FVQ(Fuzzy Vector Quantization) 사상화에 의한 화자적응 음성합성

Speaker-Adaptive Speech Synthesis by Fuzzy Vector Quantization Mapping

이 진 이*, 이 광 형*

Jin-Yi Lee*, Kwang-Hyung Lee*

요 약

본 연구에서는 퍼지사상화(fuzzy mapping)에 의한 사상된(mapped) 코드북을 사용하는 화자적응 음성합성 알고리즘을 제안한다. 입력화자와 기준화자의 코드북은 신경망 클러스터링 알고리즘인 자율경쟁 학습을 사용하여 작성된다. 사상된 코드북은 입력 음성벡터에 대한 두 화자의 대응 코드벡터의 소속값(membership value)으로 퍼지 히스토그램을 작성하여 이들을 1 차 결합함으로써 얻어지는 퍼지사상화에 의하여 작성된다. 음성합성시에는 사상된 코드북을 사용하여 입력화자의 음성을 퍼지 벡터양자화한다음, FCM 연산으로 합성함으로써 입력화자에 적응된 합성음을 얻는다. 실험에서 여러 입력화자로 30대의 남성, 20대의 남성, 20대의 여성음성을 사용하였고 기준음성으로 입력음성과는 다른 20대의 여성음성을 사용하였다. 실험에 사용된 음성데이타는 문장 /안녕하십니까/ 와 /굿모닝/이다. 실험결과 각각의 입력화자에 기준화자 음성이 적응된 합성음을 얻었다.

ABSTRACT

In this paper, we propose a speaker-adaptive speech synthesis method using mapped codebooks designed by fuzzy mapping. Competitve learning neural networks is used to design input speaker's codebook and referer.ce speaker's codebook. We used fuzzy mapping to design mapped codebooks. Fuzzy mapping replaces a codevector while preserving its fuzzy membership function. The codevector correspondence histogram is obtained from accumulating the vector correspondence along the DTW optimal path. Using each histogram as a weighting function, the mapped codebook is defined as a linear combination of reference speaker's vectors.

Speech synthesis is perform as follows : Input speaker's speech is fuzzy vector quantization using mapped codebook, and then FCM arthimetic is used to synthesis input speaker-adaptive speech. The speaker adaption experiments are carried out using speech of a male in his thirties, speech of a male in his twenties, and speech of a female in her twenties, as input speaker's speech, and speech of a other female in her twenties, as reference speaker's speech. Speech used in experiments are sentences /anyoung hasim nika/ and /good morning/. As a results of experiments, we have listened speaker-adaptive speech.

* 숭실대학교 전자공학과

* Department of Electronics Engineering, Soong Sil University

Ⅰ. 서 론

일상의 통신에서 개인 음성은 가장 중요한 정보이다. 그것은 전화선을 통하여 대화를 할때 화자를 판명하는데 특히 중요하다. 그러므로 개인음성을 조절하는 것은 중요한 역할을 하고 많은 응용분야가 있다. 이 연구의 목적은 미지화자로부터의 음질을 다른화자에 적응하는 것과 합성음에 개인음성을 부여 할 수 있는 기술을 개발하는데 있다.

음성인식의 HMM 기술은 거대한 양의 트레이닝 데이타를 필요로 하기 때문에 미지의 입력화자에 대한 인식에는 어려움이 따른다. 화자독립 트레이닝으로 가능하기는 하지만 인식율이 크게 떨어진다. 그래서 적은 단어나 문장으로 인식시스템을 트레이닝할 수 있는 화자적응기술이 사용된다. 음성인식분야에 퍼지 벡터양자화[1]의 사상화 소속함수를 사용하여 화자적응을 행함으로써 인식율을 높히고 있다[2].

음성합성 분야에서 퍼지 벡터양자화를 도입하여 합성음질을 향상시키기도 했다[3]. 그러나 본 연구에서는 지금까지의 의미전달 음성합성 방식에서 한 걸음 더 나아가 화자에 따른 질의-응답 음성의 친밀성을 높힐 수 있는 화자적응 음성합성방법을 제안한다.

음성의 개인성은 일반적으로 두개의 주요한 요소로 구성된다. 하나는 음향학적인 특징이고 다른하나는 운율적인 특징이다. 이 연구는 첫째 단계로 음향적인 특징을 조절하여 합성음질을 높힌다. 지금까지의 연구에 따르면 개인음성에 기여하는 음향적인 특징들은 포르만트 주파수, 포르만트 대역폭, 스펙트럼의 기울기, 성문의 파형들과 같은 여러가지 파라메타들의 복합으로 이루어져있다[4].

이 논문에서는 음향적인 파라메타를 분리하지 않고 음성을 합성하는 것으로, 코드북은 변화하는 음향적인 특징들에 관하여 개인음성에 대한 모든 정보를 포함하고 있다. 코드북은 대표벡터의 집단으로 VQ 의 성능을 좌우한다. 코드북 작성에는 k-means, LBG 알고리즘이 대표적이며, 최근에는 부 공간사이의 소속정도를 고려한 퍼지 클러스터링 알고리즘으로 최적의 코드북을 작성하고있다[5]. 이상의 방식은 많은 계산으로 인해 입력 데이타 벡터의 고유한 성질을 잃어버릴 수 있는 위험 부담을 갖고 있다. 이러한 문제점을 해결할 수 있는 또 다른 코드북 작성 기술이 인공 신경망을 이용한 학습 벡터 양자화(Learning VQ)[6]이다. 이 학습 벡터 양자화는 간단한 학습 알고리즘에 의해 코드북을 작성하는 것으로써 기존의 적응 벡터 양자화보다도 월등히 적은 계산양으로도 데이타의 본래 성질을 우수하게 보존하면서 데이타 압축이 가능한 기술이다. 본 연구에서는 이러한 신경망의 학습기능에 의한 코드북을 사용한다.

미지화자로 부터의 음향적인 특징들을 다른 화자로의 변환은 두 화자의 코드북 사상화에 관한 문제로 축소된다. 화자적응을 위한 코드북 사상화는 서로 다른 화자의 음성파라메타 공간사이의 사상화함수를 결정해야 하는데 매우 복잡한 비선형함수가된다[7]. 여기에는 퍼지벡터 양자화, 퍼지사상화와 퍼지히스토그램 근사화 알고리즘들이 코드북 사상화를 향상시키기 위해 사용된다. 다음 장에서는 화자적응밍 음성합성을 위한 코드북 작성을 기술하고 3 장에서는 퍼지-VQ 사상화알고리즘을 기술하고 4 장에서는 화자적응실험을 하였고 5장에서 결론을 맺는다.

Ⅱ. 화자적응과 신경망 코드북

본 연구에서는 표준 데이타베이스의 기준음성으로부터 여러화자의 음성에 적응된 음성을 정확히 합성해내기 위해 화자적응을 도입한다. 화자적응에 쓰이는 코드북작성 방법에는 K-means, LBG, 퍼지클러스터링, 신경망-VQ 등이 있다.

2.1 화자적응

화자적응은 기준화자의 코드북의 코드벡터를 입력화자의 코드북의 코드벡터로 변환함으로써 이루어지며 그 반대도 가능하다. 화자적응은 두 화자 사이의 스펙트럼의 차이를 줄이기 위해 코드북에 있는 벡터들을 다른 코드북에 있는 벡터들로 대치함으로써 수행된다. VQ는 주어진 하나의 화자에 대한 음성특징을 정확하게 근사화할 수 있지만, 여러 화자에 대해서는 문제를 갖고 있다.

화자적응 알고리즘중 하나는 학습에 의한 화자적응으로, 이것은 두 코드북의 코드벡터 사이의 대응벡터들을 찾기 위해 학습단어들을 필요로 한다. 다른 하나의 화자적응 알고리즘은 어떠한 코드북내의 벡터들의 분포는 비슷하다라는 가정에 기초하여 두개의 VQ 코드북내의 대응벡터들을 찾는다. 벡터들의 1 대 1 대응은 두 코드북 사이의 벡터 거리들의 합을 최소로 하는것으로 결정된다. 이러한 화자적응을 음성인식에 적용하여 인식율을 높힌 연구결과를 보면 화자적응이 없이, 서로 다른화자들 사이의 평균단어 인식율은 단지 64.0% 이였고, 학습에 의한 화자적응을 함으로써 83.1% 로 향상되었으며, 벡터분포에 기초한 화자적응을 함으로써 그 비율은 95.4%이었다[2][8].

본 연구는 서로 다른 화자에서 얻은 두개의 신경망 코드북 사이의 대응을 찾는 것에 기초한다. 코드북의 대응은 두 코드북내의 벡터들 사이의 거리에 의해 구해지며 일단 두 코드북 사이의 대응이 정확하게 찾아지면 정규화한다. 사상된 코드북은 기준화자의 벡터들을 1 차 결합 함으로써 얻어진다. 이때의 각 계수값들은 누적 소속값으로 주어진다. 화자적응에는 모음의 선형변환, 통계적 사상화, 스펙트럼사상화가 있다[8]. 본 연구에서는 퍼지-VQ 사상화를 사용한다.

2.2 코드북 작성

VQ-코드북 설계를 위해 신경망 기술을 사용한 대표적인 연구를 들면 다음과 같다. 음성에 대해서는 Kohonen [9], Naylor 과 Li[10]를 언급할 수 있는데, 이들은 코호넨의 자기 조직화 형상지도를 사용하였다. Nasrabadi 와 Feng[11] 는 영상 VQ를 위한 코드북을 작성하는데 신경망을 사용하였다. 음성과 영상 모두를 가변 영역 벡터 양자화하는 연구는 Matsuyama[12]에 의해 처음 시도되었다.

벡터양자화하려는 여러화자의 입력음성 벡터는 k차원의 벡터 공간에 있는 것으로 생각하고 왜곡 측정 $d(X,Y_j)$는 이 공간에서 정의된 것으로 한다. 코드북 크기를 L 로 하여 코드벡터는 Y_j, $j=1, \ldots, L$ 개로 생각한다. 이것들을 L 개의 신경 단위로 보고 이들로 구성된 신경망을 고려한다. 그리고 j 번째 코드벡터 Y_j 를 신경망 단위 j 의 하중벡터로 생각한다. 어떤 벡터 X 가 주어졌다면, X 는 모든 L 개의 신경단위에 일괄적으로 입력되며 이들 각각의 신경 단위들은 입력 벡터와 자신의 하중벡터간의 왜곡을 계산한다.

$$d_j = d(X, Y_j), \quad j=1, \ldots, L \tag{1}$$

최소 왜곡을 갖는 신경 단위 j^* 를 선택함으로써 입력 벡터는 j^* 의 인덱스로써 양자화된다.

$$j^* = \min_j \{d_j\} \tag{2}$$

일반적으로 이러한 "최근접 이웃" 양자화절차는 최적이다[13]. 승자(winner) 신경단위를 선택하고 그것의 인덱스를 결정하는 것을 제외하고는 모든 계산들이 똑같은 형태로 이루어지는 것이 특징이다. 신경망에서 승자를 선택하는데 많은 방법이 있는데, 대표적으로는 Grossberg's on-center,off-surround 방법[14][15], Lippman et al's MAXNET[16], 그리고 Winters 와 Rose의 최소거리 오토마타[17]들이 있다.

특히 Hecht-Nielson[18]는 신경망에 출력층을 하나 더 부가함으로써 승자신경 단위의 인덱스 번호를 직접적으로 계산하는 것이 가능하도록 하였다.

2.3 경쟁 학습망

벡터 양자화와 관련된 트레이닝 알고리즘으로는 승자만 학습시키는 경쟁 학습 알고리즘(Competieive learning(CL) network), 승자뿐만 아니라 그 이웃 뉴런의 하중(weight)까지 함께 학습시키는 코오넨의 자기 조직화 형상지도(Kohenen self-organizing feature map (KSFM))알고리즘, 그리고 각 뉴런이 승자가 되는 빈도를 계산하여 어느 특정 뉴런만 승자가 되는 것을 방지 하고 모든 뉴런이 거의 같은 횟수로 승자가 되게함으로써 부호화시 엔트로피를 최대로 하는 빈도-민감 경쟁 학습망(frequency-sensitive competitive Learning (FSCL) net-

work) 알고리즘이 있다[19][20].

본 연구에서는 경쟁 학습망 알고리즘을 이용하여 음성 코드북을 작성한다. 코호넨의 자기 조직화 형상 알고리즘은 이웃 반경까지 정의하여 승자를 함께 학습시키므로써 수반되는 엄청난 계산량의 증가가 문제점이 된다. 특히 트레이닝 회수가 클 경우에는 경쟁학습망의 성능과 거의 같음을 고려하였다[19].

그림 1 은 4차원 32 레벨의 코드북을 갖는 경쟁학습망의 구조를 나타낸다. 입력층에서는 입력 음성 벡터를 받아들인다. 각각의 입력음성벡터는 4 개의 음성샘플로 구성한다. 은닉층에서는 입력 벡터와 입력층의 뉴런과 은닉층의 뉴런사이의 하중값과의 거리를 계산한다. 하중벡터는 다음과 같다.

$$W=\{w_{ij}\}\ i=1, \ldots, 4,\ j=1, \ldots, 32 \tag{3}$$

출력층에서는 중간층에서 계산한 거리를 비교하여 최소거리의 뉴런을 택하여 승자로 결정한다. 이 승자 뉴런만 학습하게 되고 학습이 종료되면 입력층과 은닉층의 연결강도가 코드벡터를 형성하여 4차원 32 레벨의 코드북이 작성된다.

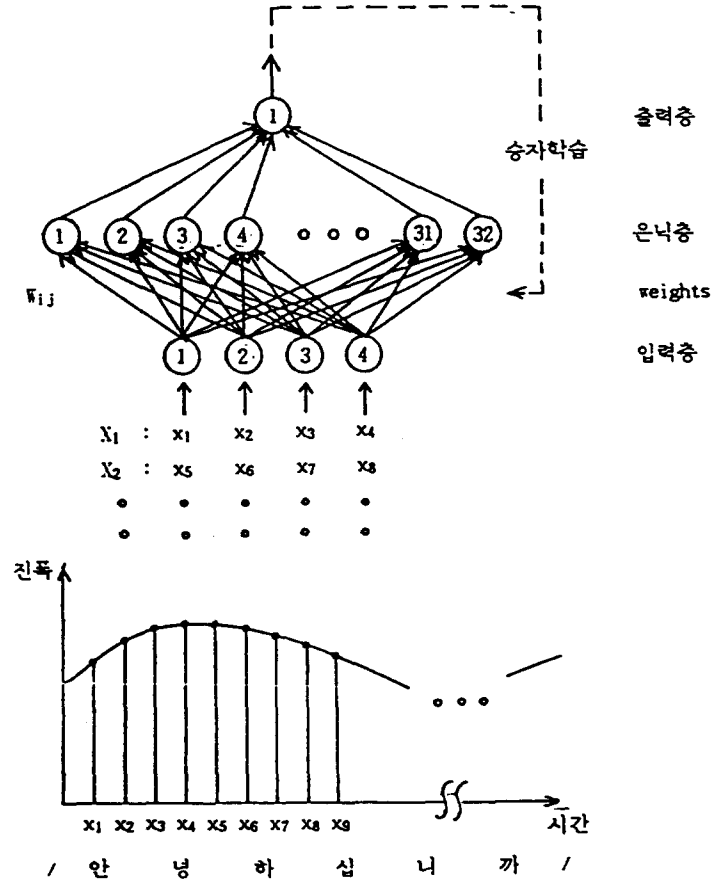


그림 1. 경쟁학습망 (4 차원 32 레벨 코드북)

Fig. 1. Competitive learning network (4 dimension 32 level codebook)

경쟁 학습 알고리즘을 그림 2 에 나타낸다. 32개의 신경 단위들은 하중 벡터 $W_i(0)$, $i=1, \ldots, 32$ 로 초기화 한다. 이들 하중값들을 랜덤하게 초기화할 수 도 있고 트레이닝 데이타 집합의 첫 번째 32 개 벡터들로 할 수 도 있다. 하중벡터를 조정하기 위해서 사용된 알고리즘은 경쟁에 의한 학습에 근거한다. 하나의 입력벡터 X 에 대해 하중벡터를 갱신하기 위한 알고리즘은 아래와 같다.

$$Z_i = \begin{cases} 1, & d(X, W_i(n)) \le d(X, W_j(n)),\ j \neq i,\ 1 \le j \le 32 \\ 0, & o.w \end{cases} \tag{4}$$

새로운 하중벡터 $W_i(n+1)$은 다음으로 계산된다.

$$W_i(n+1) = W_i(n) + \varepsilon(n)\ (X - W_i(n))\ Z_i \tag{5}$$

위의 식에서, 학습률 $\varepsilon(n)$는 벡터 패턴들이 클러스터 중심으로 접근함에 따라 시간에 따라 감소하는 변화량으로 일반적으로 다음과 같은 형태를 취한다.

$$\varepsilon(n) = A\, e^{(-n/T)} \tag{6}$$

여기서 A 는 최대 변화량을 결정하는 상수이다. T 는 각 입력이 제시된 후 감소하는 각 패턴 벡터의 이웃 벡터들에 대한 각 벡터 패턴의 영향을 조절한다. 사실상 그것은 입력 벡터공간을 부공간으로 분할 하거나 줄이는 역할을 한다.

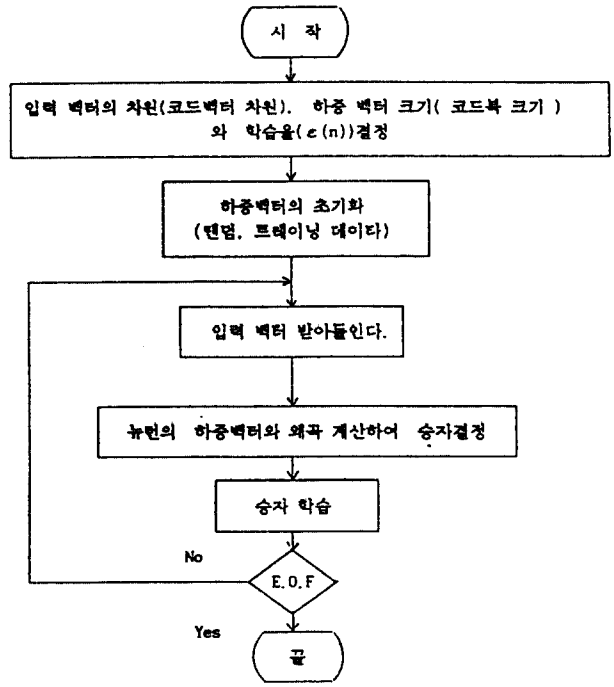


그림 2. 경쟁 학습망 알고리즘
Fig. 2. Competitive learning network algorithm

Ⅲ. 퍼지 벡터양자화를 이용한 화자 적응 음성합성

화자적응 음성합성은 두 단계로 구성된다. 훈련 단계와 적응합성 단계이다. 훈련단계는 퍼지 사상된 코드북을 생성하는 단계이다. 그리고 적응음성합성 단계는 퍼지 사상된 코드북을 사용하여 음성을 합성하는 과정이다. 퍼지 벡터양자화는 하나의 입력 벡터를 VQ-코드벡터들의 하중 선형 결합으로 나타내기때문에 벡터양자화보다는 합성음의 왜곡을 크게 줄일 수 있다.

3.1 퍼지 VQ 사상된 코드북 작성

퍼지 VQ 사상된 코드북(甲男 → 乙女)은 두 화자의 코드벡터 사이의 소속값(membership value) 의 누적 히스토그램을 계수값으로 기준화자의 코드벡터를 1 차 결합함으로써 얻어지며, 이 과정을 퍼지사상화라한다. 그림 3 은 퍼지 VQ와 퍼지사상을 보여준다. 편의상 입력화자의 코드북과 사상된 코드북은 4 차원 4레벨의 크기로 나타낸 것이다. 퍼지-VQ는 입력벡터가 각각의 코드벡터에 소속되는 정도의 소속값을 성분으로 하는 하나의 백터를

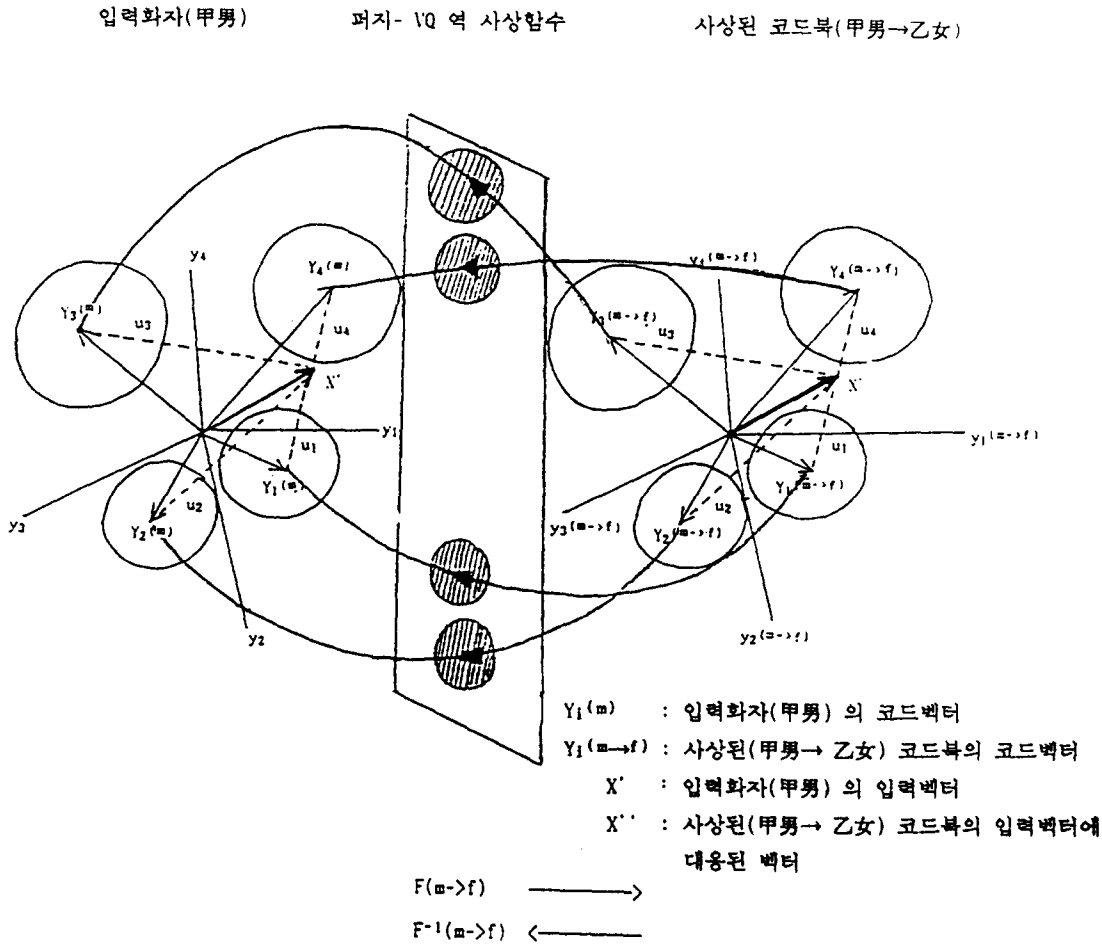


그림 3. 퍼지-VQ 역 사상화함수($F^{-1}(m \rightarrow f)$)에 의한 화자적응

Fig. 3. Speaker adaptation using inverse Fuzzy VQ Mapping Function ($F^{-1}(m \rightarrow f)$)

생성한다. 입력화자의 벡터공간과 기준화자 벡터공간사이의 퍼지 VQ-사상화는 퍼지 VQ-사상화 함수 $F(m \rightarrow f)$에 의해 이루어지고 이 함수에 의해 입력화자의 코드벡터가 기준화자의 코드벡터에 사상된 코드북이 작성된다. 이 사상된 코드북을 구성하는 코드벡터는 다음 (7) 식으로 구한다.

$$Y_i^{(m \rightarrow f)} = \frac{\sum_{j=1}^{4} h_{ij} Y_j^{(f)}}{\sum_{j=1}^{4} h_{ij}} \quad . \quad i = 1, \ldots, 4 \tag{7}$$

여기서 i, j 는 각각 입력밍 기준화자의 코드벡터의 순번을 나타낸다.

퍼지-VQ 역 사상화 함수 $F^{-1}(m \rightarrow f)$ 에 의한 퍼지-VQ 역 사상화는 다음과 같이 나타내며 입력화자에 대한 정보를 알려준다.

$$Y_i^{(m \rightarrow f)} \xrightarrow{F^{-1}(m \rightarrow f)} Y_i^{(m)} \tag{8}$$

여기서 $F^{-1}(m \rightarrow f)$은 퍼지-VQ 역 사상화함수(Fuzzy-VQ inverse mapping function)를 나타낸다.

그림 3 은 퍼지-VQ 와 사상된 코드북을 사용하여 퍼지-VQ 역 사상화 함수에 의한 화자적응을 나타낸다.

대응 코드벡터의 퍼지 히스토그램은 동일 입력벡터에 대해 선택된 입력화자의 코드벡터와 기준화자의 코드벡터사이의 DTW(Dynamic Time Warping)을 행하여 대응하는 코드벡터들의 소속값을 누적하여 얻는다. 이 값을 대응확률로 간주함으로써 퍼지 히스토그램이 비퍼지 히스토그램보다 정확하게 근사화 된다.

그림 4 는 사상된 코드북 생성과정을 나타낸다.

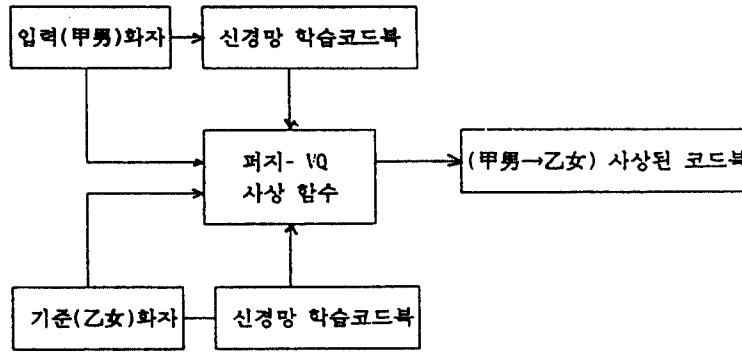


그림 4. 사상된 코드북 생성
Fig. 4. Mapped codebook generation

사상된 코드북을 얻기 위한 훈련절차(training procedure)는 다음과 같다.

단계 1) 입력화자와 기준화자의 VQ 코드북을 신경망 학습알고리즘을 사용하여 작성한다.

단계 2) 각각의 입력화자의 입력 벡터를 퍼지 벡터양자화한다.

단계 3) 퍼지벡터 양자화에 의한 대응 코드벡터를 찾는다.

단계 4) 대응코드벡터의 소속값을 누적하여 퍼지 히스토그램을 작성한다.

$$h_{ij} = h_{ij} + u_{ij} \tag{9}$$

여기서 i 는 입력화자 코드북의 코드벡터 순번, j 는 기준화자 코드북의 코드벡터 순번이다. 한편 비퍼지 VQ-사상에서는 대응벡터가 있을때 마다 1 을 누적시켜 히스토그램을 얻는다. 입력화자의 i 번째 코드벡터가 기준화자 j 번째 코드벡터에 소속된 정도의 소속값 : u_{ij} 는 (10)식과 같다.

$$u_{ij} = \frac{\left[\dfrac{1}{d(Y_i^{(m)}, Y_j^{(f)})}\right]^{\frac{1}{F-1}}}{\sum_{i=1}^{L}\left[\dfrac{1}{d(Y_i^{(m)}, Y_j^{(f)})}\right]^{\frac{1}{F-1}}} \tag{10}$$

여기서 F = 애매성 상수
L = 코드북 크기(size, level)
$Y_i^{(m)}$: 입력화자(남성)의 i 번째 코드벡터
$Y_i^{(f)}$: 기준화자(여성)의 j 번째 코드벡터
$Y_i^{(m\to f)}$: 사상된 코드북의 j 번째 코드벡터
$1 \le i \le L, \quad 1 \le j \le L$

단계 5) 다음 식을 사용하여 사상된 코드북을 작성한다.

$$Y_i^{(m)} \longrightarrow Y_i^{(m\to f)} = \frac{\sum_{j=1}^{L} h_{ij}\, Y_j^{(f)}}{\sum_{j=1}^{L} h_{ij}} \tag{11}$$

여기서 h_{ij} 는 퍼지히스토그램값을 나타낸다.

단계 6) 입력화자의 코드북을 사상된 코드북으로 대치한다.

단계 7) 모든 입력벡터에 대한 평균왜곡을 계산하여 그 값이 수렴하면 사상된코드북의 훈련을 끝내고, 그렇지 않으면 계속해서 훈련하여 정교한 사상된 코드북을 만든다.

3.2 적응 음성분석-합성

사상된 코드북을 사용하여 입력화자의 음성을 퍼지벡터 양자화하면 입력 벡터와 각 코드벡터와의 소속값을 성분으로 하는 퍼지 사상된 소속출력함수(Fuzzy Mapped Membership Output Function) O_i 를 얻는다.

$$O_i = [\, u_{i1}\, u_{i2} \ldots\ u_{iL} \,]^T \tag{12}$$

이들 출력벡터 O_i 성분들은 양의 값이고 모두 합하면 1 이 된다. 그리고 F(Fuzziness)>1 는 확률적 변화양을 나타내는 애매성 상수이다. 퍼지사상된 소속함수 O_i (12) 식은 다음 식의 퍼지 목적함수를 최소화하여 구한다.

$$\min Z(u, Y^{(m\to f)}) = \sum_{i=1}^{M} \sum_{j=1}^{L} u_{ij}^{F}\, d(X_i, Y_j) \tag{13}$$

여기서 n 은 입력음성벡터의 순번크기이다.

목적함수 (13)식에서 Y_j 를 고정시키고 u_{ij} 에 대해 편미분을 행하고, 영으로 놓아서 식 (10)을 구한다.

애매성 상수값 F 가 무한히 커짐에 따라서 O_i 의 각 성분들은 1/L 에 근접한 값들을 갖게 되고 F 가 1 에 가까운 값을 갖을 수록 어느 하나의 성분만 1 에 가까운 값을 갖게 되어 다른 모든 성분들은 0 의 값으로 접근한다. 그래서 FVQ 판정을 단순판정 ($F\to 1$)으로 할것인가, 아니면 매우 큰 퍼지판정 ($F\to\infty$)으로 행할 것인가 하는 것은 F 값의 적절한 선택에 달려 있다.

그림 5 는 사상된 코드북을 사용한 퍼지-VQ 에 의한 음성 분석-합성을 나타낸다.

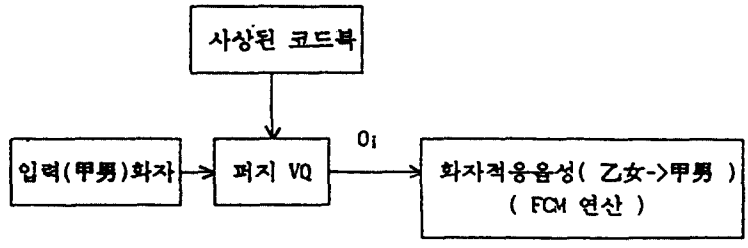


그림 5. 퍼지 VQ 에 의한 화자적응 음성분석-합성
Fig. 5. Speaker-adaptive speech analysis-synthesis by fuzzy-VQ

FVQ 음성합성은 VQ 의 패턴 매칭기법에 의한 합성과는 다르게 분석단에서 얻은 O_i 벡터와 FCM 합성규칙을 사용하여 사상된 코드북 내의 코드 벡터가 아닌 새로운 하나의 합성 벡터를 얻게 되어 좀더 입력화자에 적응된 합성음을 얻게 된다.

적응합성음성벡터 $\hat{X}_i$ 는

$$\hat{X}_i = [\,\hat{x}_{i1}, \hat{x}_{i2}, \hat{x}_{i3}, \cdots \hat{x}_{iN}\,]^T \tag{14}$$

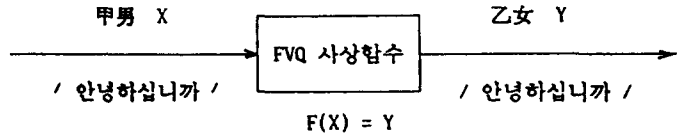
$\hat{X}_i$의 성분 $\hat{x}_{ij}$는 다음 FCM 연산으로 구한다[22].

$$\hat{x}_{ij} = \frac{\sum_{j=1}^{L} (u_{ij}^{F} \cdot Y_{ji}^{(m\to f)})}{\sum_{j=1}^{L} u_{ij}^{F}} \quad (1 \le i \le M) \tag{15}$$

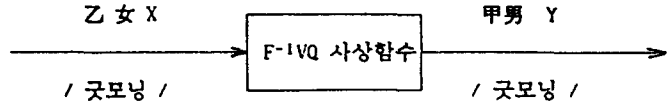
그림 6 은 화자적응 음성분석-합성을 나타낸다.

화자적응 음성 분석단에서는 甲男 의 입력 음성 /안녕하십니까/ 를 데이타 베이스의 기준화자 乙女 의 음성벡터 /안녕하십니까/ 로 퍼지사상화하는 퍼지-VQ 사상화함수 F 를 구한다.

합성단에서는 분석단에서 얻은 퍼지-VQ 사상화함수 F 의 퍼지-VQ 역 사상화함수 F^{-1} 를 구하여 데이타베이스의 기준화자인 乙女 의 음성 /굿모닝/ 에서 입력화자인 甲男 의 음성 /굿모닝/ 을 얻는다.



(a) 분 석 단



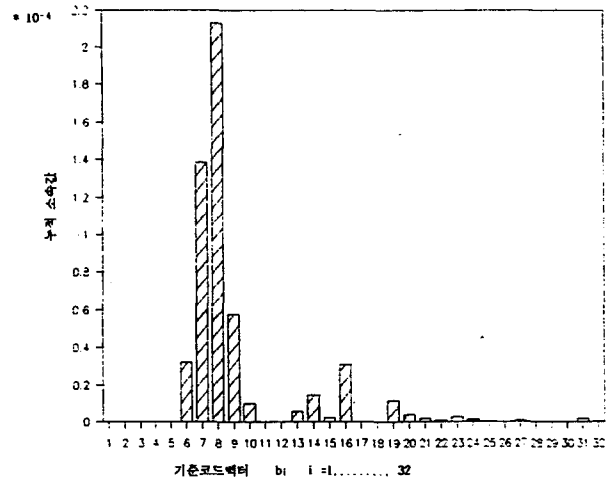
(b) 합 성 단

그림 6. 화자적응 음성 분석-합성
Fig. 6. Speaker-adaptive speech analysis-synthesis

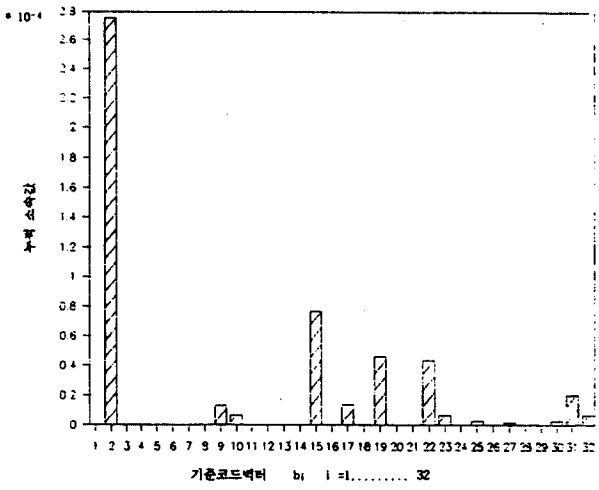
Ⅳ. 화자적응 실험

실험에서 여러 입력화자로 30대의 남성(男 1), 20대의 남성(男 2), 20대의 여성(女 1)음성을 사용하였고 기준음성으로 입력음성과는 다른 20대의 여성(女 2)음성을 사용하였다. 실험에 사용된 음성데이타는 문장 /안녕하십니까/ 와 /굿모닝/이다. 음성데이타는 8 [KHz]로 샘플링되었고, 코드북 크기는 모두 4 차원의 32 레벨을 갖는다.

입력화자와 기준화자의 학습코드북을 훈련하여 그 벡터 대응을 찾는다. 그 다음 퍼지 VQ 에 의한 소속값의 누적으로 퍼지히스토그램을 구하여 기준화자의 벡터를 1 차 결합함으로써 사상된 코드북을 작성하여 합성한다. 그림 7 은 /안녕하십니까/ 음성에 대한 입력화자(男 1)와 기준화자(女 2) 사이의 퍼지 VQ-히스토그램을 나타낸것이다. 대표적으로 입력화자의 음성벡터 a7, a10에 대한 것을 나타내고 있다.



(a) 입력벡터 X_7



(b) 입력벡터X_{10}

그림 7. 퍼지 히스토그램 (a) 입력벡터 X_7 (b) 입력벡터X_{10}
Fig. 7. Fuzzy histogram : (a) input vector X_7 (b) input vector X_{10}

그림 8은 합성시스템의 구조를 나타낸다.

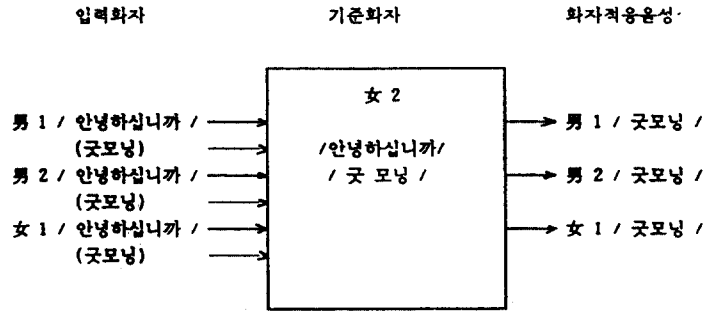


그림 8. 화자적응 음성합성 시스템
Fig. 8. Speaker-adaptive speech synthesis system

여러 화자의 FVQ-사상된 코드북을 작성하기 위해 입력화자와 기준화자의 코드북은 신경망 학습 알고리즘을 사용하여 작성하였다. 정확히 사상된 코드북을 작성하기 위하여 상호화자간의 왜곡치를 고려하여 사상된 코드북을 트레이닝하였다. 입력화자 男 1, 男 2, 女 1 의 음성 /안녕하십니까/ 가 입력되면 기준화자 女 2 의 /안녕하십니까/ 로 퍼지 사상하는 퍼지-VQ 사상함수를 구하여 놓고, 그 함수의 퍼지-VQ 역 사상함수를 계산한다. 그 다음 퍼지-VQ 역 사상함수를 이용하여 기준화자 女 2 의 /굿모닝/ 에서 입력화자 男 1, 男 2, 女 1 의 /굿모닝/ 으로 변환된 음성을 얻는다.

그림 9는 화자적응된 음성을 얻는 퍼지 VQ-사상관계를 나타낸것이다.

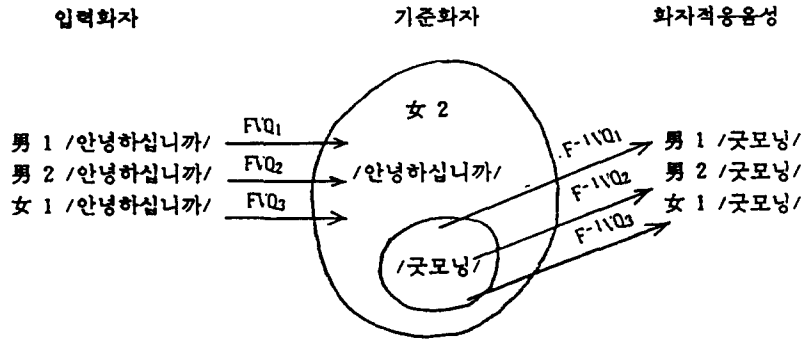


그림 9. 퍼지-VQ 사상화함수에 의한 화자적응 음성합성
Fig. 9. Speaker-adaptive speech synthesis using Fuzzy-VQ Mapping Function

각각의 화자에 대한 /안녕하십니까/의 적응파라메타(퍼지히스토그램)를 구하여 사상된 코드북을 작성한 다음 화자적응 음성합성한다.

실험에서는 편의상 男 1 의 /안녕하십니까/ 를 입력음성으로 하고 이 음성에 적응된 男 1 의 /굿모닝/ 을 얻는 화자적응을 행하여 그 결과를 보였다.

그림 10(a), (b) 는 각각 입력화자 男 1, 기준화자 女 2 의 /안녕하십니까/ 원 음성파형이다.

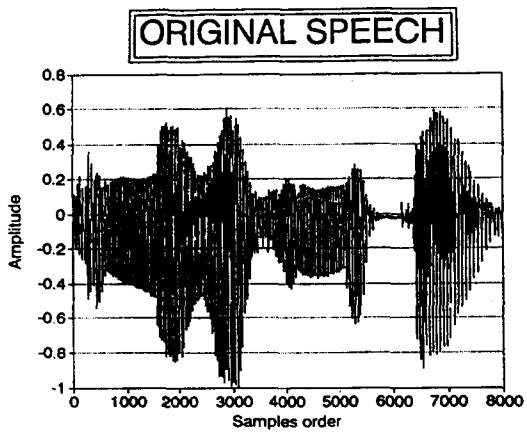


그림 10(a). 입력화자 男 1 /안녕하십니까/ 원 음성파형
Fig. 10(a). Original speech /anyoung hasim nika/ of male 1

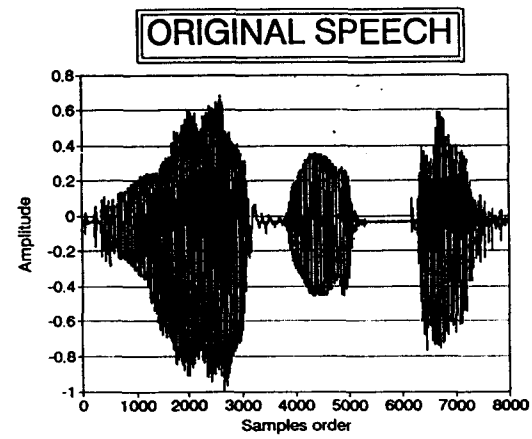


그림 10(b). 기준화자 女 2 /안녕하십니까/ 원 음성파형
Fig. 10(b). Original speech /anyoung hasim nika/ of reference speaker female 2.

그림 11(a), (b) 는 男 1, 女 2 의 /안녕하십니까/ 의 스펙트로그램이다.

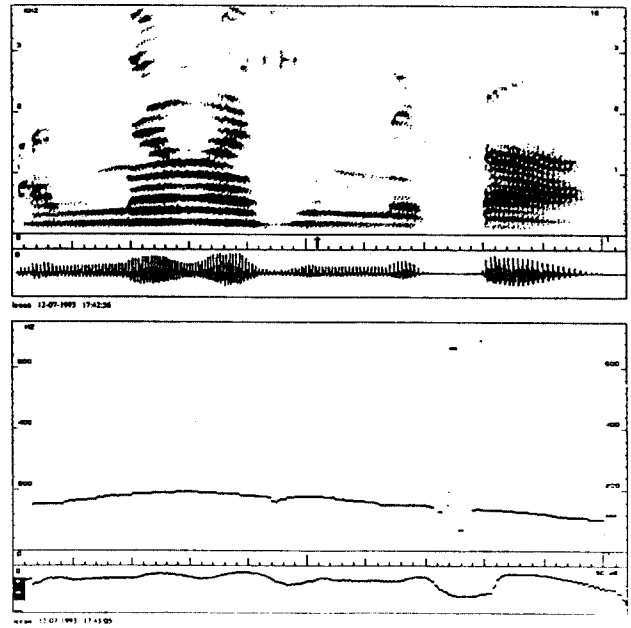


그림 11(a). 男 1 의 /안녕하십니까/ 의 스펙트로그램
Fig. 11(a). Spectrogram of original speech of male 1 /anyoung hasim nika/.

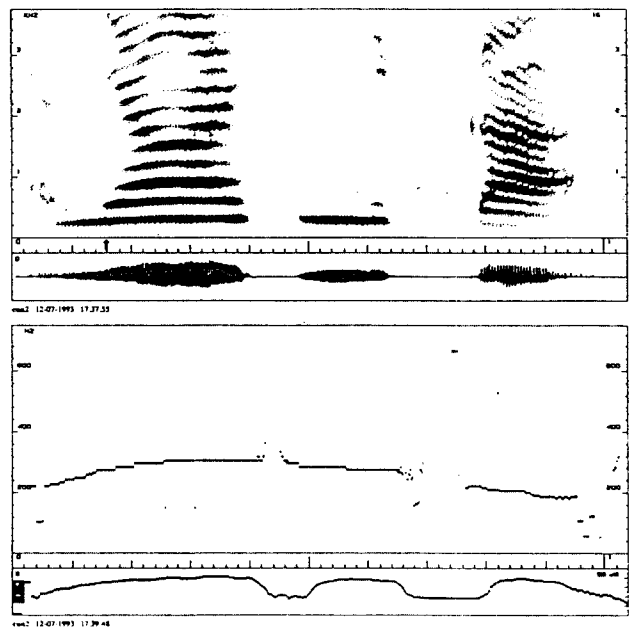


그림 11(b). 女 2 의 /안녕하십니까/ 의 스펙트로그램
Fig. 11(b). Spectrogram of original speech of female 2 /anyoung hasim nika/.

그림 12(a), (b) 는 각각 男 1, 女 2 의 /굿모닝/ 의 원 음성파형이다.

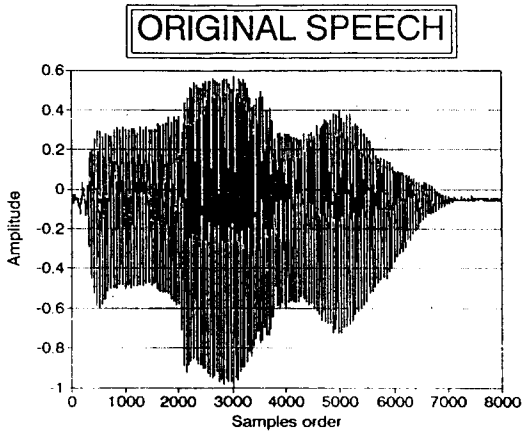


그림 12(a). 男 1 의 /굿모닝/ 의 음성파형
Fig. 12(a). Original speech /good morning/ of male 1.

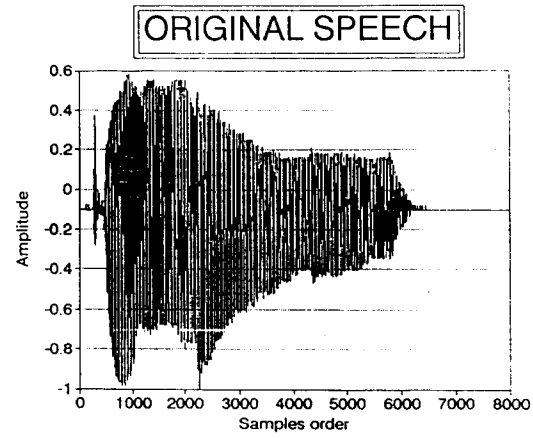


그림 12(b). 女 2 의 /굿모닝/ 의 원 음성파형
Fig. 12(b). Original speech /good morning/ of female 2.

그림 13(a), (b) 는 각각 男 1, 女 2 의 /굿모닝/ 의 스펙트로그램이다.

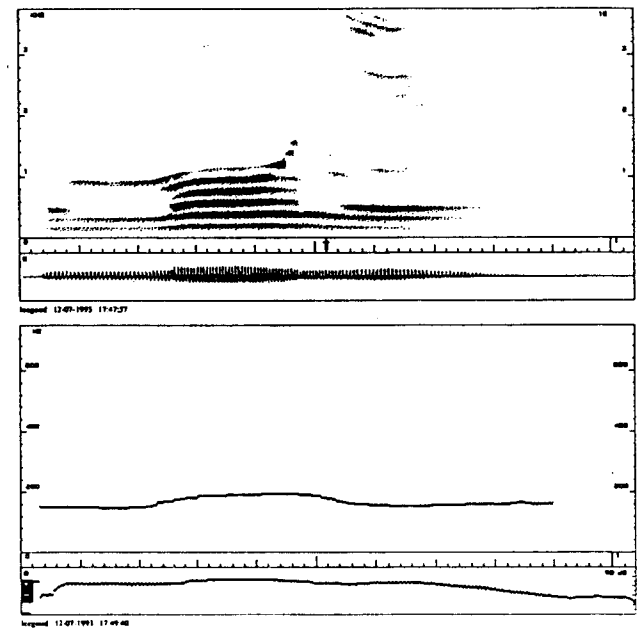


그림 13(a). 男 1 의 /굿모닝/ 의 스펙트로그램.
Fig. 13(a). Spectrogram of original speech of male 1 /good morning/

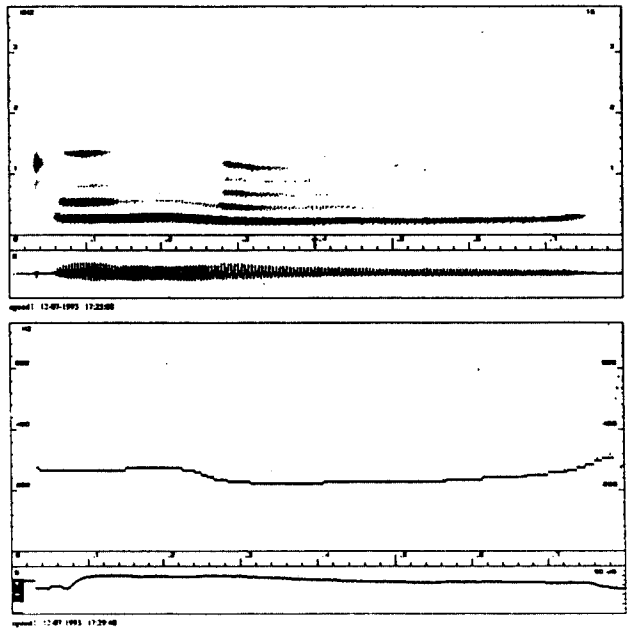


그림 13(b). 女 2 의 /굿모닝/ 의 스펙트로그램.
Fig. 13(b). Spectrogram of original speech of female 2 /good morning/

그림 14(a), (b) 는 퍼지-VQ 사상화함수 FVQ 에 의한 男 1 음성이 女 2 음성으로 변환된 합성음성파형 /안녕하십니까/ 와 그 스펙트로그램이다.

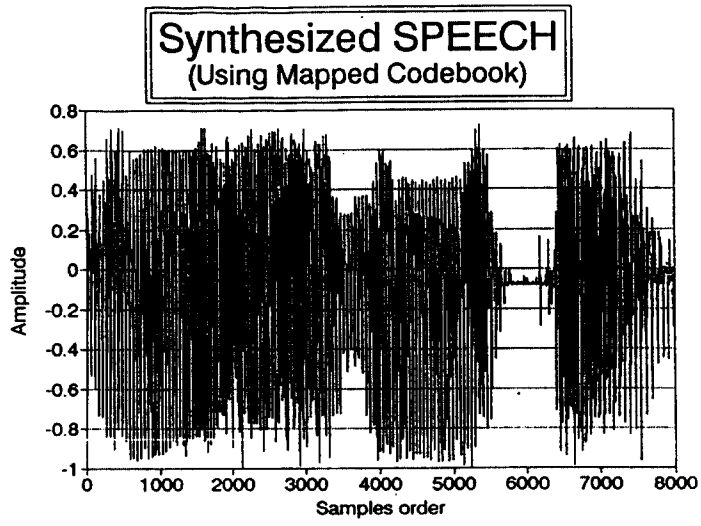


그림 14(a). 퍼지-VQ 사상함수 FVQ에 의한 男 1 음성이 女 2 음성으로 변환된 음성파형 /안녕하십니까/.
Fig. 14(a). Speech waveform /anyoung hasim nika/ of male 1 converted femal 2 by Fuzzy-VQ mapping function FVQ.

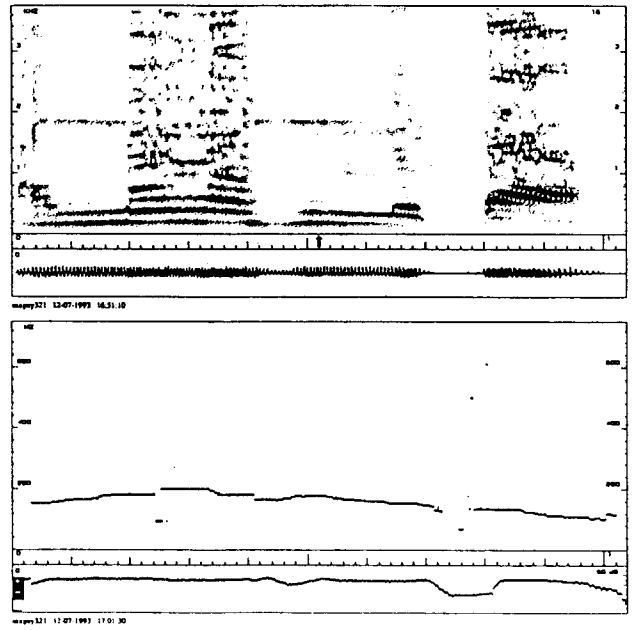


그림 14(b). 변환음성 14(a)의 스펙트로그램
Fig. 14(b). Spectrogram of converted speech 14(a) : /anyoung hasim nika/

그림 15(a), (b) 는 퍼지-VQ 역 사상화함수 $F^{-1}VQ$ 에 의한 女 2 음성이 男 1 음성으로 변환된 합성음성파형 / 굿모닝/ 과 그 스펙트로그램이다.

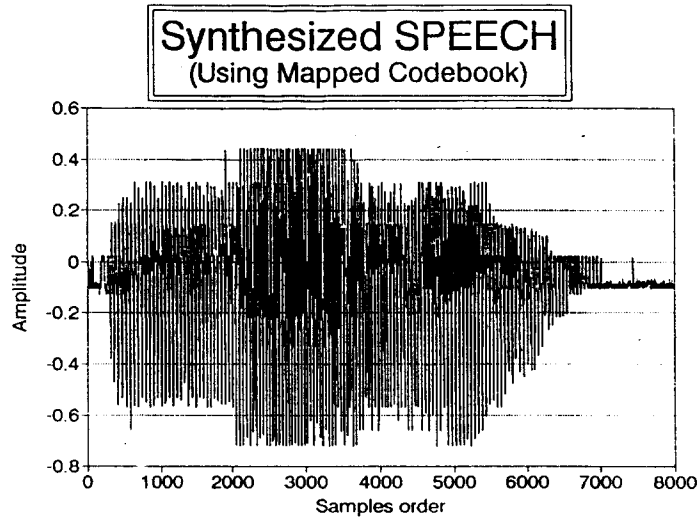


그림 15(a). 퍼지-VQ 역 사상함수 $F^{-1}VQ$에 의한 女 2 음성이 男 1 음성으로 변환된 음성파형 /굿모닝/.
Fig. 15(a). Female 2-to-mal 1 conversion speech /good morning/ by $F^{-1}VQ$: Fuzzy VQ-inverse mapping function

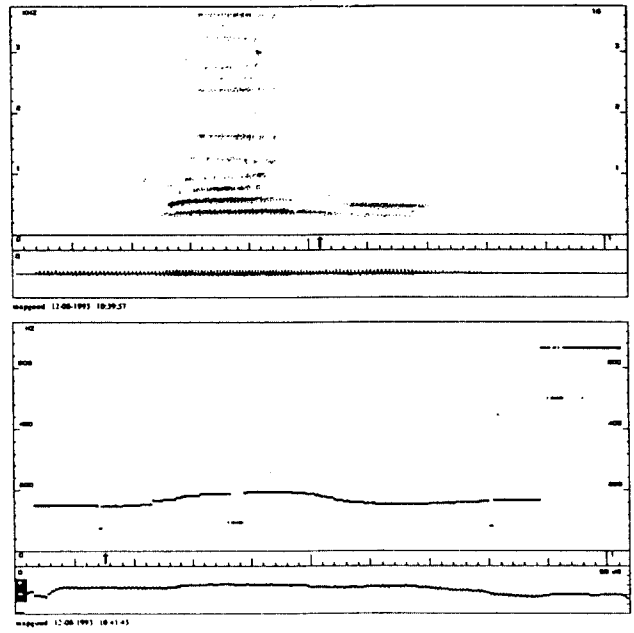


그림 15(b). 합성음성파형 15(a)의 스펙트로그램.
Fig. 15(b). Spectrogram of converted speech 15(a) : /good morning/

실험결과 입력화자 男 1 의 /안녕하십니까/에 적응된 /굿모닝/ 합성음을 얻었다. 그림 14(a) 와 그림 15(a) 의 변환된 합성파형은 입력화자 음성파형과 기준화자 음성의 복합된 파형을 보이고 있다. 그러나 코드북 크기를 늘이면 그림 14(a) 는 기준화자 女 2 의 음성파형인 그림 10(b) 에, 그림 15(a) 는 입력화자인 男 1 의 음성파형인 12(a) 에 거의 가깝도록 될 것이다.

또한 VQ-사상화에 의한 합성음보다 퍼지 VQ-사상화에 의한 합성음이 더 우수함을 알 수 있다. 이것은 기준화자의 코드벡터를 1 차 결합하여 매핑된 코드북을 작성할때 그 계수값을 퍼지 히스토그램값으로 하였기 때문이다.

V. 결 론

이 논문에서는 퍼지 VQ 사상된 코드북을 사용한 화자적응 음성합성기법을 제안하였다. 입력화자의 음성벡터 공간을 기준화자의 음성벡터 공간으로 사상화하는 사상화함수를 누적퍼지 히스토그램의 1 차 결합으로 정의하여 적응음성합성을 하였다. 입력화자와 기준화자의 코드북은 신경망 학습이론에 근거한 경쟁학습 알고리즘으로 작성하였다. 이 기술의 성능평가는 성별이 서로 다른 화자를 입력화자및 기준화자로 선정하여 합성음을 들어서 기준화자의 음성에 가까운정도로 평가 하였으며, 기준화자에 적응된 합성음을 들을 수 있었다. 또한 이 기법을 이용하여 개인음성이 조절된 합성음을 얻을 수 있음을 보였다. 코드북 크기를 늘이므로써 더욱 더 기준화자에 적응된 합성음을 들을 수 있겠다. 화자 반적응 음성합성도 계속 연구되어야할 과제이다.

참 고 문 헌

1. H. P. Tseng, M. J. Sabin, and E. A. Lee, "Fuzzy vector quantization applied to hidden Markov modeling." Proc. ICASSP 87, Paper 15. 5.

2. 이 기영, "사상 멤버쉽함수에 의한 화자적응 단어인식," 명지대학교 박사 학위 논문, 1991.
3. Lee Jin-Yi, Lee Kwang-Hyung, "The Optimum Fuzzy Vector Quantizer for Speech Synthesis," Fifth IFSA Congress, 1993, 1321-1325
4. Kuwabara, H., Takagi, T., "Quality control of speech by modifying formant frequencies and bandwith," 11th Inter. Congress of Phonetic Science, pp. 281-284, August 1987.
5. 이 진이, 김형석, 이광형, "Fuzzy-C-means 알고리즘에 의한 벡터양자화 코드북의 성능비교" 제 3회 인공지능, 신경망및 퍼지시스템 종합학술대회 1993,10,18-20.
6. 이진이, 김형석, 이광형, "신경망 학습벡터양자화기에 의한 음성합성의 성능비교" 인공지능,신경망및 퍼지관련 학술발표회, 1993, 5,1
7. Nakamura, Shikano, "A comparative study of spectral mapping for speaker adaptation." ICASSP, S3. 7, 1990.
8. Kiyohiro Shikano, Kai-Fu Lee, Raj Reddy, "Speaker adaption through vector quantization," ICASSP, pp. 2643-2646, April 1986.
9. Kohonen,T. Self-orginization and associative meomry. Berlin : Springer, Verlag.
10. Naylor. J,. & Li,K. P. "Analysis of a neural network algorithm for vector quantization of speech parameters." Proceedings of the First Annual INNS Meeting, P. 310, New York : Pergamon Press
11. Nasrabadi, N. M., & Feng, Y. "Vector quantization of images based upon the Kohonen self-organizing feature maps," IEEE International Conference on Neural Networks, pp 1101-1108, San Diego : IEEE.
12. Matsuyama, Y,. "Vector quantizationwith optimized grouping and parallel distributed processing," Journal of Neural Networks, 1988.
13. Gray. R. M. "Vector quantization," IEEE ASSP Magaine, 1(2), pp. 4-29,
14. Grossberg. S., "Adaptive pattern classification and universal recording : I. Parallel developement and coding neural feature detectors," Biological Cybernetics.
15. Grossberg. S., "Adaptive pattern classification and universal recording : Ii Feedback, expectation, olfaction, illusions," Biological Cybernetics.
16. Lippmann,R. R, "An introduction to computing with neural nets," IEEE ASSP Maganzine, 4(2), pp. 4-22.
17. Winters. J. H., & Rose, "Minimum distance automata in parallel networks for optimum classification," Neural Networks,2, pp. 127-132,1989.
18. Hecht-Nielsen, R.(1988). Applications of counterpropagation networks. Neural Networks, 1(2), 131-141.
19. Stanley C. Ahalt, Ashok K, Krishnamurthy, Prakoon Chen,and Douglas E. Melton, "Competitive Learning Algorithms for Vector Quantization," Neural Networks 4, 1990.
20. DeSieno. D.(1988). Adding a conscience to competitive learning. In IEEE International Conference on Neural Networks, 1117-1124
21. Grossberg, S.(1987). Competitive learning ; From interactive activation to adaptive resonance. Cognitive Science, 11, 23-63
22. H. J. Zimmermann, Fuzzy set theory and its applications, second edition, Kluwer academic publisers, 1991.