한국음향학회지 제21권 제2호 pp. 160~166 (2002)

VCCV단위를 이용한 어휘독립 음성인식 시스템의 구현

An Implementation of the Vocabulary Independent Speech Recognition System Using VCCV Unit

윤 재 선*, 홍 광 석*
(Jeh Seon Youn*, Kwang Seok Hong*)

*성균관대학교 전기 전자 및 컴퓨터 공학부

(접수일자: 2001년 7월 9일; 수정일자: 2001년 9월 20일; 채택일자: 2002년 1월 18일)

본 논문에서는 CV (Consonant Vowel), VCCV (Vowel Consonant Consonant Vowel), VC (Vowel Consonant) 인식 단위를 이용한 새로운 어휘 독립 음성인식 시스템을 구현하였다. 이 인식 단위는 음절의 안정된 모음 구간에서 분할하여 구성했기 때문에 분할이 용이하다. VCCV 단위가 존재하지 않을 경우에는 VC와 CV 반음절 모델을 결합하여 대체모델을 구성하였다. 모음군 군집화 (clustering)와 VCCV 모델이 존재하지 않을 경우 대체모델에 결합규칙을 적용하여 제 1후보에서 90.4% (모델 A)에서 95.6% (모델 C)로 5.2%의 인식 성능 향상을 가져왔다. 인식실험결과 제 2후보에서 98.8%의 인식률로 제안된 방법이 효율적임을 확인하였다.

핵심용어: 어휘독립인식기, VCCV 인식 단위, 대체모델 결합규칙, 모음군 군집화
투고분야: 음성처리 분야 (2.5)

In this paper, we implement a new vocabulary-independent speech recognition system that uses CV, VCCV, VC recognition unit. Since these recognition units are extracted in the vowel region of syllable, the segmentation is easy and robust. And in the case of not existing VCCV unit, the units are replaced by combining VC and CV semi-syllable model. Clustering of vowel group and applying combination rule to the substitution model in the case of not existing of VCCV model lead to 5.2% recognition performance improvement from 90.4% (Model A) to 95.6% (Model C) in the first candidate. The recognition results that is 98.8% recognition rate in the second candidate confirm the effectiveness of the proposed method.

Keywords: *Vocabulary-independent speech recognition, VCCV recognition unit, Combination rule of substitution model, Clustering of vowel group*
ASK subject classification: *Speech signal processing (2.5)*

I. 서 론

현재 상용화된 다수의 음성인식 시스템은 인식할 대상 어휘를 미리 선정하여 이 어휘들에 대한 음성 데이터베이스를 수집한다. 그리고 이 음성 데이터베이스를 사용하여 인식할 단어 또는 음소 모델을 훈련하는데, 이와 같은 방식의 음성인식 시스템은 선정한 어휘들에 대해서는 높은 인식 성능을 보이지만, 인식대상 어휘를 변경하거나 추가할 때마다 새로운 어휘들에 대하여 음성 데이터베이스를 별도로 수집하여 처음부터 다시 모델을 훈련해야 하므로 많은 시간이 소비되고 추가 비용이 드는

책임저자: 윤재선 (sunhci@netian.com)
440-746 경기도 수원시 장안구 천천동 300
성균관대학교 전기 전자 및 컴퓨터 공학부 휴먼 컴퓨터 연구실
(전화: 031-290-7196; 팩스: 031-290-7170)

문제점이 발생한다. 특히, 화자독립 음성인식 시스템의 경우 여러 명의 화자에 대해서 음성 데이터베이스를 수집해야 하기 때문에 이러한 문제점은 더욱 증대될 수 있다. 따라서 어휘독립 인식기술은 이러한 문제점을 해결하기 위한 음성인식 기술로, 별도의 훈련과정 없이 인식대상 어휘를 변경하거나 추가할 수 있는 장점이 있다. 여러 명의 화자로부터 음운 현상이 충분히 반영된 대용량 음성 데이터베이스를 사용하여 미리 구성한 기준 모델을 토대로 인식대상 어휘가 결정되면 발음 사전에 따라 해당하는 모델을 연결함으로써 모델을 만든다. 이 방법은 처음 한 번만 기준 모델을 훈련하기만 하면 인식대상 어휘를 변경하거나 추가할 때마다 별도의 음성 데이터베이스 수집이나 훈련 과정이 필요하지 않다[1].

현재 음성인식에 사용되고 있는 단위는 음소, 음절, 단어 등 언어적으로 정의된 단어들과 음향, 음성학적인 유사도에 기반한 유사음소 단위들이 사용되고 있다. 단어를 기본 단위로 사용할 경우 단어 사전을 구성하지 않아도 되기 때문에 구성상 간단하나 모든 단어에 대한 모델을 전부 구성해야 되므로 연속음성인식으로의 확장이 어렵다. 음소를 기본 단위로 사용할 경우, 데이터의 양은 적으나 인접하는 음운사이의 천이를 포함하지 못하는 단점이 있다. 음절을 기본 단위로 사용할 경우, 무제한 어휘 인식이 가능하나 실제로 발성되는 데이터는 음절과 음절, 음소와 음소 사이의 조음 현상으로 인하여 인식률이 낮다. 반음절을 기본 단위로 사용할 경우, 음절에 비해 데이터베이스의 규모를 줄일 수 있는 장점이 있다. 다이폰을 기본 단위로 사용할 경우, 음절이 가진 장점을 살리고 인식 단위의 개수를 줄일 수 있는 반면에 음성인식 과정에서 자음 부분의 안정 구간을 검출하는 방법에 어려움이 있다. 트라이폰과 같은 복합 음소열을 단위로 선택하면 인식 성능은 좋지만, 특별히 훈련된 숙련자의 음성분할 작업이 필요하며 그 결과 많은 시간과 비용이 요구된다.

이를 보완하기 위해 본 논문에서는 단음절 데이터 521개, 108개의 성이 포함된 성명데이터 1,145개, PBW (Phonetically Balanced Word)가 포함된 임의의 데이터 1,001개로부터 분맥종속형 CV, VCCV (VV, VCV포함), VC 단위를 적용하여 어휘 독립 음성인식 시스템을 구현하였다. 이 인식 단위는 음절의 안정된 모음 구간에서 분할하여 구성하기 때문에 분할이 용이하다[2]. VCCV단위가 존재하지 않을 경우에는 대체모델의 연결규칙과 모음군 군집화 방법을 적용하여 VC와 CV 반음절 단위를 결합하여 대체모델로 구성한 후, 제안한 방법의 성능을 확인하였다.

II. VCCV단위를 이용한 어휘독립 음성인식 시스템

분할 정보를 이용하여 훈련 데이터로부터 어휘독립 음성인식 시스템에 사용될 CV모델 323개, VCCV모델 1,611개, VC모델 56개를 기준 모델로 구성하고, HMM (Hidden Markov Model)을 위한 VQ 코드북 생성은 전체 음성의 특징 벡터를 가장 효율적으로 표현할 수 있는 K-Means 알고리즘을 이용하여 무제한 어휘인식을 위해 128개의 코드워드를 갖는 코드북을 구성하였다[3-5].

인식 단계에서는 인식목록 어휘를 CV, VCCV, VC모델로 연결하여 단어 모델을 구성하고, 입력 음성으로부터 16차 멜 켑스트럼 특징 파라미터를 추출하여 VQ (Vector Quantization)를 통과시켜 코드워드열로 이산화시킨 후, 각 단어 모델과 확률값을 비교하여 인식된 단어를 결정하는 HMM 어휘독립 음성인식 시스템을 그림 1에 나타내었다[6].

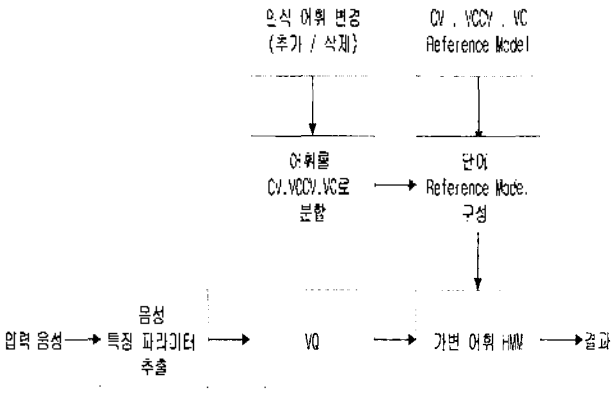

그림 1. 어휘독립 음성인식 시스템
Fig. 1. Vocabulary independent speech recognition system.

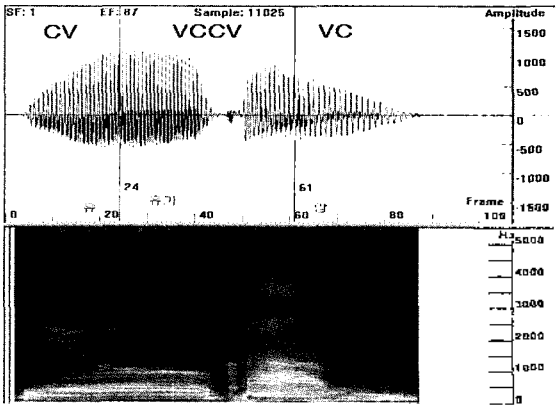

그림 2. CV, VCCV, VC 분할(/유감/)
Fig. 2. Segmentation of CV, VCCV and VC.

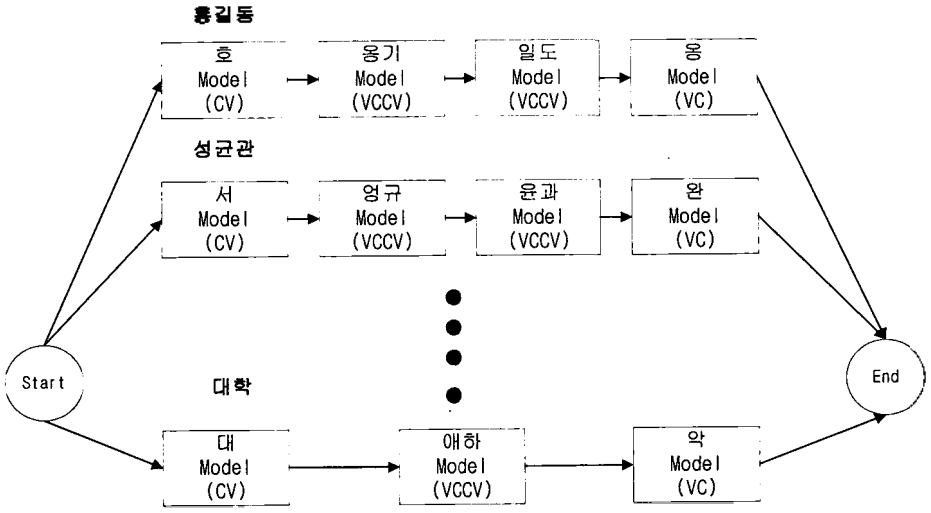

그림 3. 어휘독립 음성인식 네트워크
Fig. 3. Network of vocabulary independent speech recognition.

그림 2는 /유감/의 단어를 분할한 것으로 앞쪽 CV는 /유/, 가운데 VCCV는 /유가/, 마지막으로 뒤쪽 영역 VC는 /암/의 영역이며, 각 분할된 영역은 저장된 후 HMM에 의해 기준 모델을 구성한다.

어휘독립 음성인식에서 단어 모델의 결합 방법은 목록 텍스트로부터 CV, VCCV, VC단위로 분할한 후, 각각에 해당하는 기준 모델을 연결하여 단어 모델을 구성한다. 그림 3은 본 어휘독립 음성인식 시스템의 단어 네트워크 구성을 나타낸 것으로써 인식목록 단어가 /성균관/인 경우 앞쪽 반음절 CV /서/, VCCV /엉규/, /윤과/, 뒤쪽 반음절 /완/으로 분할한 후 CV, VCCV, VC의 기준 모델을 결합하여 단어 모델을 만드는 과정을 나타내고 있다.

일반적인 인식 시스템의 경우 입력된 단어를 발음 사전에 따라 소리나는 대로 발성한 음소열로 구분하여 인식 시스템을 구축하고 있으나, 모든 화자가 항상 같은 발음으로 발성하는 것이 아니기 때문에 한 단어에 대해서 많은 기준 모델을 구성해야 하며 또한 인식 시간이 훨씬 많이 소요되는 단점이 있다. 그러나 본 논문에서 제안한 VCCV 기준 모델 안에 모든 경우의 소리나는 발음을 포함하고 있기 때문에 단어 모델을 구성하는데 장점이 있다.

CV모델과 VC모델은 단음절 데이터 521개로부터 모두 구할 수 있지만, 종성 법칙을 고려한 VCCV모델의 개수는 67,032개 (21×8×19×21)가 필요하기 때문에 모든 VCCV 단위를 구성하는 것은 실제적으로 불가능하므로 본 논문에서는 훈련 데이터에 존재하지 않는 VCCV모델은 VC와 CV모델을 연결해 구성하도록 시스템을 설계하였다. 예를 들어 단어 /학습/에 /악스/의 VCCV모델이 없다면, /악/의 VC 모델과 /스/의 CV 모델을 연결하여 VCCV 모델을 만드는 것이다.

본 어휘독립 음성인식 시스템의 기준 모델에 사용된 상태 수는 CV, VC는 3개 또는 4개, VCCV는 7개의 상태를 두어 HMM 모델을 훈련하였다. 일반적으로 CV와 VC는 초성 + 중성, 중성 + 종성인 두 개의 음소로, VCCV는 중성 + 종성 + 초성 + 중성인 네 개의 음소로 구성되어 있다. 따라서 기준 모델의 상태수는 음소의 개수와 음소의 천이시 발생되는 조음 현상을 포함하기 위해 음소의 개수보다 많은 상태의 수를 두고 구성하였다.

III. 단어 결합 모델링

3.1. 모델 결합 방법

데이터베이스를 추가함에 따라 추출된 기준 모델의 증가는 표 1에 나타냈으며, 단음절 데이터로부터 모든 CV, VC기준 모델을 포함할 수 있도록 하였다.

본 논문에서 제안한 CV, VCCV, VC모델은 연결 부분이 모두 안정 구간에서 분할된 단위이다. CV는 무음 + 천이 + 안정 구간, VC는 안정 + 천이 + 무음 구간, 그리고 VCCV

표 1. 기준 모델의 증가
Table 1. Increment of reference model.

종류 / 데이터베이스	CV	VCCV	VC
단음절 데이터	380	0	160
성명 데이터	380	1,227	161
PBW 데이터	383	2,489	166

표 2. 출력 확률 분포값을 갖는 번호
Table 2. Index with output probability distribution value.

인식 단위	Index
CV /가/	0,3,4,7,11,12,13,14,15,16,19,20,22,23,24,25, 26,27,28,30,32,33,34,35,36,40,53,63,65,67,75, 79,85,116
CV /아/	3,6,7,11,12,14,15,16,19,20,22,23,24,25,26,27, 33,34,36,37,40,42,44,45,53,56,57,62,63,64,65, 67,70,74,75,77,116
VCCV /아겨/	4,7,12,14,16,19,20,22,23,24,25,27,31,32,33,34, 43,64,65,67,69,72,116
공통의 index	12,14,16,19,20,22,23,24,25,27,33,34,65,67

는 안정 + 천이 + 안정 구간을 가지고 있다. 즉 다른 모델과 연결되는 부분 CV의 뒤쪽 영역, VC의 앞쪽 영역 그리고 VCCV의 양쪽 영역은 안정 구간으로 구성되어져 있다. 표 2는 같은 /아/ 모음 계열 /가/와 /아/의 마지막 회귀 상태의 출력 확률값과 VCCV 모델 /아겨/의 첫 번째 회귀 상태의 출력 확률값을 가지는 VQ 번호값을 나타내고 있다.

표 2를 살펴보면, 공통의 번호값이 존재하며, 각각 모델의 공통 번호의 확률값을 합하면 0.8이상의 높은 값을 나타내기 때문에 연결 부분의 주파수 특성이 유사하다는 것을 알 수 있다. 따라서 CV의 마지막 상태, VCCV의 양쪽 끝 상태, VC의 첫 번째 상태에 결합 방법을 적용하면 무제한으로 어휘를 인식할 수 있는 시스템을 구축할 수 있다. 그림 4는 CV모델과 VCCV 모델의 결합 과정을 나타낸 것으로써 예를 들어 /겨울/의 앞쪽 반음절 CV /겨/와 VCCV /여우/의 결합 방법을 설명하고 있다. 또한 이 방법은 VCCV 모델이 존재하지 않을 때 VC모델과 CV모델의 결합시에도 동일한 방법으로 결합하여 대체 모델을 구성하였다.

표 3. 모델(I)과 모델(II)의 모델 구성 방법
Table 3. Model construction method of model(I) and model(II).

General		Special	
CV	VC	CV	VC
/ㄱ, ㄸ, ㅃ, ㅉ, ㅆ/ SF - 3 F	모든 자음 EF + 1 F	초성이 없는 경우 SF+(EF - SF)*0.25	종성이 없는 경우 (EF-SF)*0.25 + SF
		/ㄱ, ㄷ, ㅂ/ SF + 3 F	
		/ㅎ/ SF + 5 F	/ㄱ, ㄷ, ㅂ/ EF + 3 F
그 이외의 자음 SF - 1 F		/ㄲ, ㄸ, ㅃ, ㅆ, ㅉ, ㅋ, ㅌ, ㅍ, ㅈ/ SF + 1 F	
		/ㄴ, ㄹ, ㅁ/ SF + 4 F	/ㄴ, ㄹ, ㅁ, ㅇ/ EF - 3 F
		/ㅅ, ㅈ, ㅊ/ SF + 3 F	

SF: Start Frame, EF: End Frame, F: Frame

3.2. 대체모델 연결규칙

VCCV모델은 화자에 따라 다르게 발성되는 경우를 포함하여 기준 모델을 구성하였기 때문에 인식 성능에는 큰 문제점이 없으나, VC모델과 CV모델을 결합하여 VCCV 대체 모델을 구성할 경우에는 음운변동 규칙 적용과 더불어 음절의 종성과 다음 음절의 초성이 상호 작용을 고려하여 음절의 음향학적 특성을 변형시킨 후 연결해 주어야 한다[7].

두 음절을 연결시켜 주는 음절의 연결 규칙을 고려하여 CV, VC모델을 모델 (I)과 모델 (II)로 구성하였다. 모델 (I)은 앞뒤의 음절에 무음 구간을 포함한 기준 모델을 구성하였고, 모델 (II)는 조음 현상에 따른 음소 길이의 변화를 고려하여 기준 모델을 구성하였으며, 표 3에 구성 방법을 나타내었다[8]. 분할된 모델의 길이 정보에 따라 모델

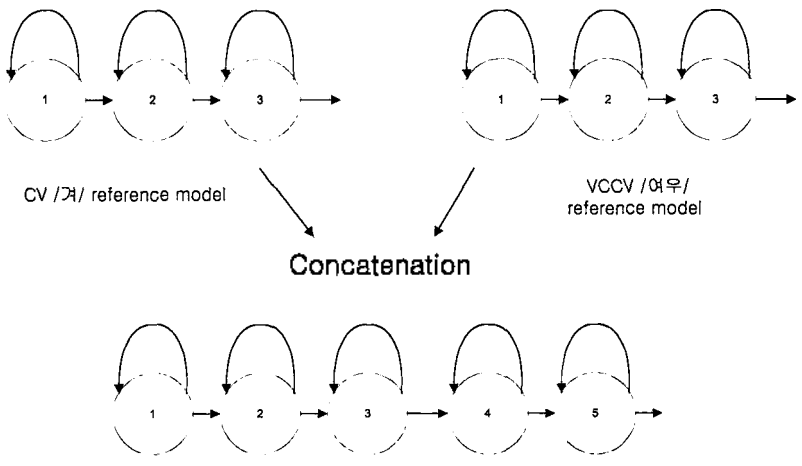

그림 4. CV모델과 VCCV모델의 결합
Fig. 4. Concatenation of CV model and VCCV model.

(I)의 상태는 4개, 그리고 모델 (II)는 VC의 /ㄱ ㄷ ㅂ/을 제외하고 3개의 상태로 기준 모델을 구성하였다. 어휘의 맨 앞과 맨 뒤의 CV와 VC모델은 모델 (I)을 연결하여 구성하고, 존재하지 않는 VCCV의 대체 모델은 규칙을 적용하여 결합하였다.

3.3. 군집화

훈련 데이터로부터 모든 VCCV 기준 모델을 구성할 수 없기 때문에 비슷한 특성을 가지는 음소를 하나의 클래스로 모델링하는 것이 필요하다. 음운 규칙에 따라 소리나는 대로 변환할 경우에는 초성과 종성의 변환만 나타나며, 중성에는 변화가 없다. 따라서 모음군에 따라 CV, VCCV, VC단위의 중성에 대해 군집화 작업을 통해 기준 모델의 감소와 단어결합률의 증가를 가져오도록 하였다[9].

앞쪽 반음절 CV기준 모델의 모음열은 ㅐ/ㅔ, ㅒ/ㅖ, 괘/괴/궤를 유사 모음군으로 구성하여 17개의 모음군을 표 4에 나타냈다.

뒤쪽 반음절 VC 기준 모델의 모음열은 이중 모음의 경우 반모음이 나타나다가 바로 단모음으로 소리가 발생되는 것을 고려하여 뒤쪽 반음절인 VC 기준 모델을 표 5와 같이 구성하였다.

군집화 작업을 적용하여 CV와 VC는 각각 323개 (19×17), 56개 (7×8)로 줄어들었다. 또한 VCCV모델도 같은 방법으로 VCCV의 앞쪽 모음은 뒤쪽 반음절 VC모음군을, 뒤쪽 음절의 모음은 앞쪽 반음절 CV모음군을 적용하여 데이터베이스들로부터 추출된 VCCV 기준 모델의 개수를 2,489개에서 1,611개로 줄일 수 있었다.

표 4. CV 모음군
Table 4. CV vowel group.

CV 모음군				
아	애/에	야	얘/예	어
여	오	와	왜/외/웨	요
우	워	위	유	으
의	이			

표 5. VC 모음군
Table 5. VC vowel group.

VC 모음군	종류
아	아, 야, 와
애	애, 에, 얘, 예, 웨, 외, 왜
어	어, 여, 워
오	오, 요
우	우, 유
으	으
이	이, 위, 의

IV. 실험 및 고찰

본 논문에서 사용한 음성 데이터는 사무실 환경에서 20~30대 남성 화자 107명이 발성한 521개의 단음절 데이터, 62명이 발성한 108개의 성이 포함된 1,145개 성명 데이터, 그리고 60명이 발성한 PBW 데이터가 포함된 1,001개의 데이터를 16 bits, 11,025 kHz로 샘플링한 후, 훈련 데이터량에 따른 인식 성능의 실험, VCCV 대체 모델에 규칙을 적용하지 않은 모델과 적용한 모델과의 인식 성능 비교, 모음의 군집화 작업을 적용한 모델과 적용하지 않은 모델의 인식 실험을 비교, 평가하였다.

4.1. 훈련 데이터에 따른 인식 성능

연구실에서 수집한 단음절 데이터로만 구성된 4.8 MB의 모델 1, 단음절 + 성명 데이터로 구성된 23.8 MB의 모델 2, 단음절 + 성명 데이터 + PBW 데이터로 구성된 43 MB의 모델 3으로 구분하여 훈련 데이터의 용량에 따른 인식 실험을 하였다.

인식 실험에 사용된 데이터는 단음절에서 종성이 없는 음절 6개, 무성종성자음의 음절 6개, 유성종성자음의 음절 6개, 총 18개의 단음절 데이터와 전화번호부에서 임의로 추출한 16개의 성명 데이터, 그리고 한국어 발음 예문인 바람과 햇님 문장으로부터 선정한 16개의 단어, 총 50개의 테스트 데이터를 가지고 실험하였으며, 그 목록을 표 6에 나타내었다.

훈련 데이터에 용량에 따른 인식실험 방법은 모음의 군집화 규칙과 VCCV 대체 모델의 결합 규칙을 적용하지 않은 방법으로 기준 모델을 구성하여 실험하였다. 모델 1로부터 추출된 반음절 CV와 VC는 각각 380개와 161개, 모델 2로부터는 CV 380개, VC 161개, VCCV 1,227개, 모델 3은 CV 383개, VC 166개, VCCV 2,489개로 기준 모델을 구성하였다. 단음절 데이터를 제외하고 각 모델에 따른 VCCV 대체모델 비율은 모델 1이 100% (32개), 모델 2는 81.25% (26개), 모델 3은 59.4% (19개)로 나타났다. 즉, 데이터베

표 6. 테스트 데이터
Table 6. Test data.

구분	데이터
단음절	노, 다, 메, 비, 화, 수, 국, 목, 습, 작, 찾, 턱, 급, 란, 월, 일, 장, 형
성명	강명숙, 김은주, 노수정, 도석훈, 류상민, 박기수, 서판주, 오지영, 윤정훈, 이지영, 전석표, 정대진, 최재수, 태용호, 표안철, 황정섭
임의	바람, 햇님, 나그네, 외투, 왔습니다, 그들은, 누구든지, 벗기는, 결정, 복풍, 단도히, 이때, 뜨거운, 가만히, 돌중에, 인정하지

표 7. 데이터베이스에 따른 인식 성능
Table 7. Recognition performance according to the database.

Model / 화자	Model 1		Model 2		Model 3	
	제1후보	제2후보	제1후보	제2후보	제1후보	제2후보
1	35/50	37/50	39/50	44/50	44/50	48/50
2	36/50	40/50	43/50	47/50	47/50	49/50
3	32/50	35/50	38/50	42/50	44/50	47/50
4	34/50	35/50	41/50	44/50	45/50	47/50
5	34/50	37/50	39/50	44/50	46/50	47/50
평균	68.4%	73.6%	80%	88.4%	90.4%	95.2%

이스의 수가 증가함에 따라 대체 모델의 비율이 감소함을 알 수 있다. 인식 방법은 남성 화자 5명의 결과를 제 1 후보와 제 2 후보로 구별하여 실험하고, 각각의 모델에 따른 실험 결과를 표 7에 나타내었다.

모델 1은 단음절 데이터베이스로 구성되었음에도 불구하고, /비/, /찾/, /메/, /금/, /월/과 같은 단음절 후보에서도 오인식되는 경우가 있었다. 그러나 모델 2부터는 /비/, /찾/ 음절을 제외하고 모든 단음절 데이터를 인식하였다. 즉 훈련 데이터의 양이 많을수록 인식률이 높아짐을 알 수 있었다. 또한 모델 2와 모델 3사이에서도 대체모델 비율의 감소와 훈련 데이터량의 증가에 따라 제 1 후보의 인식률은 80%에서 90.4%로 향상되었다.

4.2. VCCV 결합 규칙에 따른 인식 성능

VCCV 결합규칙 모델의 훈련 데이터베이스는 단음절, 성명, PBW가 포함된 데이터를 사용하였다. 따라서 모델 A는 앞에서 실험한 모델 3과 동일한 훈련 모델이며, 모델 B는 모델 (I), 모델 (II)로 나누어 기준 모델을 구성하였다. 모델 A의 기준 모델의 용량은 43 MB이며, 반음절 CV와 VC모델이 증가된 모델 B의 용량은 47 MB이다. 실험 방법은 앞의 실험과 동일하게 남성 화자 5명의 결과를 제 1 후보와 제 2 후보로 구별하여 실험하였으며 그 결과를 표 8에 나타내었다.

존재하지 않는 VCCV 모델의 경우, 음절 결합에 따른 결합 규칙을 적용하는 것이 제 1 후보와 제 2 후보에서 3.6%의 인식 성능 향상을 가져왔다. 즉, 대체모델을 구성할 경우 CV와 VC만을 이용한 연결 방법 (모델 A)보다는 연결 규칙에 의해 자연스러운 결합으로 더 좋은 인식 성능이 나타났다.

4.3. 군집화 모델과 비군집화 모델의 인식 성능

VCCV결합 규칙만을 적용한 기준 모델 (모델 B)과 VCCV결합 규칙과 군집화 작업을 거친 기준 모델 (모델 C)을 이용하여 인식 성능을 평가하였다. 표 9는 5명의 화자

표 8. 결합 규칙에 따른 인식 성능
Table 8. Recognition performance according to combinational rule.

Model / 화자	Model A		Model B	
	제1후보	제2후보	제1후보	제2후보
1	44/50	48/50	47/50	50/50
2	47/50	49/50	47/50	49/50
3	44/50	47/50	46/50	49/50
4	45/50	47/50	47/50	50/50
5	46/50	47/50	48/50	49/50
평균	90.4%	95.2%	94%	98.8%

표 9. 인식 결과
Table 9. Recognition result.

Model / 화자	Model B		Model C	
	제1후보	제2후보	제1후보	제2후보
1	47/50	50/50	48/50	50/50
2	47/50	49/50	47/50	49/50
3	46/50	49/50	48/50	49/50
4	47/50	50/50	48/50	50/50
5	48/50	49/50	48/50	49/50
평균	94%	98.8%	95.6%	98.8%

에 대해서 모델 B과 모델 C로 구성된 모델에 의한 인식 결과이다.

모델 B에서는 /찾/, /전석표/, /표인철/, /햇님/, /벗기는/과 같은 단어가 오인식되었으나, 모델 C에서는 /찾/, /햇님/, /벗기는/ 단어가 오인식되었다. 따라서 모음에 따라 기준 모델을 구성하여 적은 수의 훈련 데이터를 가지고 기준 모델을 만드는 것보다는 많은 수의 훈련 데이터를 가질 수 있는 유사 모음을 같은 클래스로 구성해 기준 모델을 만드는 것이 효과적임을 알 수 있다. 오인식된 단어를 살펴보면, /햇님/인 경우에는 /햇/, /벗기는/에서는 /벗/과 같이 음절이 짧게 발성되는 모음이 들어간 음절들이 포함된 단어가 오인식되었다.

V. 결 론

본 논문에서는 분할이 비교적 쉬우면서 문맥 종속형을 포함한 CV, VCCV, VC 인식 단위를 설정하였으며, 이 단위는 음절의 모음 구간에서 분할하여 인식 단위를 구성하기 때문에 분할이 용이하며, 또한 VCCV 단위가 존재하지 않을 경우에는 VC와 CV 반음절 단위를 결합하여 높은 인식 성능을 나타낼 수 있는 어휘독립음성인식 시스템을 구현하였다.

VCCV 대체 모델을 구성할 경우, 결합 규칙을 적용하고 모음군에 따라 기준 모델을 구성하는 것이 더 좋은 인식률이 나타남을 알 수 있었다.

향후 빈도수가 높은 단어와 외래어가 포함된 데이터베이스를 구축하면 좀더 좋은 어휘독립 음성인식 시스템을 구현할 수 있을 것이다.

참 고 문 헌

1. H. W. Hon, "Vocabulary-independent speech recognition: the VOCIND system," CMU-CS-92-108, March 16, 1992.
2. 윤재선, 홍광석, "반음절 단위 HMM을 이용한 연속 숫자 음성인식," 한국음향학회지, 제17권 제5호, pp. 73-78, 1998.
3. 윤재선, 정광우, 홍광석, "무제한 단어 인식 시스템을 위한 VCCV 분할에 관한 연구," 한국음향학회, 하계 학술발표대회 논문집, 제19권 제1호, 2000.
4. D. G. Childers, Speech Processing and Synthesis Toolboxes, John Wiley &Sons Inc, New York, Appendix 9, pp. 330-367, 1999.
5. J. G. Wilpon and L. R. Rabiner, "A Modified K-Means Clustering Algorithms for use in Isolated Word Recognition," *IEEE Trans. on Acoust. Speech and Signal Proc.*, vol. 33, no. 3, pp. 587- 594, 1985.
6. 최인정, "단어 결합 모델링을 이용한 한국어 연속 음성 인식에 관한 연구," 석사학위논문, 한국과학기술원, 1994.
7. 이용주, "반음절 단위, LSP 방식에 의한 한국어 음성의 규칙 합성에 관한 연구," 박사학위논문, 고려대학교, 1992.
8. 전효섭, "한국어 음성합성에서의 음절 연결 및 음율조절에 관한 연구," 석사학위논문, 한국과학기술원, 1988.
9. 정용주, "대용량 단어 인식에서의 모음 분류를 이용한 시간 감축에 관한 연구," 석사학위논문, 한국과학기술원, 1990.

저자 약력

● 윤 재 선 (Jeh Seon Youn)

1996년 2월: 성균관대학교 전자공학과 공학사
1998년 2월: 성균관대학교 전자공학과 공학석사
2002년 2월: 성균관대학교 전기 전자 및 컴퓨터 공학부 공학박사
※ 주관심분야: 음성 인식 및 화자 인증

● 홍 광 석 (Kwang Seok Hong)

1985년 2월: 성균관대학교 전자공학과 공학사
1988년 2월: 성균관대학교 전자공학과 공학석사
1992년 2월: 성균관대학교 전자공학과 공학박사
1990년 3월~1993년 2월: 서울보건전문대학교 전산정보처리과 전임강사
1993년 3월~1995년 2월: 제주대학교 정보공학과 전임강사
1995년 3월~1996년 2월: 성균관대학교 전자공학과 조교수
1996년 2월~현재: 성균관대학교 전기 전자 및 컴퓨터 공학부 조교수
※ 주관심분야: 음성 인식 및 신호 처리