관성을 줄어

책임저자: 이기용 (kylee@soongsil.ac.kr)  
156-743 서울시 동작구 상도 5동  
숭실대학교 정보통신전자공학부  
(전화: 02-827-0992; 팩스: 02-786-558)

는 방법이다[3]. 따라서 본 논문에서는 신호의 상관성 제거를 위해서 주성분 분석을 적용하였다. 화자 확인에서 SPRT를 사용한 방법은 다른 일반적인 방법보다 적어도 두 가지 이점을 가진다. 첫 번째는 이론적으로 확인 오류들을 제어할 수 있으며, 마지막으로 성능의 주어진 레벨을 위한 입력 샘플의 수가 어떤 다른 방법보다 작다[5]. 그러나 SPRT 방법은 입력신호가 확률밀도함수 (PDF)로부터 비상관성이라는 가정하에 수행할 수 있으며, 또한 발생길이 \$L\$가 짧은 데이터에는 적용하기에 적절하지 못하다. 제안한 방법은 시작 프레임의 샘플 시프트를 통한 새로운 테스트 데이터를 생성하는 방법이기 때문에 테스트 데이터의 길이에 상관없이 SPRT를 수행할 수 있다. 또한 SPRT 방법에서 고려해야 하는 데이터의 상관성은 주성분 분석을 이용함으로써 효과적으로 줄일 수 있다.

논문은 다음과 같이 구성되었다. II장에서는 신호의 상관성 제거를 위해서 주성분 분석을 설명하였다. III장은 화자확인 방법에서 연속밀도 HMM을 이용한 일반적인 SPRT 방법을 소개하고, 그리고 IV장에서는 논문에서 제안한 시작 프레임의 샘플 시프트를 통한 새로운 테스트 데이터를 생성하는 방법을 설명하였다. 그리고 실험 결과는 V장에서 설명되며, VI장에는 결론을 서술하였다.

## II. 주성분 분석을 이용한 신호의 상관성 제거

여러 개의 변수들에 대하여 얻어진 다변량 자료를 분석 대상으로 하는 주성분 분석 (principal component analysis)은 다차원 벡터를 축소, 요약하는 차원의 단순화와 함께 일반적으로 서로 상관성 있는 벡터들간의 복잡한 구조를 분석하는데 그 목적을 두고 있다. 이를 위하여 입력 데이터를 상관관계가 없는 새로운 좌표계로 선형변환시킨다[1,3].

여러 개의 상관성이 밀접한 벡터를 \$O = (o\_1, o\_2, \dots, o\_p)\$라 하자. 주성분 분석은 \$O\$가 서로 상관관계가 없는 새로운 벡터 \$Z = (z\_1, z\_2, \dots, z\_p)\$로 좌표변환을 시키는 기법이다. 이러한 특성 때문에 주성분 분석은 특징벡터의 추출 방법에서 사용된다. 데이터 집합에서 주성분을 찾는 일반적인 방법은 공분산 행렬 (covariance matrix)의 고유벡터 (eigen vector)를 계산하는 것이다. 먼저 프레임 길이가 \$T\$이고 \$p\$차 특징벡터 \$O\_i = (O\_i(1), O\_i(2), \dots, O\_i(T))\$로부터 선형변환 행렬 \$\Omega^T\$를 구하기 위해서 전

체 평균과 공분산 행렬을 구해야 한다[4].

$$\mu_i = \frac{1}{T} \sum_{t=1}^T O_i(t) \quad i=1, 2, \dots, p \quad (1)$$

$$\sigma_{ij} = \frac{1}{T} \sum_{t=1}^T (O_i(t) - \mu_i)^T (O_j(t) - \mu_j) \quad i, j=1, 2, \dots, p \quad (2)$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \quad (3)$$

여기서 \$\mu\_i\$는 \$i\$번째 성분의 평균, \$\Sigma\$는 \$p \times p\$로 \$\sigma\_{ij}\$를 원소로 갖는 공분산 행렬이다. \$\Sigma\$의 \$(i, j)\$번째 성분은 \$i \neq j\$일 때, \$O\$의 \$i\$번째와 \$j\$번째 성분의 공분산을 나타내고 \$i = j\$일 때는 \$O\$의 \$j\$번째 성분의 분산을 나타낸다. 공분산 \$\Sigma\$는 다음과 같이 나타낼 수 있다.

$$\Sigma = \sum_{i=1}^p \lambda_i v_i v_i^T \quad (4)$$

여기서 \$\lambda\_i\$는 \$\Sigma\$의 \$i\$번째 고유치 (eigen value)이고, \$v\_i\$는 고유값 \$\lambda\_i\$에 대응되는 정규화된 고유벡터 (eigen vector)이다. 이들은 \$p \times p\$인 직교 행렬 \$\Omega^T = I\$를 이룬다.

위의 설명과 같이 \$i\$번째 시퀀스의 \$i\$번째 특징벡터 \$O\_i(t)\$와 주성분 \$Z\_i(t)\$의 관계는

$$Z_i(t) = v_i^T O_i(t) \quad (5)$$

이고, 성분 전체의 관계를 식으로 나타내면 다음과 같다.

$$Z(t) = \Omega^T O(t) \quad (6)$$

식 (6)에서 \$\Omega^T\$는 특징벡터 \$O\$를 주성분 \$Z\$로 바꾸기 위한 변환행렬이다. 위의 식으로부터 특징벡터 \$O\$를 신호의 상관성이 제거된 새로운 특징벡터 \$Z\$를 구할 수 있다.

## III. 화자확인을 위한 SPRT

SPRT (sequential probability ratio test) 방법은 화자 확인에서 성능 개선 및 계산량 감소를 위한 통계적인 시퀀스 결정이라고 할 수 있다[5,6]. 먼저 SPRT에서 고려되어야 하는 것은 결정과정을 수행할 때, 입력신호가 확률밀도함수로부터 상호 독립적이라고 가정한다. 무엇보다도 연속적인 특징벡터 사이에는 많은 상관성이 존재하지만, 그러나 SPRT를 사용하는 화자 확인방법에서는 서로

독립적으로 다루어진다.

입력 신호로부터 얻어진 특징벡터를  $\{o_t, t=1, 2, \dots, F, \dots, T\}$ 라 하고,  $P_S(o_t)$ 와  $P_B(o_t)$ 를 각각 연속밀도 HMM에서 계산된 화자와 사칭자의 확률밀도함수라고 하자. 관측벡터  $o_t$ 가 주어지고 그리고  $o_t$ 는 상호 독립이라고 가정할 때, SPRT의 유사도 비는 다음과 같이 정의할 수 있다.

$$l_r(o_1, \dots, o_F) = \prod_{t=1}^F \frac{P_S(o_t)}{P_B(o_t)} \quad (7)$$

이런 유사도 비를 사용할 때, 결정은 다음과 같이 판별할 수 있다.

$$\begin{aligned} l_r(o_1, \dots, o_F) > Th_a & \quad \text{인증 (accept) : 화자} \\ l_r(o_1, \dots, o_F) < Th_r & \quad \text{거절 (reject) : 사칭자} \end{aligned} \quad (8)$$

여기서  $Th_a$ 와  $Th_r$ 는 각각 인증과 거절을 나타내는 판별값이고,  $Th_a > Th_r$  관계를 갖는다. 만약 유사도 비가  $Th_a$ 와  $Th_r$  사이에 존재하면 다음 프레임  $F+1$ 의 데이터에 대해서 결정을 반복 수행한다.

#### IV. 제안한 시작 프레임의 샘플 시프트를 통한 새로운 데이터 생성

본 논문에서 제안하는 방법은 테스트 길이에 상관없이 SPRT를 수행할 수 있도록 새로운 데이터를 생성하는 방법이다. 그림 1의 상측은 프레임 생성을 위한 50% 프레임 시프트를 시킨 일반적인 경우이다. 새로운 데이터 생성은 그림 1의 상측과 같은 일반적인 과정을 수행한 후 결정 로직에서 인증 판별이  $Th_r < l_r(o_1, \dots, o_T) < Th_a$ 인 경우에 그림 1의 하측과 같이 주어진 데이터에 대해서 시작

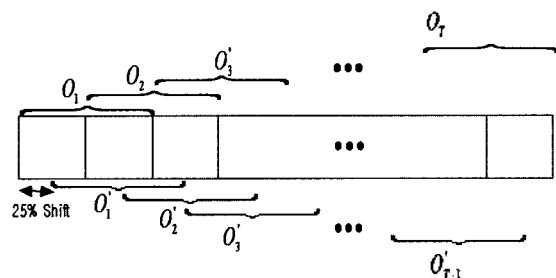


그림 1. 일반적인 프레임 생성과 제안한 25% 시작 프레임 시프트를 통한 새로운 데이터 생성 방법  
Fig. 1. The conventional frame generation and proposed method using 25% shift of the start frame.

프레임의 샘플 시프트를 통해서 새로운 데이터를 생성할 수 있다. 이때, 샘플 시프트는 기존 프레임 시프트의 값을 이용해서 구할 수 있다.

여기서 생성되는 데이터는 첫 번째 프레임에서 25% 프레임 시프트되었기 때문에  $T-1$ 개의 새로운 데이터가 생성된다. 그림 1의 하측과 같은 방법을 적용할 경우 두 배의 새로운 테스트 데이터를 생성할 수 있다.

이때 처음 프레임 시프트의 값은 식 (9)와 같이 생성하고자 하는 데이터의 반복 횟수와 관련이 있다.

$$S_{shift} = \frac{N}{2 \cdot I} \quad (9)$$

여기서  $S_{shift}$ 는 시작 프레임에서 샘플 시프트를 위한 샘플 수,  $N$ 은 한 프레임의 전체 샘플 수,  $I$ 는 반복 횟수이다. 만약  $I=1$ 이면 그림 1의 상측과 같이 일반적인 프레임 생성과정과 같고,  $I=2$ 이면 25%의 시작 샘플을 시프트한 경우이다. 분모항의 2는 분자항의 전체 프레임 길이에 대한 기존의 50%의 시프트를 고려한 것이다. 만약 데이터의 생성 반복 횟수가  $I=3$ , 즉 각 샘플 시프트를 17%와 34%를 사용한다면, 데이터 생성과정은 그림 2와 같이 진행된다.

제안한 방법에서는 위와 같은 과정으로 결정로직을 위한 테스트 데이터를 생성할 수 있으며, 첫 번째 결정로직을 위한 유사도 비는 그림 1의 상측에서 얻어진 데이터를 이용해서 유사도 비를 얻을 수 있다.

$$l_r(o_1, \dots, o_T) = \prod_{t=1}^T \frac{P_S(o_t)}{P_B(o_t)} \quad (10)$$

여기서 유사도 비가 식 (8)에 속할 경우에 인증판별을 결정할 수 있지만,  $Th_r < l_r(o_1, \dots, o_T) < Th_a$ 인 경우에는 그림 1의 하측 또는 그림 2와 같은 과정을 통해서 다음과 같이 판별을 계속할 수 있다.

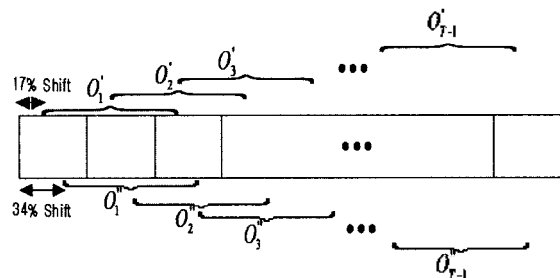


그림 2. 제안한 17%와 34% 시작 프레임 시프트를 통한 새로운 데이터 생성 방법  
Fig. 2. The proposed method using 17% and 34% shift of the start frame.

$$l_r(o_1', \dots, o_{T-1}') = \prod_{i=1}^{T-1} \frac{P_S(o_i')}{P_B(o_i')} \quad (11)$$

$$l_r(o_1'', \dots, o_{T-1}'') = \prod_{i=1}^{T-1} \frac{p_S(o_i'')}{p_B(o_i'')} \quad (12)$$

식 (11)은 프레임 시프트를 통한 두 번째 유사도 비를 구한 것이다. 여기서 고려해야 하는 것은 식 (11)과 (12)를 통해서 얻어진 유사도 비가 여전히  $Th_r < l_r(o_1', \dots, o_{T-1}') < Th_a$ 와  $Th_r < l_r(o_1'', \dots, o_{T-1}'') < Th_a$ 인 경우이다. 이런 경우에는 임의의 판별값  $Th_l$ 으로 결정판별을 내려야 한다. 따라서 최종 결정판별 과정은 식 (13) 또는 (14)와 같이 결론지을 수 있다.

$$-\frac{1}{T} \left( l_r(o_1', \dots, o_T') + \sum_{i=2}^T l_r(o_1', \dots, o_{i-1}') \right) \geq Th_l$$

인증(accept) : 화자 (13)

$$-\frac{1}{T} \left( l_r(o_1', \dots, o_T') + \sum_{i=2}^T l_r(o_1', \dots, o_{i-1}') \right) < Th_l$$

거절(reject) : 사칭자 (14)

이러한 과정으로 진행하면 그림 3과 같이 일반적인 화자 확인 방법과 기존의 SPRT보다 계산량은 증가하지만, 많은 데이터에 대해서 확인을 수행할 수 있기 때문에 성능 오류는 줄일 수 있다. 그림 3은 화자확인에서 제안한 방법을 이용했을 때, 새로운 테스트 데이터를 이용해서 오류를 줄이는 과정을 나타낸 것이다. 여기서  $o$ 와  $*$ 는 각각 화자와 사칭자의 유사도 비를 나타낸 것이다.  $i=2$ 는 시작 프레임 이서 25% 샘플 시프트를 시킨 경우이고  $i=3$ 은 시작 프레임에서 17%와 34% 샘플 시프트를 이용한 경우이다. 특히 마지막 확인 판별에서 기존의 방법은 첫 번째 생성 데이터에 대해서만 수행을 할 수 있지만, 제안한 방법에서는 식

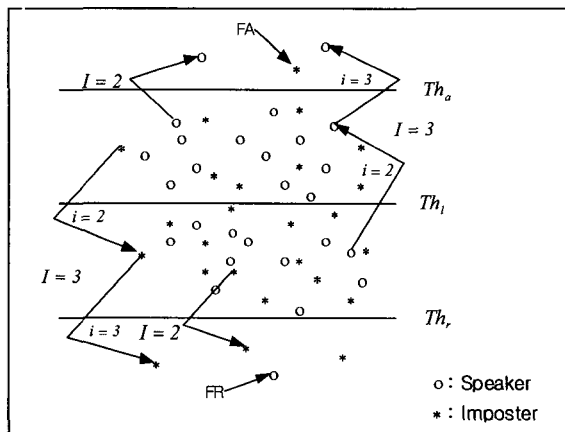


그림 3. 새로운 테스트 데이터에 의한 화자 확인 방법  
Fig. 3. Speaker verification method by a new test data.

(13), (14)와 같이 기존의 생성된 데이터와 제안된 방법에 의해서 생성된 데이터의 합에 의해서 임의의 판별값  $Th_l$ 으로 결정할 수 있기 때문에 좋은 성능 결과를 얻을 수 있다.

## V. 실험 결과

실험을 위해서 사용된 데이터는 대학원 실험실 환경에서 수집하였으며, 한국어 문장 종속 연속음 화자인식을 위해서 "얼러라 참깨"를 사용하였다. 수집된 데이터는 1주 간격의 시간차를 가지고 있으며, 3주에 걸쳐서 수집하였다. 매주 1회 발성에서는 각 5번 발성을 하였으며, 개인별 전체 발성된 데이터 수는 15개이다. 수집된 데이터의 화자 인원 수는 200명으로 남/여 각각 100명이다. 따라서 수집된 전체 데이터는 3,000개이며 샘플링 주파수는 16 kHz이고 분해능은 16 bit이다. 학습을 위해서 사용한 데이터는 처음 2주간 10번 발성한 데이터를 사용하였고 테스트 데이터는 마지막 주의 5번 발성한 데이터를 사용하였다. 따라서 오거절률 (FRR)을 위한 참여 데이터는 1,000개이고, 오인식률 (FAR)을 위해서 사용된 데이터는 199,000개가 된다.

먼저 실험에서 사용된 음성의 프레임 길이는 256 샘플을 한 프레임으로 간주하고 그리고 프레임간 50% 중첩을 적용했다. 특징벡터는 MFCC를 사용했으며, 이때 FFT는 256 point를 사용하고 그리고 필터뱅크는 24개를 사용하였다. 전체 특징벡터는 MFCC를 통해서 얻어진 12차 캡스트럼 계수, 델타 에너지 그리고 12차 델타 캡스트럼의 25차원으로 구성되었다. 또한 MFCC를 통해서 얻어진 캡스트럼은 채널의 영향을 없애기 위해서 캡스트럼 평균 차감법 (CMS: cepstral mean subtraction)과 고차 성분을 강조하기 위해서 캡스트럼 가중치를 적용하였다.

실험에서 연속밀도 HMM의 학습 파라미터를 추정하기 위해서 사용한 방법은 Baum Welch 재추정 알고리즘을 사용하였으며 그리고 각 상태별 혼합 성분은 5개를 사용하였다[7]. 또한 사칭자의 확률밀도함수  $P_B(o_i)$ 를 위해서 UBM (universal background model)을 사용하였다[8]. UBM은 남/여의 성 (gender)을 구분하였으며, 각 상태별 혼합 성분의 개수는 50개를 사용하였다. SPRT를 위한 결정 판별의 프레임 수는 전체 데이터에서 80%에 해당하는 부분을 기준으로 실험을 시작하였다.

실험을 통한 제안한 방법의 우수성을 확인하기 위해서 비교한 방법은 일반적인 화자확인, SPRT 방법 그리고 논문에서 제안한 방법을 FRR과 FAR의 오류를 표를 통해서 나타내었다. 또한 정확한 성능 확인을 위해서 유사도

표 1. 상태 수 5개에서의 일반적인 HMM, SPRT, 그리고 제안한 방법의 성능 비교

Table 1. Performance comparison between the conventional HMM, SPRT, and proposed method for state 5.

States : 5	Speaker Verification (%)		
	FRR	FAR	EER
Conventional HMM	7.51	1.217	3.43
SPRT	8.2	1.154	3.65
Frame Shift ( $l=2$ )	7.2	1.188	3.13
Frame Shift ( $l=3$ )	7.06	1.102	2.93

표 2. 상태 수 6개에서의 일반적인 HMM, SPRT 그리고 제안한 방법의 성능 비교

Table 2. Performance comparison between the conventional HMM, SPRT, and proposed method for state 6.

States : 6	Speaker Verification (%)		
	FRR	FAR	EER
Conventional HMM	7.3	1.100	3.27
SPRT	7.7	1.124	3.35
PCA+Frame Shift ( $l=2$ )	6.8	1.091	2.76
PCA+Frame Shift ( $l=3$ )	6.5	1.060	2.61

비의 결과를 함께 표에 기술하였다.

표 1은 연속밀도 HMM에서 상태 수 5개를 사용했을 때의 실험 결과를 나타낸 것이다. 실험 결과에서 제안한 방법은  $l=2$ , 즉 25% 시작 프레임 시프트를 시킨 경우 유사도 바에서 일반적인 방법보다 0.3% 정도 향상된 결과를 보였다. 그리고  $l=3$ 인 경우 일반적인 화자확인 방법보다 약 0.4% 이상 좋은 결과를 얻을 수 있었다. 특히 제안한 방법에서  $l=3$ 인 경우 그림 1과 같이  $l=2$ 인 경우만 사용한 경우보다 약 0.15% 좋은 결과를 보였다. 그러나 기존의 SPRT 방법은 일반적인 확인 방법보다 약 0.2% 정도 오류가 증가하였다. 이것은 발성 문장이 짧은 경우에서 나타난 오류라고 할 수 있다.

표 2는 상태 수가 증가된 6개일 때의 실험결과이다. 실험 결과에서 알 수 있듯이 주성분 분석을 적용했을 때, 제안한 방법이 기존의 방법보다 약 0.5%와 0.6%의 오류를 줄일 수 있었다. 일반적인 SPRT는 역시 기존의 방법보다 약 0.1% 정도의 오류가 증가하였다.

표 3은 상태 수 7개로 증가해서 실험한 결과를 나타낸 것이다. 제안한 방법에서의 실험 결과는 일반적인 HMM 보다 주성분 분석을 이용했을 때, 각각 0.7%와 0.8%의 오류가 줄었으며, 단지 시작 프레임 시프트만을 적용했을 때도 0.3% 이상 오류를 줄일 수 있었다. 그러나 SPRT 방법에서는 일반적인 HMM 방법보다 오류가 약간 증가

표 3. 상태 수 7개에서의 일반적인 HMM, SPRT, 그리고 제안한 방법의 성능 비교

Table 3. Performance comparison between the conventional HMM, SPRT, and proposed method for state 7.

States : 7	Speaker Verification (%)		
	FRR	FAR	EER
Conventional HMM	7.1	1.273	3.20
SPRT	7.6	1.124	3.29
Frame Shift ( $l=2$ )	7.0	1.136	2.87
Frame Shift ( $l=3$ )	6.8	1.06	2.75
PCA+Frame Shift ( $l=2$ )	6.6	0.985	2.53
PCA+Frame Shift ( $l=3$ )	6.5	0.915	2.41

표 4. 계산량 비교

Table 4. Comparison results of the computational complexity.

Methods	Computation Comparison (%)
Conventional HMM	-
SPRT	-13.2
Frame Shift ( $l=2$ )	+7.0
Frame Shift ( $l=3$ )	+9.5
PCA+Frame Shift ( $l=2$ )	+6.7
PCA+Frame Shift ( $l=3$ )	+8.6

했다. 이것은 짧은 발성에서 SPRT 방법을 적용하는 것은 적절하지 못하기 때문에 나타난 문제점이라고 할 수 있다. 그리고 계산량은 표 4에서 일반적인 HMM을 기준으로 나타냈다. 일반적인 HMM을 기준으로 했을 때, 기존의 SPRT 방법은 약 13.2% 정도 줄었다. 그리고 제안한 방법에서  $l=2$ 일 때 평균 6.8% 정도 증가했으며,  $l=3$ 일 때는 9% 정도 증가했다. 계산량 비교에서 제안한 방법이 비록 계산량은 약간 증가했지만, 위의 실험 결과를 통해서 알 수 있듯이 비교적 큰 성능향상을 얻을 수 있었다.

따라서 실험 결과로부터 제안한 시작 프레임 시프트를 통한 새로운 테스트 데이터 생성이 화자확인 오류를 줄이는데 효과적인 알고리즘임을 알 수 있다. 그리고 새로운 테스트 데이터 생성에 있어서 주성분 분석이 신호의 상관성을 효율적으로 제거하고 있음을 나타내고 있다.

## VI. 결론

본 논문에서 제안하는 방법은 화자확인에서 SPRT (sequential probability ratio test)를 위한 시작 프레임의

샘플 시프트를 이용해서 새로운 테스트 데이터를 생성하는 방법이다. SPRT는 테스트 계산량을 줄일 수 있는 효과적인 알고리즘이다. 그러나 테스트의 결정과정에서 SPRT 방법은 입력신호가 확률밀도함수로부터 비상관적이고 균일하게 분포되어 있다는 가정하에 수행할 수 있으며, 또한 발생길이가 짧은 데이터에는 적용하기에 적절하지 못하다. 제안한 방법은 시작 프레임의 샘플 시프트를 통한 새로운 테스트 데이터를 생성하는 방법이기 때문에 테스트 데이터의 길이에 상관없이 SPRT를 수행할 수 있다. 또한 SPRT 방법에서 고려해야 하는 데이터의 상관성은 주성분 분석을 이용함으로써 효과적으로 제거하였다. 실험결과 제안한 방법은 기존의 방법보다 샘플 시프트를 위한 데이터의 계산량은 약간 증가하였지만, 평균 0.7% 이상 좋은 성능결과를 보였다.

**감사의 글**

This work was supported (in part) by Biometrics Engineering Research Center, (KOSEF).

**참고 문헌**

1. Y. Ariki, S. Tagashira, and M. Nishijima, "Speaker recognition and speaker normalization by projection to speaker subspace," *ICASSP 96*, Atlanta, Georgia, USA, 319-322, 1996.
2. C.-W. Seo, K.-Y. Lee, and J.-H. Lee, "GMM based on local PCA for speaker identification," *IEEE Electronic Letters*, 37 (24), 1486-1488, Nov., 2001.

3. L. Liu and J. He, "On the use of orthogonal GMM in speaker recognition," *Proc. ICASSP*, 845-849, 1999.
4. C. Croux and G. Haesbroeck, "Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence function and efficiencies," *Biometrika*, 87 (3), 603-618, 2000.
5. H. Noda, K. Harada, and E. Kawaguchi, "A context-dependent sequential decision for speaker verification," *IEICE Trans. INF. & syst.*, E82-D (10) Oct., 1999.
6. M. A. Lund and C. C. Lee, "A robust sequential test for text-independent speaker verification," *J. Acoust. Soc. Am.*, 99 (1), 609-621, 1996.
7. L. E. Baum, C. S. T. Petrie, and N. Weiss, "Maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, 41, 164-171, Jan, 1970.
8. T. Matsui and S. Furui, "Likelihood normalization for speaker verification using a phoneme and speaker-independent model," *Speech Communication*, 17, 109-116, August 1995.

**저자 약력**

● 서 창 우 (Chang-Woo Seo)



1996년 2월: 창원대학교 전자공학과 (공학사)  
 1998년 2월: 창원대학교 전기전자제어공학부 (석사)  
 1999년 3월~현재: 숭실대학교 정보통신전자공학 박사  
 수료  
 ※ 주관심분야: 화자인식, 음성인식, 신경망

● 이 기 용 (Ki-Yong Lee)

1991년 2월: 서울대학교 전자공학과 (박사)  
 1991년 9월~1997년 8월: 국립창원대학교 조교수  
 1994년 8월~1995년 6월: 일본 와세다대학 초빙연구원  
 1996년 1월~3월: 영국 에든버러대학 박사후과정  
 1997년 6월~ 8월: 독일문헌공대 초빙연구원  
 1997년 9월~ 현재: 숭실대학교 정보통신전자공학부 부교수  
 ※ 주관심분야: 음성신호 향상, 화자인식, 음성인식