논문 2007-44CI-1-4

양자화 된 범용 화자모델을 이용한 연속적 화자분류

(Sequential Speaker Classification Using Quantized Generic Speaker Models)

권 순 일*

(Soonil Kwon)

요 약

연속적 화자 분류에 있어서 분류 대상이 되는 화자에 대한 정보가 없거나 부족할 경우 정확한 연속적 분류가 어렵다. 이러한 문제를 해결하기 위해 표본 화자모델을 이용하는 방법이 제안되었는데, 이 방법을 이용하면 미리 준비된 화자의 데이터가 없이 화자모델 초기화와 화자분류가 가능해진다. 하지만 여전히 화자모델의 표본을 얻는 방법에 어려움이 따른다. 이 문제를 해결하기 위해 벡터 양자화에서 비롯된 화자 양자화를 제안한다. 유선전화 데이터를 이용한 실험에서 화자 양자화를 이용한 표본 화자모델 방법은 무작위 표본추출 방법을 이용할 경우 보다 25%의 성능 향상을 보였다.

Abstract

In sequential speaker classification, the lack of prior information about the speakers poses a challenge for model initialization. To address the challenge, a predetermined generic model set, called Sample Speaker Models, was previously proposed. This approach can be useful for accurate speaker modeling without requiring initial speaker data. However, an optimal method for sampling the models from a generic model pool is still required. To solve this problem, the Speaker Quantization method, motivated by vector quantization, is proposed. Experimental results showed that the new approach outperformed the random sampling approach with 25% relative improvement in error rate on switchboard telephone conversations.

Keywords: 연속적 화자분류, 범용 화자모델, 표본 화자모델, 화자 양자화.

I. 서 론

화자인덱싱(Speaker Indexing)은 다자간 대화를 포함하고 있는 음성신호로부터 누가 언제 말을 하고 있는지를 알려준다. 화자인덱싱은 화자인식을 연속적으로 수행한다고 볼 수 있기 때문에, 화자인식을 수행할 때와 마찬가지로 미리 준비된 화자들의 정보가 필요하다. 하지만, 어떤 경우에는 인식 대상이 되는 화자들에 대한 정보를 미리 갖고 있을 수 없다. 예를 들어, 방송 뉴스를 대상으로 화자인덱싱을 할 경우, 앵커들이나 기자, 리포터들에 대한 정보는 미리 갖고 있을 수 있으나, 인터뷰를 하는 불특정한 사람들에 대한 정보는 일일이 미리 준비될 수가 없다. 그래서 화자인덱싱을 수행할 때, 미리 알 수 없는 화자들에 대한 대비책이 필요한 것이다.

연속적인 화자분류는 화자인덱싱의 일종이지만, 특히 미리 준비된 정보가 없는 화자들만으로 이루어진 경우의 화자인덱싱을 보완하기 위한 것이다. 화자들을 인식할 수 없는 대신 서로 다른 화자들이라는 것을 구분해 내는 작업이라고 할 수 있다. 하지만, 화자들에 대한 정보가 없을 때 화자들의 모델을 미리 만들어 놓지 못한 상황에서 그것들을 대체할 모델이 필요하게 된다. 순차적으로 들어오는 음성정보를 이용하여 화자모델들을 초기화 시킬 수도 있겠지만, 이럴 경우 화자모델을 만드는데 필요한 절대적인 데이터의 양이 부족하여 도리어 오류를 유발시키는 원인이 된다[1][9][10][11][13][17]. 이러한

* 정회원, 한국과학기술연구원 시스템연구부
(Division of Systems Technology, Korea Institute of Science and Technology)
접수일자: 2006년12월6일, 수정완료일: 2007년1월11일

문제를 해결하기 위해 제안된 것이 범용 화자모델이다.

범용 화자모델은 분류의 대상이 되는 화자들이 아닌 다른 화자들의 정보를 가지고 만든 모델이다. 지금까지 제안되어온 범용 화자모델에는 Universal Background Model(UBM)과 Gender Model(GM)이 있다. 먼저 범용 화자모델을 만들기 위해서는 많은 수의 화자들로 이루어진 화자 풀(Pool)이 미리 준비되어 있어야 한다. UBM은 범용 모델을 만들기 위해 준비되어 있는 화자 풀에 존재하는 모든 화자들의 정보를 한데 묶어 만든 화자 모델이고, GM은 화자 풀에 있는 화자들을 성별로 나누어 만든 남성과 여성모델로 이루어진 것이다. 최근에 발표한 논문에 제안된 범용 화자모델로 표본 화자모델(Sample Speaker Models)이 있는데, 이것은 화자 풀에서 일정 수의 화자를 추출하여 각각의 화자모델을 만든 것이다[9][11][16].

표본 화자모델은 기존의 범용 화자모델들에 비해 그 성능이 우수한 것으로 나타났다. 하지만 다른 범용모델들과 달리 추가적으로 필요한 것이 있는데, 화자 풀에서 어떤 방법으로 표본을 추출하느냐는 것이다. 기본적으로 무작위 표본추출 방법을 이용할 수도 있겠지만, 이보다 더 나은 방법에 대한 연구가 필요하다. 이 논문에서는 표본 화자모델을 만들기 위한 표본추출 방법에 있어서 벡터 양자화(Vector Quantization)에서 비롯된 화자모델 양자화를 이용함으로써 연속적인 화자분류의 성능을 높이는데 기여하고자 한다.

이번 논문에서는 기본적으로 사용되는 Universal Background Model(UBM)과 무작위 추출을 이용한 표본 화자모델, 그리고 양자화 된 화자모델을 이용한 표본 화자모델의 연속적 화자분류에 대한 기여도를 실험을 통해 비교해 보고자 한다. 실험 대상으로는 두 화자의 유선전화 음성대화와 방송용 뉴스가 이용되었다. 실험에서 양자화 된 표본화자 모델을 이용한 연속적 화자분류가 가장 우수한 성능을 보였다.

최근까지 벡터 양자화를 이용한 화자인식, 화자분류에 대하여 연구되어 오고 있다. Kinnunen은 벡터 양자화 방법을 이용하여 화자인식을 하였는데, 미리 벡터 코드북(Codebook)을 만들어 놓고 인식대상이 되는 화자의 입력 벡터와 코드북사이의 Euclidean Distance를 측정하여 비교하는 방식 이었다 [8]. 한편, Nishida and Kawahara는 화자인덱싱에 있어서 Gaussian Mixture Model(GMM)과 벡터 양자화 방법의 장단점을 이용하여 환경에 따라 유리한 방법을 선택할 수 있도록 하였다. 즉, 화자모델을 만들기 위한 데이터의 양이 적을 때는 벡터 양자화가 유리하고, 그렇지 않을 때는 GMM을 이용한 방법이 유리한데, 이를 Bayesian Information Criterion(BIC)을 활용하여 선택적으로 사용함으로써 화자인덱싱 성능을 향상 시킬 수 있었다[14]. 기존에 연구된 논문들을 분석해 볼 때, 지금까지의 연구들은 벡터 양자화를 이용하여 단순히 화자인식을 하는 것에 지나지 않았지만, 이 논문에서는 벡터 양자화의 방법을 이용하여 다수의 화자들을 양자화 함으로써 적절한 표본을 추출할 수 있게 만들었다.

이 논문은 다음과 같이 구성되어 있다. 다음 장인 본론은 크게 세 절로 나누어져 있는데, 첫 번째 절에서는 연속적 화자분류에 대해 소개되어 있고, 두 번째 절에서는 표본 화자모델에 대한 설명과 기존 범용 화자모델과의 비교가 되어 있다. 세 번째 절에는 화자 양자화에 대한 제안과 방법 등에 대해 자세한 설명이 되어있다. 그 다음 장에서는 실험의 취지와 방법, 그리고 결과에 대한 분석 및 토의가 서술되어 있고, 마지막 장에서는 이번 논문에 대한 결론이 기술되어 있다.

Ⅱ. 본 론

1. 연속적 화자분류

연속적으로 입력되는 복수의 화자 음성데이터를 화자 별로 분류해 주는 것을 연속적 화자분류라고 한다. 이러한 프로세스는 일반적으로 화자인덱싱이나 연속적 화자인식이라고도 일컫지만, 인식의 대상이 되는 화자

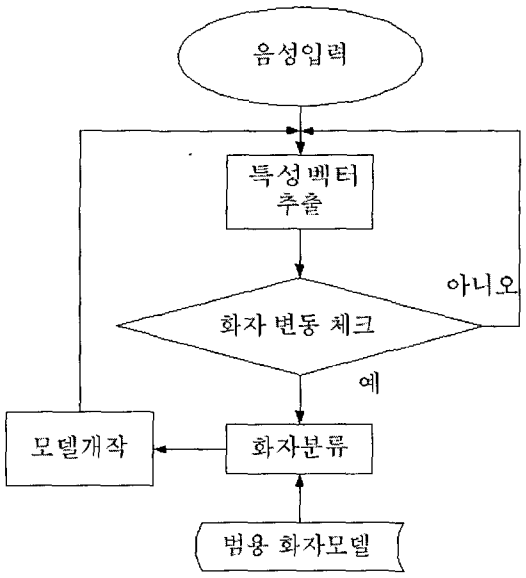

그림 1. 연속적 화자분류 블록도

Fig. 1. Block diagram of Sequential Speaker Classification.

에 대한 정보가 없을 경우 화자를 인식을 한다고 말할 수 없기 때문에, 그 대신 서로 다른 화자들의 음성을 구분하여 분류해 준다는 취지에서 화자분류라고 명명하였다.

그림 1은 일반적으로 연속적 화자분류가 이루어지는 과정에 대해 보여준다. 화자 간 대화의 음성신호가 입력되면, 먼저 신호로부터 화자 음성의 특성을 잘 표현해 줄 수 있는 벡터들이 추출되고, 그 벡터들을 순차적으로 관찰하여 화자가 바뀌는 시점을 찾는다. 화자가 바뀌는 시점에서 그 이전까지 얻은 음성 데이터를 이용하여 현재 화자와 특성이 가장 비슷하게 범용 화자모델을 변화시키거나 표본 화자모델들 중에서 찾는다. 음성 입력이 종료될 때까지 위의 과정이 반복 된다 [11].

2. 표본 화자모델

범용 화자모델은 화자인식에 있어서 그 대상이 되는 화자들에 대한 데이터, 모델 등의 정보가 없을 경우 그 대안으로 사용된다. 최근까지 제안된 범용 화자모델에는 Universal Background Model(UBM)과 Gender Model (GM)이 있다. 범용 화자모델들은 인식 대상의 화자들이 아닌 별도의 화자들을 이용하여 만들어진다. UBM은 범용 모델을 만들기 위해 준비되어 있는 모든 화자들의 정보를 가지고 하나의 화자모델을 만드는 것인데, 성별 구분 없이 수백 명에서 수천 명의 화자정보가 사용된다. 이렇듯 많고 다양한 화자정보가 사용되다보니, 화자모델의 분산이 매우 크다[그림 2.(a)]. GM은 범용 모델을 만들기 위해 준비되어 있는 모든 화자들을 성별로 구분하여 두개의 화자모델을 만든 것으로 UBM에 비해 화자모델의 분산은 작아지지만, 이 역시 다수의 화자 정보를 가지고 소수의 화자모델이 만들어 진다는 점에 있어서 UBM과 비슷하다고 볼 수 있다[그림 2.(b)].

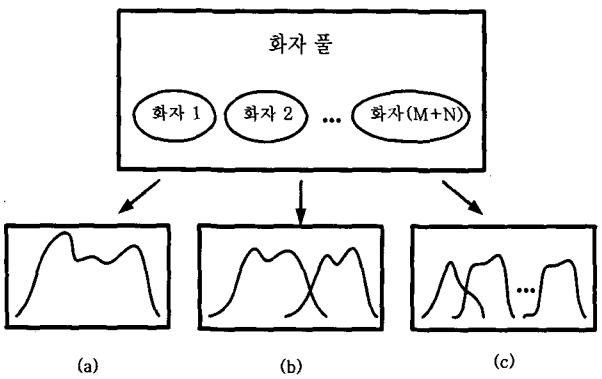

그림 2. 범용 화자모델의 예 :
(a) UBM, (b) GM, (c) 표본 화자모델
Fig. 2. Example of generic models.

지금까지 설명한 범용 화자모델들의 공통점은 다수의 화자정보가 혼합되어 한두 개의 화자모델을 만든다는 것이다. 이러한 방법을 사용할 경우 화자모델의 분산이 커지므로, 범용 화자모델로부터 파생(Adaptation)되어 나오는 화자모델들의 서로 겹침(overlap)이 심화될 수밖에 없다. 또한 화자모델 자체도 평균화 된 화자들을 표현하기 때문에 화자분류 능력이 저하 된다. 이러한 문제점을 해결해 보려고 제안된 방법이 표본 화자모델(Sample Speaker Models)이다.

표본 화자모델은 화자 풀에서 필요로 하는 수의 화자를 추출하여 각각의 화자마다 자신의 화자모델을 만든 것이다 [그림 2.(c)]. 표본 화자모델이 포함하고 있는 각각의 화자모델들은 다수의 화자에 대한 정보가 섞여 있지 않아 인위적이 아닌 실질적인 화자모델들로 이루어져 있다. 그렇기 때문에 각 화자모델의 분산이 작고, 서로간의 겹침도 적어질 뿐 아니라, 범용 화자모델로부터 파생(Adaptation)의 효과도 커질 수 있다 [9][10][11][12][15].

표본 화자모델이 다른 범용 화자모델들에 비해 화자분류를 위한 화자모델 초기화 측면에서 우수함을 보여줄 수 있는 방법 중의 하나가 Kullback-Leibler(KL) Distance 측정해 보는 것이다. KL distance 측정을 통해 범용 화자모델이 실제 분류될 화자모델과 얼마나 유사한 지를 정보이론 측면에서 알아볼 수 있기 때문이다.

KL Distance는 두개의 통계적 모델의 유사도를 측정하기 위한 것인데, Gaussian Mixture Model(GMM)을

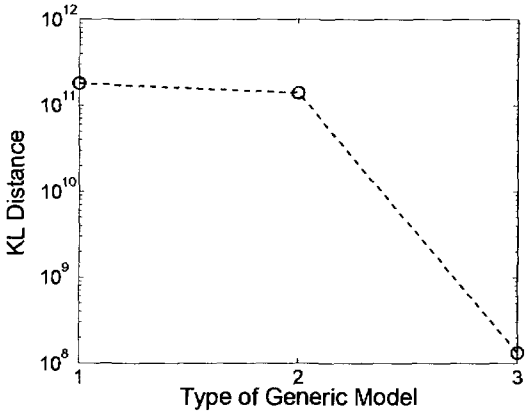

그림 3. 범용 화자모델들과 분류 대상 화자들의 실제 화자모델간의 KL Distance 측정결과 : 1. UBM, 2. GM, 3. 표본 화자모델
Fig. 3. KL distances between generic models and target speaker models.

사용하는 화자모델간의 정확한 KL Distance 측정은 거의 불가능하기 때문에 아래와 같은 수식을 사용하여 그것의 대략적인 거리를 측정하게 된다 [2][3][5][7]. 수식 (1)에서 M은 GMM의 Mixture 개수이고, N_{ij}는 모델 i 의 j 번째 Gaussian 모델을 의미하며, w는 가중치이다.

$$D_{KL} = D(\sum_{i=1}^{M} w_{1i}N_{1i} \mid \sum_{j=1}^{M} w_{2j}N_{2j})$$

$$\leq \sum_{i,j=1}^{M} w_{1i}w_{2j}D(N_{1i} \mid N_{2j}) \qquad (1)$$

화자 500명을 대상으로 사전실험을 한 결과로는 표본 화자모델이 실제 분류대상의 화자모델과 가장 짧은 KL Distance, 다시 말해 가장 유사한 것으로 나타났다 [그림 3].

3. 화자 양자화

표본 화자모델이 분류 대상 화자의 정보 없이 연속적 화자분류를 하는데 있어서 다른 범용 화자모델들에 비해 더 나은 성능을 보여줄 수 있다. 하지만, 표본 화자모델을 사용하는데 있어서 다른 범용 화자모델을 사용할 때와 달리 표본추출 과정이 필요한데, 어떤 표본추출 방법을 사용할 것인가가 문제가 된다.

화자모델이 존재하는 공간에는 많은 화자모델들이 서로 겹쳐서 위치해 있다. 이들 중 어떤 상대적 또는 절대적 위치를 차지하고 있는, 몇 개의 표본을 추출해야 하는가는 쉽지 않은 문제이다. 너무 많은 수의 표본들은 한 화자가 두 개 이상의 표본과 유사하다고 여겨질 확률이 높아져 오류를 발생시킬 여지가 커질 것이고, 반대로 너무 적은 수의 표본들은 서로 다른 화자들이

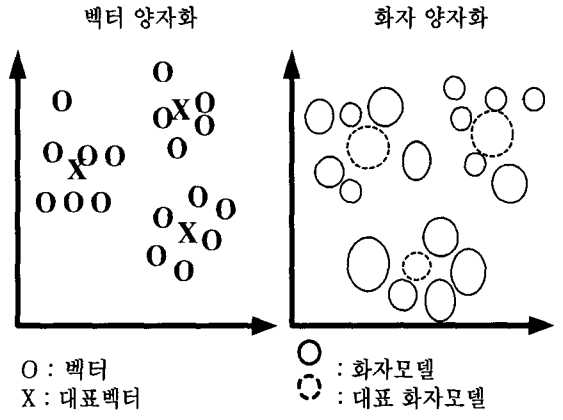

그림 4. 벡터 양자화와 화자 양자화 비교

Fig. 4. Illustration of conventional vector quantization and the proposed Speaker Quantization.

한 개의 표본과 유사하다고 분류될 확률을 높이는 문제를 야기할 수 있다. 그런데, 두 가지 경우에 각각 증가하는 오류는 서로 상반되어, 두 가지 오류확률을 동시에 낮추는 것은 거의 불가능하다고 할 수 있다 [6]. 그렇다면 가장 이상적인 표본은 어떻게 추출할 수 있을지가 관건이라 말할 수 있다.

이 논문에서는 단순하게 무작위로 표본을 추출하는 방법보다 더 좋은 화자분류 성능을 보여줄 수 있는 방법을 제시하고자 한다. 화자모델들의 공간에서 적당한 간격을 가지고 있는 적당한 수의 표본을 추출하기 위해서, 기존의 벡터 양자화의 개념을 응용하여 화자모델들의 표본을 추출하는 화자 양자화라는 방법을 제안한다. 기본적인 벡터 양자화의 경우 양자화 될 벡터의 수를 미리 알고 있는 반면, 이 논문에서는 연속적으로 분류될 화자의 수를 정확히 모른다는 가정 하에 있기 때문에, 일반적인 벡터 양자화 방법이 아닌 Tree Structured Vector Quantization (TSVQ)의 방법을 활용한다. TSVQ를 실행할 때에는 최종적으로 남는 양자화 된 벡터의 수를 미리 정할 필요가 없다. 하지만 어떤 벡터그룹이 분할될 필요가 있는지를 결정하기 위해 문턱값(Threshold)을 미리 정해야 한다 [4].

화자 양자화는 벡터 양자화를 활용한 것이지만, 그 개념에는 차이가 있다. 벡터 양자화는 벡터들을 어떤 기준에 의해 몇 개의 벡터 그룹으로 나누고, 각 벡터 그룹을 대표 해 줄 수 있는 벡터를 구하는 것이다. 이와 달리 화자 양자화는 화자모델들을 어떤 기준에 따라 몇 개의 화자모델 그룹으로 나누고, 각 그룹에 속하는 화자모델 중에서 그 그룹을 대표해 줄 수 있는 화자모델을 선정한다 [그림 4]. 또 다른 차이점은, 벡터 양자화를 위해서는 벡터들 간의 거리를 측정하게 되는데, 이때 벡터 양자화에서는 일반적으로 유클리디언(Euclidean) Distance를 이용 하지만, 화자 양자화에서는 벡터 간의 거리 계산이 아닌 통계적 모델 간의 유사도 측정을 해야 하므로, 다른 측정방법이 요구된다. 이런 이유로 화자 양자화에서는 모델 간의 유사도 측정방법으로 앞 절에서 소개한 KL Distance를 사용한다.

이진트리를 이용하여 화자를 양자화하기 위해서는 먼저 모든 화자들의 모델을 만든다. 화자 양자화에서는 화자모델들을 이용하여 화자들을 그룹화 한다. 처음 모든 화자모델들을 두 그룹으로 나누고, 각 그룹마다 그 그룹을 대표할 수 있는 대표화자모델을 선택하는데, 이 모든 과정은 KL Distance를 이용하게 된다. 즉 그룹 내의 모든 화자모델 간의 평균 KL Distance가 최소가 될

수 있도록 그룹을 나누고, 대표화자모델을 선택한다. 이러한 이진분할 과정이 반복이 되는데, 분할 전과 후의 평균 KL Distance의 상대적 변화량(R)에 따라 분할이 정지된다. 수식 (2)에서 D_r은 분할 전의 평균 KL Distance 이고, D_l은 분할 후의 평균 KL Distance 이며, ϵ은 분할 진행 여부를 결정하기 위한 문턱값이다[10].

$$R = \frac{D_r - D_l}{D_l} \geq \epsilon \qquad (2)$$

구체적인 화자 양자화 절차는 아래와 같다.

- Step 1: 각 화자모델그룹을 Lloyd algorithm에 따라 두 그룹으로 나눈다.
- Step 2: 평균 KL Distance의 상대적 변화량(R)을 계산한다.
- Step 3: 만약 평균 KL Distance의 상대적 변화량(R)이 ϵ보다 크면 분할 성공적인 것으로 보고 Step 1으로 돌아가 모든 Step들을 차례로 반복한다. 하지만 R이 ϵ보다 작으면 분할이 의미가 없는 것으로 보고 분할을 취소하고 멈춘다.

위의 방법을 이해하는데 있어서 유의할 점은 처음 시작단계에서는 그룹이 한 개이지만, 그 다음부터는 복수개의 그룹이 존재하고 분할이 계속되고 멈추는 것은 각 그룹별로 이루어지기 때문에, 어느 한쪽은 분할이 멈추더라도 다른 한쪽은 분할이 계속될 수 있고, 모든 그룹의 분할이 멈출 때까지 절차가 계속된다.

III. 실 험

이 논문에서 제안된 방법을 검증하기 위한 실험을 위해 화자데이터는 Speaker Recognition Benchmark NIST Speech (1999) Corpus에 있는 400명(여자 240명, 남자 160명)을 사용하였다. 각 화자별 50초 길이의 데이터로 GMM(16 mixtures) 화자모델을 만들었고, 약 8초의 데이터를 테스트용으로 사용하였다. 화자 양자화를 위한 ϵ은 0에 가까운 값으로 정해서 얻은 양자화 된 화자의 수는 70이었다. 비교실험을 위해 무작위 표본추출 방법을 이용하여 400명의 화자 중에서 70명의 화자를 추출하였다.

먼저 화자 양자화를 통해 얻은 표본이 무작위로 추출된 화자에 비해 얼마나 적절하게 추출 되었는지 알아보기 위해 표본 화자모델들과 실제 분류 대상이 되는 화자들의 모델들 사이의 평균 KL Distance를 계산하여 비교해 보았다. 표 1에서 볼 수 있듯이 화자 양자화를 이용한 방법이 무작위 추출방법에 비해 더 작은 평균 KL Distance를 가지고 있고 이것은 표본추출에 있어서 성능이 뛰어나다는 것을 의미한다.

표 1. 무작위 추출방법과 화자 양자화 방법으로 얻어진 표본 화자모델과 실제 분류대상 화자모델 간의 평균 KL Distance 비교

Table 1. Average KL distances: random selection vs. Speaker Quantization.

	평균 KL Distance(x 10^{10})
무작위 표본추출	0.75
화자 양자화	0.70

표 2. 연속적 화자분류의 오류률 비교: 괄호 안은 UBM기준 상대적 오류률

Table 2. Error rates of Sequential Speaker Classification: the figures in parentheses show the relative improvement from the baseline.

	전화	방송 뉴스
UBM	14.6 %(-)	25.2 %(-)
표본 화자모델 (무작위 표본추출)	12.2 %(16.4 %)	19.1 %(24.2 %)
표본 화자모델 (화자 양자화)	9.1 %(37.7 %)	14.5 %(42.5 %)

다음으로 표본 화자모델과 더불어 화자 양자화 방법을 이용할 경우 다른 범용 화자모델이나 표본 화자모델을 무작위 표본추출 방법과 함께 사용할 경우에 비하여 실제 연속적 화자분류에 있어서 얼마나 성능을 향상시킬 수 있는지 실험을 통하여 조사해 보았다. 비교 대상이 되는 범용 화자모델로는 400의 화자 데이터들로 만든 UBM을 사용하였다. 연속적 화자분류의 대상으로는 약 24분 정도의 분량의 다양한 화자들 간의 유선전화 통화 데이터(Speaker Recognition Benchmark NIST Speech)와 약 45분 길이의 방송뉴스 데이터(HUB-4 Broadcast News Evaluation English Test Material)를 이용하였다. 연속적 화자분류 성능평가를 위한 오류률은 화자분류에 실패한 일련의 음성 데이터의 시간적 길이를 테스트한 전체 음성 데이터의 시간적 길이로 나눈 값으로 계산하였다.

표 2는 연속적 화자분류에 대한 결과는 화자 양자화를 이용한 표본 화자모델이 가장 우수하다는 것을 보여준다. 실험 결과로 볼 때, 표본 화자모델을 이용한 방법이 UBM을 사용할 때보다 최소 16.4%에서 최고 42.5%의 상대적 성능향상을 보였다. 표본 화자모델에 있어서 표본 추출방법에 대한 비교를 해볼 때, 무작위 표본 추출에 대해 화자 양자화 방법이 전화통화 데이터에 있어

서는 3.1% 절대적 향상(25.4% 상대적 향상)을 보였고, 방송뉴스 데이터에 있어서는 4.6% 절대적 향상(24.1% 상대적 향상)을 보였다.

Ⅳ. 결 론

연속적인 화자인식이 필요한 화자인덱싱의 문제점을 해결하기 위한 방안으로 활용되는 기존의 범용 화자모델들은 준비된 화자들의 평균화된 정보를 이용하기 때문에 실제 단독 화자의 정보가 보이는 특성을 제대로 반영하지 못한다. 이러한 점을 보완하기 위해 제안된 것이 표본 화자모델이고, 실험적 평가에 의해 기존의 범용 화자모델들에 비해 우수한 것으로 나타났다. 하지만 표본 화자모델은 표본을 추출하는 과정이 필요한데, 서로 겹쳐 있는 수많은 화자모델들 중에서 최소로 겹쳐 있는 필요한 수만큼의 화자모델을 추출하는 것에 어려움이 있다. 무작위로 표본을 추출하는 것보다는 화자 양자화 방법을 이용함으로써 원하는 조건의 표본 화자 모델들을 얻을 수 있게 되었다.

연속적 화자분류의 실험 결과로 볼 때, 화자 양자화를 이용한 표본 화자모델 방법은 기존의 범용 모델인 UBM 뿐만 아니라 무작위 추출방법을 이용한 표본 화자모델 방법보다도 우수한 것으로 나타났다. 하지만 앞으로도 계속 연구해야 할 과제들이 있다. 첫째는 벡터 양자화와는 달리 화자 양자화에 있어서는 화자 모델들을 양자화 하여 표본 모델들을 선택하는 것인데, 이때 화자 모델들 간의 거리 또는 유사성을 측정하는데 사용되는 KL Distance에 개선되어야 할 여지가 있다. 복수개의 Gaussian 모델들의 결합으로 이루어진 GMM들 간의 거리를 KL Distance로 정확하게 측정하는 것이 무척 힘들기 때문이다. 둘째로는 화자 양자화에 의해 정해진 표본 화자모델들이 분류 될 화자들의 수나 특성에 관계없이 항상 이상적일 수 있는가 하는 점이다. 이러한 문제들을 해결하기 위해서는 정보 이론적(Information Theoretic) 고찰과 알고리즘적인 보완이 필요할 것으로 본다. 또한 위의 방법을 일반화시키기 위해 더 많은 화자들을 대상으로 실험해 보는 것도 필요할 것이다.

참 고 문 헌

[1] J. P. Campbell, "Speaker recognition: A tutorial," in Proc. of *IEEE*, Vol. 85, pp. 1436-1462, 1997.

[2] T. M. Cover and J.~A. Thomas, *"Elements of Information Theory,"* Wiley Interscience, New York, pp. 18- 19, 1991.

[3] M. Do, "Fast Approximation of Kullback-Leibler Distance for Dependence Trees and Hidden Markov Models," *IEEE Signal Processing Letters*, Vol. 10, pp. 115-118, 2003.

[4] R.M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. on Information Theory*, Vol. 44, pp. 2325-2383, 1998.

[5] T. Hastie, H. R. Tibshirani and J. Friedman, *"The Elements of Statistical Learning,"* Springer, New York, pp. 496-498, 2001.

[6] R. V. Hogg and E. A. Tanis, *"Probability and Statistical Inference,"* 6th ed. Prentice Hall, New Jersey, pp.85-102, 2001.

[7] A. Jain, P. Moulin, M. I. Miller and K. Ramchandran, "Information-Theoretic Bounds on Target Recognition Performance Based on Degraded Image Data," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 1153-1166, 2002.

[8] T. Kinnunen, T. Kilpelainen and P. Franti, "Comparison of Clustering Algorithms in Speaker Identification," in Proc. of *International Conf. of Signal Processing and Communications (SPC 2000)*, pp. 222-227, 2000.

[9] S. Kwon and S. Narayanan, "A Study of Generic Models for Unsupervised On-Line Speaker Indexing," in Proc. of *IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 423-428, St. Thomas, U.S. Virgin Islands, 2003.

[10] S. Kwon and S. Narayanan, "Speaker Model Quantization for Unsupervised Speaker Indexing," in Proc. of *International Conf. Spoken Language Processing*, WeC2102p.18, Jeju, Korea, 2004.

[11] S. Kwon and S. Narayanan, "Unsupervised Speaker Indexing Using Generic Models," *IEEE Trans. on Speech and Audio Processing*, Vol. 13, Issue 5, Part 2, pp.1004-1013, 2005.

[12] M. Liu, E. Chang and B. Q. Dai, "Hierarchical Gaussian Mixture Model for Speaker Verification," in Proc. of *International Conf. on Spoken Language Processing*, Vol. 2, pp. 1353-1356, Denver, U.S.A., 2002.

[13] L. Lu, H. J. Zhang and H. Jiang, "Content Analysis for Audio Classification and Segmemtation," *IEEE Trans. on Speech and Audio Processing*, Vol. 10, pp. 504-516, 2002.

[14] M. Nishida and T. Kawahara, "Unsupervised Speaker Indexing Using Speaker Model Selection Based on Bayesian Information Criterion," in

Proc. of *IEEE International Conf. on Acoustics, Speech and Signal Processing*, Vol. 1, pp. 172-175, Hong Kong, China, 2003.

[15] J. Wu and E. Chang, "Cohorts Based Custom Models for Rapid Speaker and Dialect Adaptation," in Proc. of *Eurospeech*, pp. 1261-1264, Aalborg, Denmark, 2001.

[16] T. Wu, L. Lu, K. Chen and H. Zhang, "UBM-Based Real-Time Speaker Segmentation for Broadcasting News," in Proc. of *IEEE International Conf. on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 193-196, Hong Kong, China, 2003.

[17] J. Yang, X. Zhu, R. Gross, J. Kominek, Y. Pan and A. Waibel, "Multimodal People ID for a Multimedia Meeting Browser," in Proc. of 7th *ACM International Conf. on Multimedia*, Part 1, pp. 159-168, 1999.

저 자 소 개

권 순 일(정회원)

1998년 연세대학교 전자공학과 학사 졸업.

2000년 University of Southern California 전기공학과 석사 졸업.

2005년 University of Southern California 전기공학과 박사 졸업.

2005년~2006년 삼성전자 정보통신총괄 통신연구소 책임연구원

현재 한국과학기술연구원 재직.

<주관심분야 : 음성신호처리, 음성인식, 화자인식, HCI, HRI>