

문맥종속 화자인식을 위한 준비반복 벡터 양자기 설계 알고리즘

임 동 철[†] · 이 행 세^{††}

요 약

이 논문은 문맥 종속 화자인식에 사용될 벡터 양자기의 설계법 개선에 관한 연구이다. 구체적으로 벡터 양자기 코드북 생성 과정에서 기준 화자를 제외한 모든 비기준 화자에 대해 비반복적 학습 방법을 사용하여, 학습에 필요한 계산 복잡도를 획기적으로 줄이는 방법을 제안한다. 이 제안된 준비반복 벡터 양자기 설계법은, 종래의 설계법이 모든 화자의 코드북 생성에 반복적 학습 설계를 사용한다는 것과 대조를 이룬다. 준비반복 벡터 양자기 설계법의 특징은 다음과 같다. 첫째, 이 설계법은 단지 기준 화자에 대하여만 반복 학습을 수행하고 비기준 화자에 대하여는 반복 학습을 하지 않는다. 둘째, 설계된 비기준 화자의 양자 영역은 기준화자의 양자 영역을 원용하며, 양자점은 자신의 통계 분포에 대해 최적점으로 설정된다. 수치 실험은 화자 20명에 대하여 멜켄스트럼 12차 특징벡터를 사용하였고 코드북 크기를 2부터 32까지 변화시키면서 기존의 벡터 양자기 인식법과 비교 하였다. 제안된 방법은 코드북 크기가 적절하고 학습 데이터 길이가 충분한 경우 인식을 100%로 기존의 방법과 같은 결과를 보였다. 따라서 제안된 준비반복 벡터 양자기 설계법은, 설계에 필요한 학습 횟수가 획기적으로 줄면서 인식률은 보존되어, 새로운 대안이 될 것으로 사료된다.

A Semi-Noniterative VQ Design Algorithm for Text Dependent Speaker Recognition

Dong Chul Lim[†] · Haing Sei Lee^{††}

ABSTRACT

In this paper, we study the enhancement of VQ (Vector Quantization) design for text dependent speaker recognition. In a concrete way, we present the non-iterative method which makes a vector quantization codebook and this method is not iterative learning so that the computational complexity is epochally reduced. The proposed semi-noniterative VQ design method contrasts with the existing design method which uses the iterative learning algorithm for every training speaker. The characteristics of a semi-noniterative VQ design is as follows. First, the proposed method performs the iterative learning only for the reference speaker, but the existing method performs the iterative learning for every speaker. Second, the quantization region of the non-reference speaker is equivalent for a quantization region of the reference speaker. And the quantization point of the non-reference speaker is the optimal point for the statistical distribution of the non-reference speaker. In the numerical experiment, we use the 12th mel-cepstrum feature vectors of 20 speakers and compare it with the existing method, changing the codebook size from 2 to 32. The recognition rate of the proposed method is 100% for suitable codebook size and adequate training data. It is equal to the recognition rate of the existing method. Therefore the proposed semi-noniterative VQ design method is, reducing computational complexity and maintaining the recognition rate, new alternative proposal.

키워드 : 벡터 양자화(Vector Quantization), 클러스터링(Clustering), 화자인식(Speaker Recognition), 계산 복잡도(Computational Complexity), 학습(Learning)

1. 서 론

1.1 논문의 목적

이 논문은 문맥 종속 화자 인식을 위한 벡터 양자화기 설계에 있어 반복 학습 없이 화자를 변별할 수 있는 코드북을 생성하는 설계 방법을 제안한다. 이 논문에서 제안된

방법을 준비반복 벡터 양자기 설계법이라 부르기로 한다. 일반적으로 화자인식에서 벡터 양자화를 이용하는 목적은 양자화 오차를 이용하여 각 화자를 변별하는 데 있다. 벡터 양자화가 인식기로서의 성능을 발휘하는 것은 벡터 공간상에서 화자의 분포를 적절히 평가함으로써 가능해진다. 이를 위해 벡터 양자화는 반복 학습에 의해 분포를 평가하게 되고 따라서 많은 학습 계산량을 필요로 하게 된다. 준비반복 벡터 양자기 설계의 목적은 이러한 반복 학습과정을 생략하고도 적절히 화자의 분포를 평가할 수 있는 학습 과정을

[†] 준회원 : 아주대학교 대학원 전자공학부

^{††} 정회원 : 아주대학교 교수

논문접수 : 2002년 9월 27일, 심사완료 : 2003년 1월 21일

설계하고 이를 이용하여 화자 인식기로서 동작하도록 하는 데 있다.

둘째로 이 논문에서 제시하는 바는 준비반복 벡터 양자기는 제안된 인식법을 사용할 경우 각 화자에 대해 최적의 양자점을 가지게 되고 따라서 효과적으로 인식이 가능하다는 것을 보이는 것이다. 준비반복 벡터 양자기는 동일 문맥에 대해 화자의 특징 벡터의 분포가 대부분 겹쳐져 있다는 가정 아래서 기준화자의 양자영역을 이용해 비기준 화자를 적절히 군집화할 수 있다. 따라서 기준화자의 군집 안에서 비기준 화자의 최적 양자점을 구할 수 있다는 것을 보임으로 제안된 방법이 이론적으로 성립함을 설명한다. 새로 정의한 용어인 기준 화자란 임의로 선택된 한 화자이다. 그리고 비기준 화자란 기준화자 이외에 학습될 모든 화자를 말한다. 기준 화자의 코드북은 기존의 벡터 양자기 방법과 같이 반복학습에 의해 생성되고 비기준 화자의 코드북은 제안된 비반복 학습과정을 거쳐 생성된다.

세째로 이 논문에서는 준비반복 벡터 양자기 설계 알고리즘을 구현하여 실험한다. 그리고 실험 결과를 분석하고 성능을 평가하여 위의 제시한 사실을 검증한다. 실험된 결과의 성능 평가를 위해 기존의 벡터 양자화기 설계법과 비교 실험을 한다. 실험 내용은 코드북 크기의 변화에 따른 인식율과 양자화 오차 그리고 반복 학습에 필요한 계산량이다.

1.2 벡터 양자기를 이용한 화자 인식법

벡터 양자기를 이용한 화자인식법은 각 화자의 고유 분포 특성을 나타내는 특징 파라미터 벡터 집합을 대표하기 위해 벡터 양자기를 사용한다. 이 기법은 각 화자의 코드북 내의 코드 벡터들이 특징 벡터 공간상에서 화자의 특징을 표현하는 분포의 중심이므로 입력된 특징 벡터가 이 영역의 중심 즉 대표점 근처에 많이 분포할수록 해당 화자의 음성일 가능성이 높다는 점을 이용한다[1].

벡터 양자기관 자료의 신뢰성을 되도록 유지하면서 소수의 벡터들로 다수의 벡터들을 대표하는 방법을 말한다. 이때 소수의 대표 벡터를 코드북이라 부르며, 이는 군집화 알고리즘에 의해 구하여진다[2]. K-means 알고리즘은 군집화를 하는데 잘 알려진 학습 알고리즘이다. 이 알고리즘은 코드북과 학습 벡터 사이의 양자화 오차를 반복 학습에 의해 감소시켜 나가는 기법이다.

벡터 양자기를 이용한 화자 인식법의 장점은 첫째, 패턴 열의 시간축 정합이 필요 없어 화자인식 시스템을 구성하는데 가장 쉬운 방법 중에 하나이며 둘째, 화자 정보의 기억 장소와 인식 과정의 계산량이 다른 인식기에 비해 적고 셋째, 문맥 종속형과 독립형 모두 구현이 가능하다는 것이다. 반면 단점은 학습 자료가 충분하지 않거나 코드북의 크기가 작을 경우 코드북은 화자 특성을 표현하기 어려워지고 코드

북을 반복 학습법을 이용하여 구하므로 국부 최소점에 빠지는 경우가 있다. 또한 새로운 화자의 등록시에 코드북을 생성하기 위하여 많은 계산량을 필요로 하는 K-means 알고리즘과 같은 반복적 클러스터링 알고리즘을 적용해야 한다 [3, 4]. 인식률은 일반적으로 코드북 크기에 비례하여 증가하며 일정 이상 코드북의 크기가 커지면 수렴하게 된다.

1.3 방법론

본 논문에서 제안된 방법은 저자의 조사한 범위에서는 아직까지 보고된 바 없는 새로운 방법이다. 일반적으로 벡터 양자기를 이용한 화자 인식법은 화자 특징 벡터의 대표를 코드북으로 생성하기 위해 K-means 알고리즘과 같은 반복 학습법을 사용한다. 이에 반하여 제안된 준비반복 벡터 양자기 설계법은 기준 화자에 대하여만 K-means 알고리즘을 통하여 코드북을 생성한다. 비기준화자에 대하여는 벡터 공간상에서 화자의 음성 특징의 분포가 동일 문맥에 대해 겹쳐지는 특징을 이용하여 코드북을 생성하게 된다. 즉 비기준 화자의 특징 벡터들은 기준화자의 코드북을 이용하여 분류된다. 그 후 비기준 화자의 코드북은 분류된 기준 화자 군집 안에서 비기준 화자의 특징 벡터들을 대표하도록 생성된다. 이렇게 구하여진 각 화자의 코드북은 기준 화자의 군집으로 분류된 경우에 각 화자에 최적화 되어 각 화자를 변별하게 된다. 인식 과정 즉 복호화 과정에서 오차를 구하기 위해서는 먼저 기준화자의 코드북을 이용하여 분류한 후 각 화자의 코드북을 이용해 화자마다의 양자화 오차를 구하게 된다. 이때 구해진 각 화자의 양자화 오차는 반복 학습을 수행하지 않은 각 화자의 코드북 임에도 최적 양자화 오차의 결과를 가지게 되어 각 화자를 인식할 수 있게 된다.

1.4 수치 결과 및 기여도

세 종류의 문맥에 대해 성인 남성 20명의 10회 발성 한 문맥종속 화자인식 실험 결과 제안된 준비반복 벡터 양자기 설계법은 코드북 크기와 학습량이 적절한 경우 기존의 벡터 양자기 설계법과 같은 인식율을 나타낸다. 또한 반복 학습에 필요한 계산량의 경우 제안된 방법은 화자수의 증가에 무관하게 일정한 반면 기존의 방법은 화자 수에 비례해 증가함을 보인다. 따라서 본 논문은 새로운 벡터 양자기 설계법을 제시하였고 화자의 특징 분포를 적절히 평가함으로써 반복 학습 없이도 인식이 가능하다는 것을 보였다. 이 방법은 기준 화자의 군집 안에서 각 화자의 변별력이 떨어질 때 까지는 화자수가 증가하여도 성능을 나타내며 적은 계산 능력을 필요로 하는 인식기에 적용될 수 있을 것이다.

1.5 논문의 구성

이 논문의 구성은 다음과 같다. 제 2장에서는 제안된 준

비반복 벡터 양자기의 코드북 생성법에 대한 설명과 이렇게 생성된 코드북이 결과적으로 기준 화자의 보로노이 영역(Voronoi Cell), 즉 양자영역 안에서 최적의 양자화 오차를 가진다는 것을 보이고 제 3장에서는 제안된 준비반복 벡터 양자기의 학습과정과 인식과정의 알고리즘을 소개한다. 제 4장에서는 실험 및 결과를 나타내고, 제 5장에서는 결론을 맺는다.

2. 코드북 생성 방법

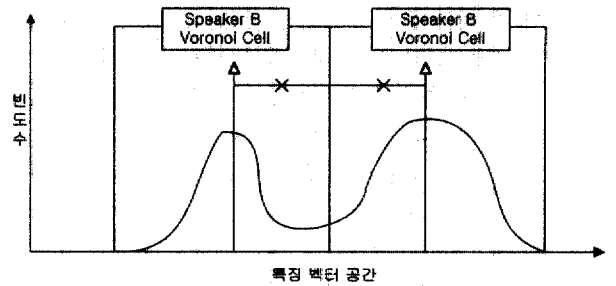
벡터 양자기를 이용한 문맥 중속 화자인식에서 생성된 코드북은 문맥의 분포특성과 발성화자의 음성의 분포특성을 나타내는 대푯값으로 볼 수 있다. 그러므로 각 화자마다 생성된 코드북은 문맥에 대한 각 사람의 음성 특징을 모형화한 것이 된다. 즉 각 화자의 특징 분포는 학습과정을 통해 대표점들로 표현되며 이 대표점들은 각 화자마다 변별된다. 또한 변별될 수 있는 새로운 화자는 기존의 화자의 대표점과 구분되는 새로운 대표점을 가져야만 한다[5].

이때 학습과정을 거친 한 화자의 특징 벡터들과 이를 대표하는 대표점들은 문맥에 종속되어 있으므로 다차원 벡터 공간상에서 문맥을 나타내는 국부적 지역에 분포하게 된다. 그리고 아직 학습되지 않은 비학습 화자들의 특징 벡터들 또한 같은 문맥을 사용하였다면 같은 국부적 지역에 분포하게 된다. 이렇게 화자의 특징 벡터들이 겹쳐져 분포하더라도 벡터 양자기가 각 화자를 변별할 수 있는 것은 각 화자의 분포가 다르고 이를 대표하는 각 화자의 대표점은 각 화자의 분포에 최적화 되어있기 때문이다[6].

본 논문의 착안점은 다음과 같다. 기준화자 - 학습된 화자 - 의 분포와 비기준 화자 - 비반복 학습 화자 - 의 분포가 겹쳐진다는 사실을 이용하여 비기준 화자를 대표하는 대표점을 반복 학습법에 의한 코드북으로 생성하지 않고 이미 군집화 되어있는 기준화자의 양자영역을 이용하여 구할 수 있다는 것이다.

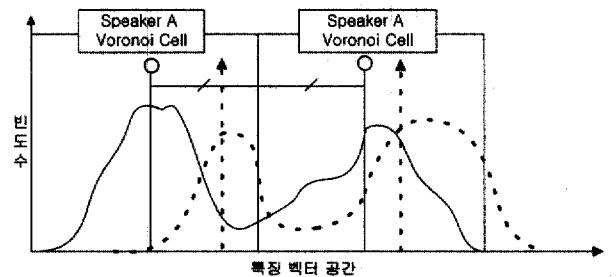
설명의 도움을 위해 먼저 일반적 벡터 양자기의 경우를 (그림 1)로 보여겠다. (그림 1)은 화자 B에 대해 K-means 클러스터링 과정을 거쳐 생성된 다차원 벡터들의 군집화된 모양을 일차원으로 투영하였을 때의 분포를 보여주고 있다. (그림 1)의 가로축은 일차원 특징 벡터 축이고 세로축은 각 특징 벡터의 빈도를 나타낸다. (그림 1)의 흰 삼각형은 각 군집, 즉 보로노이 영역의 대표점을 나타낸다. 코드북으로 생성될 이 대표점, 즉 K-means 클러스터링 과정을 마치고 생성된 양자점은 특징 벡터의 분포를 대표하는 최적화된 양자영역을 형성하며 따라서 다음 두 가지 조건을 만족한다.

- ① 각각의 대표점은 보로노이 영역을 형성한다.
- ② 각각의 대표점은 보로노이 영역의 무게중심값이다.



(그림 1) 1차원 투영된 화자 B의 특징 벡터 분포와 대표점

(그림 2)는 이 논문의 개념을 설명하기 위한 그림이다. (그림 2)는 동일한 문맥은 같은 공간상에 분포한다는 점을 보여준다. 그리고 (그림 2)는 이를 이용하면 학습되지 않은 화자의 특징 벡터들은 반복 학습된 화자의 군집으로 분할할 수 있다는 것을 보여주고 있다. (그림 2)의 실선은 화자 A의 분포를 보여주고 있다. (그림 2)의 동그라미로 표현되는 대표점은 화자 A의 대표점이며 반복학습에 의해 화자 A의 분포에 최적화 되어 있고 각 대표점은 보로노이 영역을 형성하며 보로노이 영역의 무게중심임을 볼 수 있다.



(그림 2) 학습된 화자 A의 군집에 의해 구해진 화자 B의 대표점

(그림 2)의 점선은 (그림 1)의 화자 B의 특징 벡터 분포를 나타낸다. 이런 분포의 겹쳐짐은 특징벡터가 같은 문맥에 종속되어 국부적 영역에 나타남으로서 일어나는 현상이다. 이런 현상을 이용하면 학습된 화자 A의 양자영역으로 화자 B의 분포를 군집화시킬 수 있다. 이렇게 화자 A의 양자영역에 의해 군집화된 화자 B의 반복 학습되지 않은 특징 벡터들의 분포는 반복 학습된 화자 A의 특징 벡터들의 분포와 다른 분포를 나타나게 된다. 따라서 화자 A의 양자영역에 의해 군집화된 화자 B의 특징 벡터들에 대해 무게중심점을 구하면 반복 학습된 특징벡터들의 분포를 대표하는 기존 화자 A의 대표점(흰 동그라미)과 변별되는 새로운 대표점(검은 삼각형)을 생성할 수 있다.

이때 구해진 화자 B의 대표점(검은 삼각형)은 화자 B의 전체 분포에 최적화 되어있지 않다. 즉 화자 A의 양자영역에 의한 화자 B의 대표점(검은 삼각형)은 보로노이 영역을 형성하지 않는다. 다시 말하면 (그림 1)에서 화자 B의 전체 분포에 최적화된 대표점인 흰 삼각형과 (그림 2)에서 화자 B의 대표점인 검은 삼각형은 일치하지 않는다. 그러나 (그

림 2)에서의 화자 B의 대표점인 검은 삼각형은 화자 A의 군집 안에서는 화자 B에 최적화된 양자점인 대표점이다. 즉 이 대표점은 화자 A의 양자영역 내에서 화자 B의 분포에 대해 무게중심점이고 가장 작은 오차값을 가진다.

그림을 통해 설명한 바와 같이 한번 반복학습 된 화자의 대표점을 통해 구해진 군집을 이용하면 각 화자마다 반복 학습 과정을 거치지 않더라도 반복 학습되지 않은 화자의 군집을 대표하는 변별력 있는 대표점을 찾아낼 수 있다.

3. 준비반복 벡터 양자기 설계 알고리즘

본 논문에서 제안된 알고리즘은 벡터 양자기 설계 알고리즘을 사용하나 기준 화자에 대하여만 반복학습을 수행하고 비기준 화자에게는 반복학습을 수행하지 않으므로 이를 준비반복 벡터 양자기 설계 알고리즘(A Semi-Noniterative VQ Design Algorithm)이라 명명하였다.

3.1 준비반복 벡터 양자기 학습 알고리즘

- ① 결정된 문맥에 대해 화자 S_1 으로부터 특징 벡터로 구하여 특징벡터 집합 X_1 을 추출해낸다.
- ② X_1 에 대해 K-means 클러스터링을 사용하여 화자 S_1 에 대한 코드북 C_1 을 생성한다.
- ③ 새로 등록하고자 하는 화자 S_i 에 각각에 대해 다음 과정(①~④)을 통해 코드북 C_i 를 생성한다. ($i=2, \dots, N$)
 - ① 화자 S_i 에 대해 특징 벡터 집합 X_i 를 추출해낸다.
 - ② 코드북 C_1 의 대표점 c_{1j} 로 특징 벡터 집합 X_i 을 보로노이 영역으로 구분한다.
($j=1, \dots, k$), K 는 코드북의 크기이다.
 - ③ 각 대표점 c_{1j} 에 의해 보로노이 영역으로 구분된 X_i 의 부분 집합에 대해 무게 중심점 c_{ij} 을 구한다.
 - ④ 이렇게 구해진 무게중심점 c_{ij} 를 화자 S_i 에 대한 코드북 C_i 에 등록한다.

3.2 준비반복 벡터 양자기 인식 알고리즘

- ① 입력 화자 S_i 에 대해 특징벡터 집합 $X_i = \{x_1, \dots, x_L\}$ 를 추출한다.
- ② 각각의 화자 모델 C_i 에 대해 다음 과정(①~③)을 수행한다. ($i=1, \dots, N$)
 - ① 코드북 C_i 의 대표점 c_{1j} 에 대해 특징벡터 집합 X_i 를 보로노이 영역으로 구분한다.
 - ② 분류된 부분집합 $subX_i$ 의 원소 각각에 대해 c_{ij} 와의 오차 $D_{ij} = d(x_n, C_{ij})$ 를 구한다.
 $subX_i = \{subX_{ij} | subX_{ij} \text{는 } c_{1j} \text{에 속하는 } x \text{들의 집합, } j=1, \dots, K\}$

$$d(A_j, c_{ij}) = \frac{1}{M} \sum_{p=1}^M d_E(a_{pj}, c_{ij}), (A_j = \{a_{1j}, \dots, a_{Mj}\}),$$

A_j 는 임의의 집합

여기서 d_E 는 특징 벡터간 유클리드 거리를 의미한다.

$$d_E(x, y) = \sqrt{\sum_{i=1}^{\dim} (x_i - y_i)^2}$$

③ 화자모델 C_i 와 X_i 의 오차 $D_i = \sum_{j=1}^K D_{ij}$ 를 계산한다.

④ 가장 작은 오차를 나타내는 등록된 화자의 색인 i 을 입력 화자 색인 Id 로 식별한다.

$$Id = \underset{i=1, \dots, N}{\operatorname{argmin}} \{D_i\}$$

4. 수치 실험 결과

준비반복 벡터 양자기 설계 알고리즘의 성능을 평가하기 위하여 기존의 벡터 양자기 설계법과 비교수치 실험을 하였다. 실험 환경은 조용한 실험실에서 채집되었고 인식기는 문맥 종속 인식기로 구현하였다. 실험에 사용된 문맥은 고립단어 '좋아요'와 '제 이름은 김필립입니다.', '오늘밤 목을 싱글룸을 예약하고 싶습니다.'라는 2개의 문장을 사용하였다. 실험에 참여한 화자는 20대 남성 20명으로 각 문맥 당 10회 발생하여 총 600개의 음성 자료에 대해 인식 실험을 하였다. 음성 데이터는 14.4KHz 표본화율과 16-비트로 양자화 되어 저장되었다.

실험 방법은 다음과 같다. 우선 저장된 데이터로부터 12차 멜 캡스트럼 특징 벡터를 추출하였다[7]. 그 후 기존의 벡터 양자기 학습법은 K-means 알고리즘에 의해 반복 학습을 수행하여 각 화자의 코드북을 생성하였다. 이때 학습에 사용된 각 화자의 자료는 5개이고 수행된 반복 학습 회수는 각 문맥 당 평균 10~12회였다. 준비반복 벡터 양자기 학습법은 기준화자 외에는 반복학습 없이 각 화자의 코드북을 생성하였다. 그리고 모든 자료에 대해 제안된 방법과 기존의 방법을 적용하여 각각 인식 실험을 수행하였다.

<표 1>~<표 3>은 20명이 발음한 200개의 음성을 가지고 각 문맥에 대해 실험하였을 때 코드북 증가에 따른 화자 인식율을 기존 알고리즘과 비교하여 보여주고 있다. <표 1>의 코드북의 크기가 2인 경우 제안된 알고리즘은 화자의 특징을 나타내는 대표점으로서의 변별력이 떨어져 낮은 인식률을 보인다. 즉 화자의 분포에 대한 변별을 제대로 못한다고 볼 수 있다. 코드북의 크기가 4 이상인 경우 기존 알고리즘과 비슷한 인식률을 보이나 코드북 증가에 따라 인식률이 수렴하지 않는 특성을 보이는데 이는 분포를 평가하기 위한 학습 데이터가 부족하여 발생하는 현상으로 보인다. 이는 특징벡터 수가 2~3배 되는 <표 2>와 <표 3>에서는 코드북 증가에 따라 안정된 인식률이 나타나는 것에서 알 수 있다. <표 1>의 코드북의 크기가 32인 경우는

실험에서 제외하였다. 이는 전체 특징벡터는 평균 70개이고 코드북 크기는 32개로 코드북 당 평균 특징 벡터수가 작아 화자의 분포 특성을 평가하기에 부적합하기 때문이다. <표 2>와 <표 3>에서 제안된 알고리즘은 기존 알고리즘과 비교하여 코드북 크기가 증가함에 따라 안정적으로 동일한 100%의 인식률을 나타내고 있다. 이것의 의미는 화자의 특징 분포를 평가하기 위한 문맥 정보가 충분할 경우 변별력있고 안정적인 대표점을 구할 수 있다는 것을 나타낸다.

<표 1> 화자 인식 결과 비교('좋아요')

코드북 크기	2	4	8	16	32
기존 알고리즘	94%	99%	99%	99.5%	99.5%
제안된 알고리즘	77%	98%	96%	98.5%	

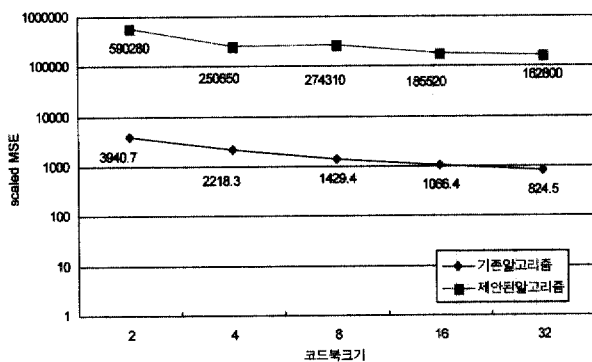
<표 2> 화자 인식 결과 비교('제 이름은 김필립 입니다.')

코드북 크기	2	4	8	16	32
기존 알고리즘	97%	99%	100%	100%	100%
제안된 알고리즘	89%	99%	99%	100%	99.5%

<표 3> 화자 인식 결과 비교('오늘밤 묵을 싱글룸을 예약하고 싶습니다.')

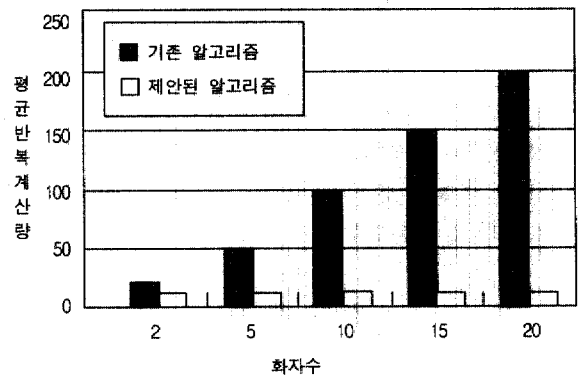
코드북 크기	2	4	8	16	32
기존 알고리즘	98%	100%	100%	100%	100%
제안된 알고리즘	91%	99.5%	100%	100%	100%

(그림 3)은 '좋아요'의 경우 코드북 증가에 따른 에러값의 누적치를 보여주고 있다. 기존 알고리즘은 각 화자의 분포에 최적화된 군집에 의해 계산되어 누적 에러값이 적은 반면 제안된 알고리즘은 최적화된 군집에 의한 누적 에러값이 아니므로 상당히 큰 에러값을 나타내는 것을 볼 수 있다. 그럼에도 인식이 가능한 것은 제안된 인식 과정에서 큰 에러값 끼리 상대 비교되기 때문이다. 또 기존 알고리즘이 코드북 크기가 증가함에 따라 에러가 감소하는 것처럼 제안된 알고리즘도 코드북 크기가 증가함에 따라 에러가 감소하는 것을 볼 수 있고 이는 반복 학습된 코드북을 이용한 새 화자의 군집화가 효과적으로 이루어진다는 것을 나타낸다.



(그림 3) 코드북 크기에 따른 누적된 scaled MSE 비교

(그림 4)는 화자의 증가에 따른 평균 반복 계산량을 비교하여 보여주고 있다. 여기서 1회 반복학습을 위해서는 코드북 크기가 n이고 특징 벡터 수가 m일 경우 $n \times m$ 의 비교 연산 후 m번의 덧셈과 n번의 곱셈을 필요로 한다. 따라서 1회 반복학습을 위한 계산량은 $n \times m$ 이라 볼 수 있다. 한 화자의 코드북이 수립되기 위한 평균 반복학습 횟수는 K-means 클러스터링에 의해 생성될 경우 각 코드북의 크기에 대해 평균 10회 정도이다. 그러므로 한 화자의 평균 반복 계산량은 $10 \times n \times m$ 이 된다. $n \times m$ 을 기본 단위로 보면 기존 알고리즘의 경우 각 화자마다 반복학습 과정을 통해 코드북이 생성되므로 평균 반복 계산량은 화자수 $\times 10$ 이 된다. 이에 반해 제안된 알고리즘은 화자 1명 $\times 10$ 이면 된다.



(그림 4) 화자수 증가에 따른 평균 반복 계산량 비교

5. 결 론

본 논문에서는 화자 인식에서 문맥 중속된 경우에 비기준 화자의 반복학습을 생략하고 기준 화자의 양자 영역을 이용하여 모든 비기준 화자의 코드북을 생성하는 새로운 벡터 양자기 설계법을 제안하였다. 또한 이론적으로 기준 화자의 양자영역 안에서 생성된 비기준 화자의 코드북은 화자의 최적 양자점이 된다는 것을 보임으로 인식기로서 성능을 나타낼 수 있다는 것을 설명하였다. 그리고 인식기의 구현을 통하여 준비반복 벡터 양자기 설계법을 시험한 결과 문맥이 중속되어 화자들의 특징이 겹쳐져 있을 경우에 각 화자의 특징 벡터집합에 최적화된 대표점을 찾는 것 없이도 각 화자의 변별력은 화자의 분포를 적절히 평가함으로써 구해줄 수 있다는 것을 확인하였다. 이 제안된 설계법의 수치 결과는 기존의 벡터 양자기 설계법과 유사한 성능을 가지면서 반복학습 계산량은 화자수 증가와 무관한 장점이 있음을 보여준다.

수치 실험 결과 적절한 코드북 크기와 화자의 분포를 평가하기에 충분한 자료를 가진 경우 인식 성능은 100%로 기존의 벡터 양자기 설계법과 같은 성능을 나타냈다. 이는 벡터 양자기의 동작 원리가 화자간 양자화 거리의 판별 기능보다는 화자의 특징 분포의 평가에 있음을 나타낸다. 이러

한 사실은 기존의 논문에서 GMM을 통한 화자인식 방법이 벡터 양자화 방법보다 강한 성능을 나타낸다는 것에서도 유추해 볼 수 있다[8].

제안된 설계법은 화자의 특징 분포를 평가하는데 한 군집 안에 평가된 각 화자의 양자점 사이의 변별력이 있는 경우 적은 계산 복잡도로 성능을 발휘한다. 물론 K-means 알고리즘과 같은 반복학습은 화자의 전체 분포에 최적화된 코드북을 생성하므로 계산 복잡도는 화자수 만큼 증가하고 그만큼 큰 변별력을 가질 수 있다. 그러나 제안된 설계법의 변별력은 적절한 코드북의 크기와 분포를 평가하기에 충분한 자료의 길이가 주어진 경우 기존의 벡터 양자화 설계법과 동일한 성능을 나타낸다. 따라서 준비반복 벡터 양자화 설계법은 적절한 규모의 화자 인식의 경우 적은 계산 능력을 필요로 하는 인식기의 구현에 도움이 될 것으로 사료된다.

앞으로의 연구 과제는 화자 수 증가에 따른 변별력 강화를 위한 연구와 문맥 독립된 경우에 대하여 확장하여 화자의 분포를 적절히 평가해 내면서 계산량을 줄일 수 있는 방법에 대한 연구가 진행 되어야 할 것이다.

참 고 문 헌

[1] T. Kinnunen, T. Kilpelinen and P. Frnti, "Comparison of clustering algorithms in speaker identification," Proc. IAS-TED Int. Conf. Signal Processing and Communications (SPC 2000), Marbella, Spain, pp.222-227, 2000.

[2] H. Gish and M. Schmidt, "Text-independent speaker identification," IEEE Signal Processing Mag., Vol.11, p.1832, 1994.

[3] Y. Linde, A. Buzo and Gray R. M., "An algorithm for vector quantizer design," IEEE Trans. On Communications, 28(1), pp.84-95, January, 1980.

[4] A. K. Jain, R. P. W. Duin and J. Mao, "Statistical pattern recognition : A review," IEEE Trans. Pattern Anal. Machine Intell., Vol.22, p.437, Jan., 2000.

[5] D. A Reynolds, "An overview of automatic speaker recog-

inition technology," Acoustics, Speech, and Signal Processing, 2002 IEEE International Conference on, Vol.4, pp.4072-4075, 2002.

[6] T. Kinnunen, I. Krikkinen and P. Frnti, "Is speech data clustered? - Statistical analysis of cepstral features," Proc. 7th European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg, Denmark, Vol.4, pp.2627-2630, 2001.

[7] S. Furui, "Cepstral analysis technique for automatic speaker verification.," IEEE Trans. on Acoustics, Speech and Signal Processing, 29(2), pp.254-272, 1981.

[8] D. A Reynolds and Robust, "Text-independent speaker identification using gaussian mixture speaker models," IEEE Transactions on speech and audio processing, Vol.3, No.1, January, 1995.



임 동 철

e-mail : thedove@hanmail.net

1996년 아주대학교 전자공학(학사)

1996년 LG 반도체 근무

1999년 아주대학교 전자공학(공학석사)

1999년~현재 아주대학교 전자공학부

박사 수료

관심분야 : 음성 및 영상 신호처리, 패턴인식



이 행 세

e-mail : haingsei@madang.ajou.ac.kr

1966년 전북대학교 전기공학과(학사)

1972년 서울대학교 전자공학과(공학석사)

1984년 고려대학교 전자공학과(공학박사)

1968년~1970년 해군사관학교 교관

1982년~1983년 Columbia Univ.n.y 객원

교수

1987년~1988년 INRIA PARIS 객원교수

1992년~1994년 거제전문대학장

1973년~현재 아주대학교 교수

관심분야 : 음성인식, 인공지능 및 신경망