

문장종속 화자확인 시스템을 위한 개선된 군집화 알고리즘에 관한 연구

A Study on Modified Clustering Algorithm for Text-Dependent Speaker Verification System

정 희 석*, 강 철 호*
(Hee-Suk Jeong*, Chul-Ho Kang*)

*광운대학교 전자통신학과

(접수일자: 2004년 8월 2일; 수정일자: 2004년 9월 8일; 채택일자: 2004년 10월 1일)

본 연구에서는 집단화 오차를 최소화 하기위해 개선된 LBG 알고리즘을 제안한다. 기존의 LBG 알고리즘은 화자확인 시스템에 적용시 소량의 학습 데이터의 분포가 가지는 특수성으로부터 기인하는 문제점들이 발생한다. 즉, 개인별 특성을 무시하고 항상 일정한 크기의 코드북을 생성해야 하는데서 기인하는 군집화 오류와 분할할 (Splitting) 방향을 잘못 선택하면서 발생하는 집단화의 오류가 전체 화자 인식을 저하의 원인이 된다. 따라서, 본 연구에서는 개인별로 최적의 크기를 가지는 가변길이 코드북 생성 기법과 중심값으로부터 최외곽의 멤버 벡터 인덱스를 찾고 다시 최외곽 멤버 벡터에서 가장 먼 멤버 벡터 인덱스를 찾음으로써 분할할 방향을 인위적으로 지정해 주는 개선된 군집화 알고리즘을 제안한다. 실험 결과, 제안된 방식을 적용한 화자확인 시스템이 기존의 LBG 알고리즘을 사용한 시스템보다 오거부율 (FR)은 3.165%, 오수락율 (FA)는 0.06%씩 각각 향상 되었다.

핵심용어: LBG 알고리즘, 화자 확인, 집단화, HMM

투고분야: 음성처리 분야 (2.5)

In this paper, we propose modified LBG algorithm to minimize quantization errors. When we apply conventional LBG algorithm for speaker verification system, problems that result from small amount of training data can be generated. That is, quantization error comes from fixed-sized codebook without any consideration for speaker characteristics and splitting vector in the wrong direction worsen performance of speaker verification system. So, we propose modified clustering method that has variable sized codebook according to speaker characteristics and makes right splitting direction by finding the farthest member away from mean and then find another member from the member. Simulation results show effectiveness of the proposed algorithm.

Keywords: LBG algorithm, Speaker verification, Clustering, HMM

ASK subject classification: Speech signal processing (2.5)

I. 서론

본 연구에서는 소규모 음성 데이터로 화자 모델을 생성해야만 하는 Mixed model VQ/HMM 기반 화자확인 시스템의 성능 향상을 위한 가변길이를 갖는 수정된 LBG 알고리즘 제안한다. 기존의 K-means 알고리즘과

LBG (Linde-Buzo-Gray) 알고리즘은 학습 과정에서 항상 일정한 크기의 코드북을 생성하므로 사용자 편이성을 고려하여 학습 데이터가 충분치 못한 화자 확인 시스템에서는 많은 집단화오차를 발생시킬 수 있다[1]. 또한, 학습 데이터 분포가 특수한 형태를 나타낼 때, 분할 벡터의 방향을 잘못 선택하게 되면 어쩔 수 없이 양자화에러 및 수렴 속도는 크게 증가하게 된다. 이는 전체 화자확인 시스템 성능 열화에 가장 큰 영향을 미치는 요소이다. 본 연구에서는 개인마다 최적의 크기를 가지는 가

변질이 코드북 생성 기법을 제안하고 또한 학습 데이터의 중심값 으로부터 최외곽의 멤버 벡터 인덱스 (index) 를 찾고 다시 최외곽 멤버 벡터에서 가장 거리가 먼 멤버 벡터 인덱스를 찾은 후 분할 벡터 (perturbation vector), ϵ 의 방향을 제시해 줌으로써 양자화 오차를 극복하고 동시에 수렴 속도의 향상을 꾀하였다.

본 논문의 구성은 전체 5장으로 구성되어 있다. 2장에서는 기존의 집단화 알고리즘에 대하여 기술하고 3장은 본 논문에서 코드북 생성을 위해 제안한 개선된 LBG 알고리즘을 제시하였고, 4장에서는 모의실험 및 성능을 평가했으며 마지막 5장에서 결론을 이끌어 냈다.

II. 기존의 집단화 알고리즘

일반적인 클러스터링 알고리즘으로는 K-means 알고리즘과 LBG 알고리즘이 대표적이는데 먼저 Lloyd 알고리즘으로 잘 알려진 K-means 알고리즘의 기본이론은 무한히 많은 수의 입력벡터를 미리 정해진 K개의 중심값으로 사상시키는 집단화 알고리즘으로써 임의의 입력벡터에 대한 벡터 양자화 에러를 최소화 하기 위해 반복적으로 클러스터의 중심값을 갱신하는 알고리즘이다.

K-means 알고리즘의 경우 유일한 코드워드로 집단화 되지 못하고 초기치의 설정에 따라 국부최소치를 갖기도 하는 등 단점을 갖는다. 따라서 최적의 벡터 양자화를 위한 개선된 집단화 알고리즘들이 발표되었고 Linde, Buzo, Gray에 의한 LBG 알고리즘이 대표적인 예이다 [2-4].

LBG 알고리즘은 반복적으로 클러스터를 분할하는 알고리즘으로써 K-means 알고리즘에서의 초기치 설정 문제를 해결하기 위해 모든 입력벡터를 하나의 클러스터로 집단화하여 초기 중심값을 설정한 후 매 반복횟수마다 현재의 클러스터들의 중심값을 이동시켜 분할한다. 결국 분할과정이 m번 진행되면 2^m 개의 클러스터를 만들게 되고 미리 정해진 수 K개의 안정된 분할을 이를 때까지 반복된다.

2.1. 문맥 중속 화자 확인 시스템을 위한 기존 LBG 알고리즘의 문제점

2.1.1. 일정 크기를 찾아야만 하는 군집화 기법

다량의 음성 데이터를 이용하여 군집화 하는 음성 인

식 시스템에서는 기존의 LBG알고리즘은 초기치에 민감한 K-means알고리즘의 문제점을 해결하여 양자화 오류를 최소화 하는 이점이 가진다. 그러나, 음성 인식 시스템에 비해 극소량의 학습 데이터를 갖는 화자확인 시스템에서 화자모델 생성시 학습 과정에서 수집된 음성데이터는 화자의 발성에 따라 그 크기가 다양하다. 이러한 다양한 크기를 가지는 특징 벡터들을 K개나 2^m 의 일정한 크기의 클러스터로 표현한다면 화자마다 상이한 특성을 제대로 표현하기 어렵게 된다. 즉, 화자확인 시스템의 구성상 공통코드북에 의한 월드모델과의 비교과정에서 공통코드북의 크기에 맞추어 화자의 개인코드북을 분할하여 형성한다면 매우 적은 수의 관측데이터에 대한 무의미한 클러스터를 형성하거나 공통코드북의 한 클러스터 영역에서 매우 세분화된 개인 코드북의 클러스터들을 형성하게 되어 화자간 변별력을 저하시키고 화자내 변이에 따른 적응성을 저하시킨다. 또한, 공통코드북에 대한 상대적인 화자의 관측데이터의 분포가 상이한 점을 고려할 때 일정한 수의 분할은 이러한 특성을 오히려 완화시킴으로써 화자간의 변별력을 저하시키는 원인이 되기도 한다.

2.2.2. 분할 벡터의 방향 오류에 의한 군집화 오류

화자확인 시스템에서 화자모델 생성시 학습 과정에서 수집된 음성데이터의 특징 벡터가 그림 1과 같이 분포될 때, 분할 벡터, ϵ 이 그림 1과 같이 설정될 경우, 기존의 LBG알고리즘에 의한 집단화 과정을 거친 클러스터는 특징 벡터를 잘 표현할 수 없게 된다. 이렇게 잘못된 군집화는 화자 확인 시스템의 인식율에 큰 영향을 미칠뿐 아니라 수렴 속도에도 많은 영향을 미치게 된다.

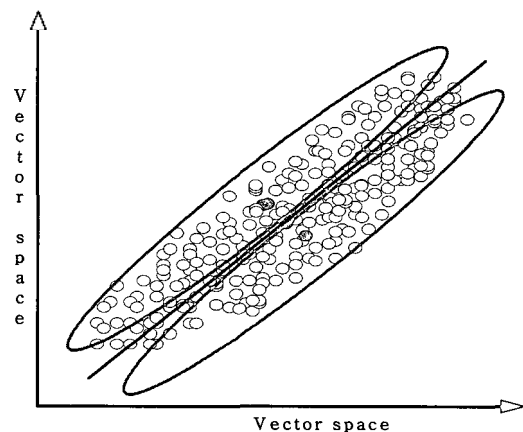


그림 1. 기존 LBG 알고리즘의 코드북 생성
Fig. 1. Codebook generation of conventional LBG algorithm.

III. 제안한 LBG 알고리즘

본 연구에서는 소량의 학습데이터를 갖는 문장중속 화자확인 시스템을 위한 기존 LBG알고리즘의 문제점을 해결하기 위해 화자별 가변크기의 클러스터를 갖는 군집화 기법을 제안하고, 분할 벡터의 방향 오류에서 오는 군집화 오류를 최소화하기 위한 방법을 제안한다.

3.1. 제안한 개선된 LBG 알고리즘

본 연구에서는 LBG 알고리즘을 이용하여 특정단어모델에 대한 일반적인 화자들의 특징벡터를 표현하는 공통코드북을 생성하였다. 또한 거리측정방식에서 임의의 발성음성이 이러한 공통 코드북과 비교대상으로 지목되는 개인화자의 코드북 벡터들과의 형평을 유지하기 위해 개인화자의 코드북을 생성하는 방식을 변형하였다. 즉, 공통화자의 각 클러스터들의 평균분산을 경계치로 하여 개인화자의 코드북을 생성한다. 따라서 기존의 K-means 나 LBG 등의 클러스터링 알고리즘이 일정한 수의 클러스터를 생성하는 반면 본 연구에서 수정된 클러스터링 알고리즘의 경우 LBG 알고리즘과 유사한 과정을 통해 집단화하는 분할 (splitting) 알고리즘이라 할 수 있으나 분할 조건을 반복횟수마다 무조건적으로 클러스터의 수를 두배수로 증가시키는 분할기법과는 달리 분산이 최대인 클러스터만을 재분할하고 각 클러스터들의 분산값이 공통코드북의 평균분산 미만이 될 때까지 분할 과정과 클러스터링을 수행하게 함으로써 일반적인 공통코드북의 수보다 작은 수의 클러스터를 생성하게 된다. 이는 또한 개인코드북의 분할이 공통코드북의 조건과 동일시되므로써 화자확인인 인식률에 크게 영향을 미치는 코드

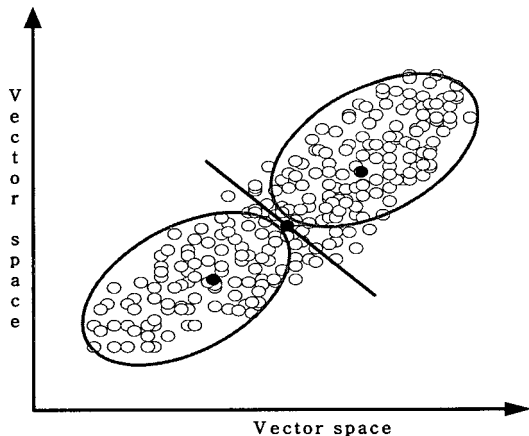


그림 2. 제안한 LBG 알고리즘의 코드북 생성
Fig. 2. Codebook generation of proposed LBG algorithm

북의 변별력을 향상시키게 된다. 즉, 전체 공통화자들에 의해 형성된 공통코드북의 크기를 128개의 분할된 클러스터로 형성하였다면 이들 클러스터들의 평균 분산을 고려하여 평균 320개 정도의 프레임으로 분석되는 개인화자의 코드북을 가변적으로 생성하게 된다. 이 경우 개인화자의 코드북은 화자의 발성음에 따라 약 36~48 정도의 크기를 갖는 코드북이 형성되고 이는 공통코드북의 각 클러스터와 비슷한 정도의 분산을 갖도록 구성되므로 코드북간의 변별력을 향상시키게 된다.

또한, 일반적인 LBG와 같은 분할 (split) 알고리즘의 경우 현재의 중심값을 일정한 방향으로 이동시켜 중심점을 재추정하는 알고리즘인데 반해 제안하는 가변길이 코드북생성 알고리즘의 경우 분할하고자하는 클러스터 (집단)의 중심점을 다차원 클러스터의 분산환경을 고려하여 클러스터 멤버의 구성이 양분될 수 있도록 중심값의 분할방향을 가변적으로 이동시킨다. 그림 2와 같이 학습데이터의 중심값 으로부터 최외곽의 멤버 벡터 인덱스를 찾고 다시 최외곽 멤버 벡터에서 가장 먼 멤버 벡터 인덱스를 찾아 분할 벡터의 방향을 결정해 줌으로써 분할 벡터의 방향 오류에서 가져올 수 있는 집단화 오류를 최소화 하는 방식을 제안한다. 이는 인식률에 크게 영향을 미치는 코드북의 변별력을 향상시키게 될 뿐 아니라 군집화 수렴속도를 줄일 수 있다.

다음은 본 연구에서 개인 코드북을 생성하기 위한 수정된 클러스터링 알고리즘의 수행 과정을 살펴본 것이다.

step 1) Initialization : 모든 입력벡터에 대한 하나의 중심값을 설정한다.

$$c_1^i = \frac{\sum_{all} x^{(i)}}{N_T} \quad (5)$$

where, N_T : 벡터의 수

step 2) Searching : 현재의 모든 클러스터 중 최대 분산을 갖는 클러스터를 검출한다.

$$\bar{k} = \arg \max_k \Sigma(S_k) \quad (6)$$

$\Sigma(S_k)$: k번째 클러스터의 분산 $k=1,2,\dots,K$

S_k : k번째클러스터

step 3) Splitting : 최대 분산을 갖는 클러스터의 중심

값을 이동시켜 둘로 분할한다. 이때 분할 방향은 다음과 같이 분산을 고려하여 분할하고자하는 클러스터의 멤버를 양분하도록 한다.

$$n = \arg \max_{all j} d(c_i, x_j) \quad (7)$$

$$m = \arg \max_{all j} d(x_n, x_j) \quad (8)$$

$$c_k^{(p)+} = \alpha * c_k^{(p)} + (1-\alpha) * x_n \quad (9)$$

$$c_k^{(p)-} = \alpha * c_k^{(p)} + (1-\alpha) * x_m \quad (10)$$

where, $0.8 \leq \alpha \leq 0.9$

c_k 는 k 번째 클러스터의 mean값

step 4) Clustering : 모든 입력벡터에 대한 분할된 각 클러스터와의 유클리드 거리를 측정하여 가장 작은 거리를 갖는 클러스터의 멤버벡터로 집단화한다.

$$k^* = \arg \min_k d(x^{(p)}, c_k^{(p)}) \quad (11)$$

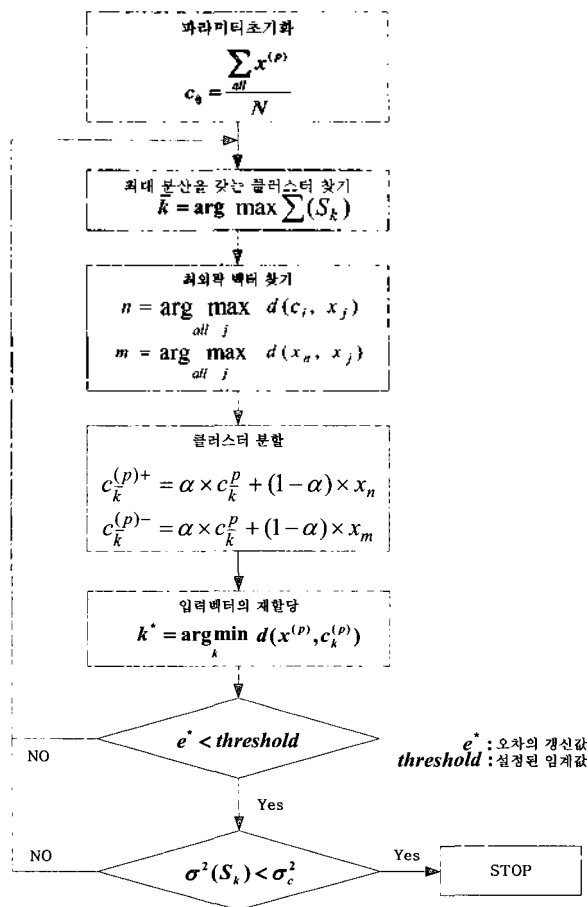


그림 3. 제안한 LBG 알고리즘 흐름도
Fig. 3. Flowchart of the Proposed LBG Algorithm.

step 5) Centroid update : 각 클러스터의 멤버벡터를 통해 클러스터의 중심점을 갱신한다.

$$c_{k^*} = \frac{1}{N_{k^*}} \sum_{x \in S_{k^*}} x \quad (12)$$

where, N_{k^*} : S_{k^*} 에 소속된 멤버의 수

step 6) Termination 1 : 오차의 갱신값이 설정된 임계값 이하이면 step 7 을 수행하고 그렇지 않으면 step 4 로 되돌아가간다.

step 7) Termination 2 : 각 클러스터들의 분산이 임계값 (공통코드북의 평균분산)이하이면 작업을 종료하고, 그렇지 않으면 step 2 로 되돌아가 반복하여 수행한다.

그림 3은 제안한 LBG 알고리즘의 흐름도를 나타낸 것이다.

IV. 모의 실험 및 결과

본 논문에 적용된 화자확인 시스템에서는 우도비 정규화 기법으로 월드모델 방식을 사용하였다. 또 화자 개인의 특징 패턴들 간의 변별력을 높이기 위해 개인의 코드북과 월드모델에 적용되는 코드북을 분리하여 설계하였고, 인식실험에서 이를 통합하였다.

공통 코드북과 월드모델은 20~30대의 남녀 각 100명의 화자로부터 10번씩 발음한 데이터 베이스를 이용하여 사전에 작성하였다. 공통 코드북은 128개의 코드워드로 구성된 코드북 벡터를 가지고 있으며, 학습에서는 공통 코드북보다 적은 개인 코드북을 생성하게 된다. HMM 과정에서는 이산 HMM을 이용하여 HMM 상태수 (State number)는 5개로 하고, left-to-right 모델을 적용하였다. Baum-Welch re-estimation의 수렴조건은 0.005를 임계값으로 설정하고 개인마다 상이한 사전 개인별 문턱치를 구하였다[5].

실험에 사용된 음성 데이터는 11,025kHz 16bit 로 샘플링 되었고, 음성 분석 구간은 한 프레임을 20msec로 하고 1/3 중첩시켜 해밍 윈도우 (hamming window)을 취한 후 1차의 에너지 켈스트럼과 19차의 LPC 켈스트럼을 특징 파라미터로 구성하였다.

모의 실험용 데이터 베이스는 20~30대 남녀 화자 각 30명씩에 의해서 일주일에 걸쳐 3회 10번씩 발음한 데이

표 1. 기존 LBG 알고리즘의 시뮬레이션 결과

Table 1. Simulation results of conventional LBG algorithm.

Conventional LBG algorithm						
male 30	after training		after 3days		after 7days	
	FR	FA	FR	FA	FR	FA
total	24/300	36/8700	28/600	33/8700	23/600	38/8700
%	8.0	0.41	9.33	0.38	7.67	0.44
female 30	after training		after 3days		after 7days	
	FR	FA	FR	FA	FR	FA
total	29/300	26/8700	28/600	34/8700	30/600	21/8700
%	9.67	0.3	9.33	0.39	10.0	0.24
total error rate	male FR : 8.33%, FA : 0.41% female FR : 9.67%, FA : 0.31%					

표 2. 제안한 LBG 알고리즘의 시뮬레이션 결과

Table 2. Simulation results of proposed LBG algorithm.

제안한 LBG algorithm						
male 30	after training		after 3days		after 7days	
	FR	FA	FR	FA	FR	FA
total	14/300	30/8700	21/600	28/8700	18/600	21/8700
%	4.67	0.34	7.0	0.32	6.0	0.24
female 30	after training		after 3days		after 7days	
	FR	FA	FR	FA	FR	FA
total	15/300	27/8700	19/600	31/8700	18/600	19/8700
%	5.0	0.31	6.33	0.36	6.0	0.22
total error rate	male FR : 5.89%, FA : 0.3% female FR : 5.78%, FA : 0.3%					

터를 이용하여 실험하였다. 학습 단계는 등록 과정의 편의성을 고려하여 3회 발음한 음성을 이용하여 학습을 진행하였다. 화자확인 시스템의 인식단계에서는 입력 음성의 특징 파라미터를 추출하여 개인 코드북과 공통 코드북을 합성한 코드북을 통해 VQ 과정을 수행하고 HMM 모델을 통해서 개인모델과 월드모델의 비터비 계산(Viterbi scoring)으로 얻어진 우도비는 결정 규칙에 따라서 인증여부를 판단하게 된다.

그림 4와 5는 남녀 화자에 대한 코드북 수렴 속도를 비교한 것으로 제안한 LBG 알고리즘이 기존의 LBG 알고리즘보다 평균 0.021초 빨리 수렴한다. 이는 분할 벡터의 방향 오류로 인한 수렴 속도 지연을 제안한 방식이

항상시켜줌을 증명한다.

표 1은 기존의 LBG 알고리즘의 성능 결과이고, 표 2는 수정된 LBG 알고리즘의 성능 결과이다. 본 실험에서는 3회 발성한 개인 학습 데이터에 대한 기존의 LBG 알고리즘과 제안한 LBG 알고리즘에 의한 클러스터링 결과가 인식율에 미치는 영향을 비교, 분석하였다.

표1, 표2는 FR과 FA의 실험결과를 나타낸 것으로 표1은 기존의 LBG알고리즘을 사용한 경우 남여 평균 오거부율 (FRR)은 9.0%, 오수락율 (FAR)는 0.36%를 각각 나타냈으며 표2는 제안한 LBG알고리즘을 사용한 경우로 남여 평균 오거부율 (FRR)은 5.835%, 오수락율 (FAR)는 0.3%를 나타냈다. 표1과 2에서 보여주듯이 제

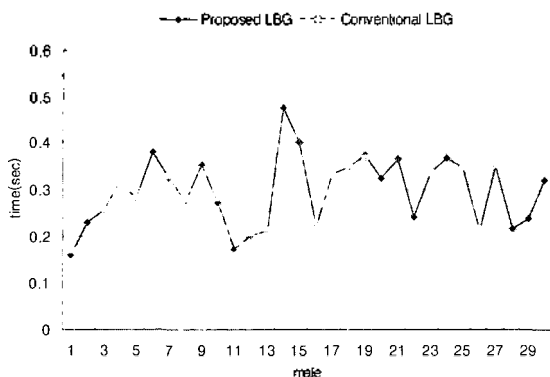


그림 4. 남성 화자 코드북 수렴속도
Fig. 4. Codebook convergence rate.

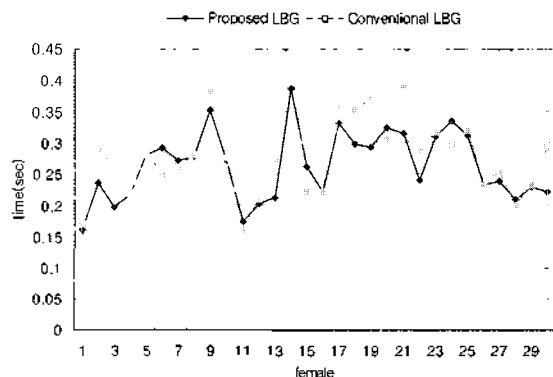


그림 5. 여성 화자 코드북 수렴속도
Fig. 5. Codebook convergence rate.

안한 LBG 알고리즘을 화자확인 시스템에 적용하였을 경우 3.165%의 오거부율 (FRR) 감소와 0.06%의 오수락율 (FAR)가 감소됨을 확인하였다.

모의 실험 결과, 제안한 알고리즘은 실제 응용분야에서 가져오는 학습 데이터의 불충분으로 인한 화자 확인 오류율을 보상해 줌을 알 수 있다. 그림 6은 제안한 LBG 알고리즘을 사용한 화자 확인 시스템의 성능 평가 비교 결과를 나타낸 것이다.

V. 결론

본 논문에서는 소량의 학습데이터를 사용하는 화자 확인 시스템에서 발생하는 군집화 알고리즘의 오류로 인한 인식 성능의 저하를 극복하기 위해서 개인별 최적 크기 코드북을 생성하기 위한 개선된 LBG 알고리즘을 제안하였다. 즉, 사용자 편의성을 고려한 화자확인시스템에서는 최소한의 발성음으로 최적의 화자모델을 등록하도록 설계하는데 본 연구에서는 이러한 조건에 만족하도록 "안녕하세요" 와 같은 간단하고 자연스러운 발성어절을 3회 반복하여 등록과정을 수행하였다. 이 경우 한 어절의 프레임 길이는 73샘플 간격으로 윈도우를 이동할 경우 사용자마다 약 80~130 프레임 정도로 분석된다. 따라서 이를 3회 반복하여 얻은 전체 분석프레임은 약 320 프레임 정도가 되므로 기존의 LBG나 K-mean와 같은 집단화 방식을 통한 VQ/HMM 인식기에서는 많은 집단화 오류에 의한 인식성능 저하현상을 가져오게 된다.

실험 결과 기존의 LBG 알고리즘을 사용 했을 때의 경우보다 제안한 LBG 알고리즘을 사용한 경우가 오거부율 (FRR)은 3.165%, 오수락율 (FAR)는 0.06%씩 각각 향상 되었다. 또한, 수렴 속도면에서도 제안한 방식이

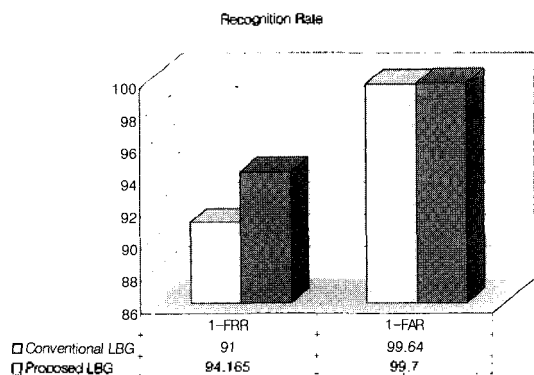


그림 6. 제안한 LBG 알고리즘의 성능 비교도
Fig. 6. Performance chart of the proposed algorithm.

기존 방식과 비교할 때 평균 0.021초 빠르게 수렴함을 알 수 있었다. 수정된 LBG 알고리즘은 학습 데이터가 적은 화자 확인 시스템에서 성능과 계산량 측면에서 모두 우수한 성능을 나타내었다.

감사의 글

본 논문은 2003년도 광운대학교 교내학술연구비 지원으로 수행되었음.

참고 문헌

1. F. K. Soong, A. E. Rosenberg, L. R. Rabiner, and B. H. Juang, "A vector quantization approach to speaker recognition", Proc. ICASSP'85, pp.387-390, March 1985.
2. Yoseph Linde, Andrés Buzo, Robert M. Gray, "An Algorithm for Vector Quantizer Design", IEEE Trans. Communications, vol. COM-28, pp. 84 - 95, January 1980
3. Allen Gersho and Robert M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.
4. M. R. Anderberg, *Cluster Analysis for Applications*, Academic, New York, 1973.
5. 진세훈, 이재희, 강철호, "화자확인 시스템을 위한 적응적 모델 갱신과 사전 문턱치 결정에 관한 연구", 한국 음향학회지, 19(5), pp.20-26, 2000년 7월.

저자 이력

◦ 정희석 (Hee-Suk Jeong)



1996년 8월: 광운대학교 전자통신공학과 (공학사)
1998년 8월: 광운대학교 일반대학원 전자통신공학과 (공학석사)
1999년 3월~현재: 광운대학교 전자통신공학과 박사과정
2002년 1월~현재: (주)한국파워보이스 대표이사
*주요관심분야: 음성인식, 화자인식, 적응신호처리

◦ 강철호 (Chul-Ho Kang)



1975년 2월: 한양대학교 전자공학과 졸업 (공학사)
1979년 2월: 서울대학교 대학원 전자공학과 졸업 (공학석사)
1988년 2월: 서울대학교 대학원 전자공학과 졸업 (공학박사)
1977년 3월~1982년 2월: 국방과학연구소 연구원
1991년 1월~1992년 1월: 미국 일리노이대학교 객원교수
2000년 3월~2001년 2월: 중국 연변 과학기술대학교 교환교수

1983년 3월~현재: 광운대학교 전자통신공학과 정교수
*주요관심분야: 음성신호처리, 적응신호처리, 통신신호처리