모음열과 VCCV단위 HMM을 이용한 연속 숫자 음성인식

윤재선°, 정광우*, 홍광석°
°성균관대학교 전기전자컴퓨터공학부 휴먼컴퓨터연구실
*한국철도대학 운전기전과

A Continuous Digits Speech Recognition Applied Vowel Sequence and VCCV Unit HMM

Jeh-Seon Youn°, , Kwang-Woo Chung*, Kwang-Seok Hong°
°HCI Lab, Electrical & Computer Engineering, Sungkyunkwan University
*Dept. of Operation-Mechatronics, KOREA Railroad College
sunhci@ece.skku.ac.kr°, ckw@chuldo.krc.ac.kr*, kshong@yurim.skku.ac.kr°

요 약

본 논문에서는 조음 효과에 대처할 수 있는 반음절, 반음절 + 반음절 단위 HMM과 모음열 정보를 적용하여 연속 숫자 음성인식을 구현하였다. 모음열 정보를 적용하여 기준모델을 모음이 포함된 HMM단위로만 구성한 시스템과 모든 기준모델과 비교하는 시스템과 성능을 비교하였다. 인식실험결과 인식률의 향상으로 제안된 방법이 효율적임을 확인하였다.

1. 서 론

컴퓨터의 발전과 통신을 이용한 정보 및 금융 서비스가 확대됨에 따라서 주민등록번호, 비밀번호, 통장번호, 회원번호 등 많은 분야에서 무제한 연속 숫자열에 대한 인식을 필요로 하고 있다. 이들 연속 숫자열에 대한 인식은 키보드 입력뿐만 아니라 음성인식 등의 수단에 의해서 확인할 필요성이 증가하고 있다. 연속 숫자 음성인식은 숫자음의 경계가 불명확한 곳이 많고, 조음 현상으로 인해, 같은 음소라도 다르게 발음되는 경우가 있다. 따라서 오인식율도 고립 숫자음의 경우에 비해서 훨씬 높게 된다. 그러나, 이런 문제에도 불구하고 연속 숫자 음성인식의 중요성은 명백하다. 그 이유는 연속 음성인식을 통해서만 인간과 기계사이의 통신수단의 자연스러움과 원하는 속도를 얻을 수 있기 때문이다. 기존의 숫자음 인식은 주로 음절 또는 음소 등의 부단어 단위를 이용하여 인식을 하여 왔지만 연속 발성된 경우 음성의 특성상 음절, 음소 단위로 정확한 분할을 하기가 매우 어려우며, 음성인식에서 오인식률은 부정확한 경계 분할이 원인이다.

따라서 본 논문에서는 모음열의 정보를 이용하여 반음절, 반음절 + 반음절을 인식단위로 하는 연속 숫자 음성인식 시스템을 제안하고 그 성능을 확인하였다.

2. 연속 숫자 음성인식 시스템

연속 숫자 음성인식에서 단어 단위로 인식을 할 경우, 그 기준패턴을 위한 기억용량 및 계산 시간을 상당히 많이 필요로 한다. 그러나, 음소나 반음절 등의 부단어 단위로 경계를 구분할 수 있다면 대용량의 고성능 인식 시스템의 구성이 가능하다.[1] 한국어 연속 숫자 음성의 경우 0에서 9까지 10개의 단음절의 조합으로 구성되어 있다. 그러나, 연속된 숫자음의 발성시에는 숫자와 숫자 사이에 음절의 구분이 없이 연결되어 나타나는 경우를 음성 파형 관찰에서 쉽게 확인할 수 있다. 일례로써 그림 1에 사연 숫자음 5235/오이삼오/라고 발성한 음성 파형과 멜스케일의 스펙트로그램을 나타내었다. 그림

1과 같이 자연스러운 발성으로 인하여 오와 이사이, 삼과 오사이가 연결되어 있고, 이와 같은 경우 정확한 음절 구분이 쉽지 않게 된다.[2]

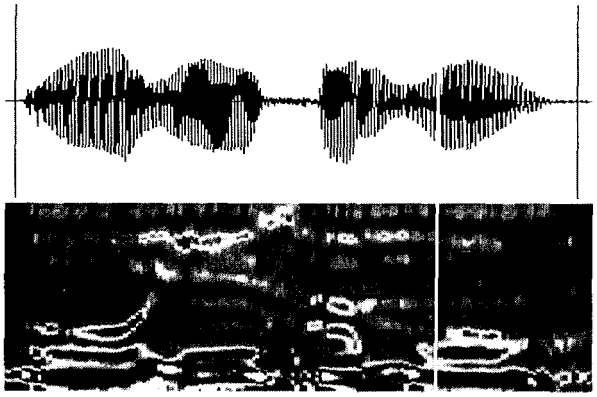

그림 1. 숫자음 5235의 음성파형과 스펙트로그램

따라서 연속 숫자음성인식 시스템에 사용되는 기준모델은 모음을 기준으로 한 반음절 CV, VC 단위와 VCCV 인식단위를 사용하며 이때 사용되는 후보는 표 1과 같다.

표 1 인식시스템에 사용된 기준모델

인식단위	종류	개수
CV	고, 구, 사, 오, 유, 이, 치, 파	8
VC	아, 알, 암, 오, 웅 우, 욱, 이, 일	9
VCCV	아고, 아구, 아사, 아오, 아유, 아이, 아치, 아파, 알고, 알구, 알사, 알오, 알유, 알이, 알치, 알파, 암고, 암구, 암사, 암오, 암유, 암이, 암치, 암파, 오고, 오구, 오사, 오오, 오유, 오이, 오치, 오파, 웅고, 웅구, 웅사 ,웅오, 웅유, 웅이, 웅치, 웅파, 우고, 우구, 우사, 우오, 우유, 우이, 우치, 우파, 욱고, 욱구, 욱사, 욱오, 욱유, 욱이, 욱치, 욱파, 이고, 이구, 이사, 이오, 이유, 이이, 이치, 이파, 일고, 일구, 일사, 일오, 일유, 일이, 일치, 일파	72

숫자음 /일/과 /이/의 앞쪽 반음절 CV는 /이/로, /삼/과 /사/의 앞쪽 반음절 CV는 /사/이기 때문에 CV의 인식 단위의 개수는 8개이며, /일/과 /칠/의 뒤쪽 반음절 VC가 /일/이므로 VC인식 단위의 개수는 9개이다.

연속숫자음 인식시스템의 순서도는 그림 2와 같다.

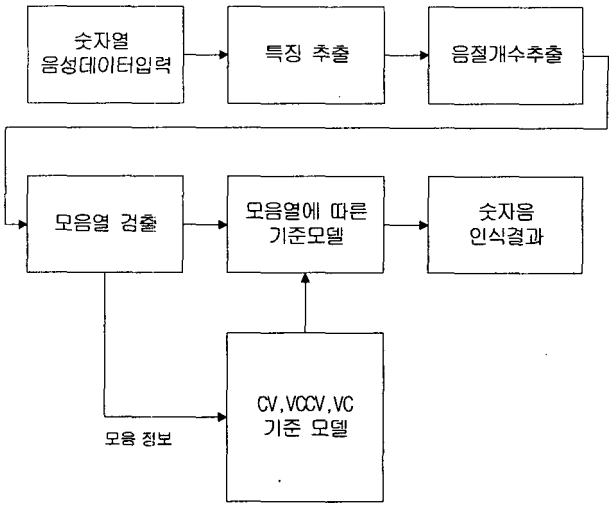

그림 2 연속 숫자음 인식 시스템

입력된 음성데이터로부터 먼저 음절 개수를 추출한 후, 분할된 분할 영역으로부터 모음열을 검출한다. 검출된 모음으로부터 CV, VCCV, VC 기준모델의 결합에 의해 후보 모델을 구성한 후, 인식하도록 하였다.[3]

2.1 음절 개수 추출

음절 개수 추출 방법은 기준모델 작성시 후보의 개수에 큰 영향을 미치기 때문에 중요한 과정이다. 먼저 입력 데이터로부터 에너지와 zerocrossing rate을 이용하여 유성음과 무성음영역을 검출한다. 또한 좀더 정확한 유성음 영역을 검출하기 위해 제 1 포먼트, 제 2 포먼트가 존재하는 215Hz와 2,756Hz사이의 에너지 정보 VB 파라미터를 이용하여 안정된 유성음 영역을 추출한다. VB 파라미터의 포락은 비교적 완만하게 변화하는 모음 영역을 검출에 유효하다. 따라서 VB 파라미터의 시간에 따른 미분 정보를 이용하여 좀더 정확한 음절분할에 이용하였다. 그림 3은 /5212345/의 음성데이터의 음절분할 결과를 나타내었다.

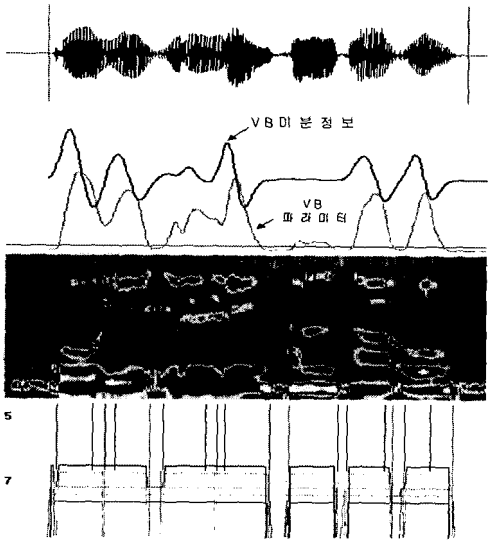

그림 3 /5212345/ 분할 결과

먼저 에너지와 Zerocrossing rate, VB를 이용하여 분할된 음절의 개수는 5개가 나타났으며, VB의 미분정보를 이용하여 /오/와 /이/사이, /일/과 /이/사이가 분할하여 총 음절의 개수는 7개가 나타났다.

2.2 모음열 추출

모음열을 추출하기 위해 /삼/, /사/, /팔/은 /아/모음군에 /공/, /오/, /육/, /구/는 /오/모음군에, /일/, /이/, /칠/은 /이/모음군으로 구별하였다. 분할된 음절영역을 24개의 주파수 영역으로 분할한 후, 각각 주파수 영역에 대해 각 모음군의 에너지 분포를 성명데이터, 단음절데이터, PBW가 포함된 TDB, PRW가 포함된 FDB로부터 추출하였다.[3]

음절 개수 추출 후에도 추출되지 않는 연속된 숫자열인 /이이/, /오오/는 시간 정보를 이용하여 두 개의 영역으로 분할하도록 하였고, 숫자음 /팔/인 경우에는 /아/모음 다음에 유성종성자음 /ㄹ/이 /이/모음군으로 맵핑되기 때문에 규칙을 적용하여 한 음절안에 끊김없이 /아/, /이/ 올경우에는 뒤쪽에 위치한 /이/모음군을 삭제하도록 구성하였다. 또한 모음열이 짧은 시간의 /이/모음군과 /오/모음군이 연속적으로 발생될 경우에는 숫자음 /육/의 발생된 것으로 하여 단어 결합 HMM의 수를 줄이도록 하였다.

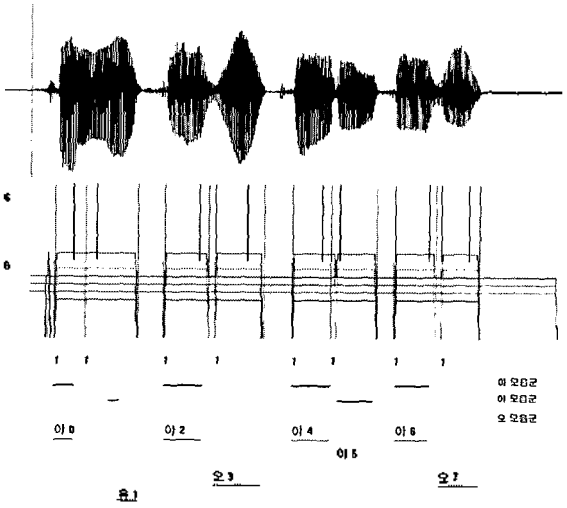

그림 4 숫자음 /36458245/ 모음군 추출 결과

그림 4는 숫자음 /36458245/를 음절개수를 8개로 분할한 후, 모음군 추출한 결과를 나타낸 것으로써 /6/은 /이/와 /오/모음군이 연속으로 나타냈기 때문에 /유/모음군으로 결과로 맵핑된 것을 보여 주고 있다.

2.3 모음군에 따른 기준 모델 구성

모음열 추출 결과에 따라 숫자음 기준 모델을 결합하도록 하여 후보의 수를 줄여 인식 속도를 향상하도록 구성하였으며, 그림 5는 모음열 /이오육/에 따른 기준 모델구성의 결과를 나타내었다.

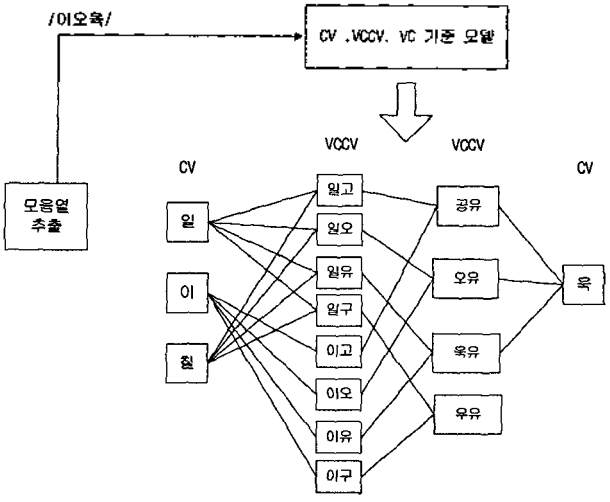

그림 5 모음열 /이오육/ 기준모델 구성

세 음절의 단어모델은 1,000개가 필요하지만, /이오육/의 모음열을 이용해서 기준모델을 구성할 경우에는 12($3\times4\times1\times1$)개만 필요하기 때문에 인식시간에서 상당한 단축을 가져올 수 있게 되었다.

3. 실험 및 결과

본 장에서는 모음열 정보를 이용한 기준모델 구성에 따른 연속숫자음 인식 시스템의 유효성을 검증하기 위해 화자 5명이 모든 조합이 고려한 4연속 숫자음 35개를 이용하여 4음절의 음절 개수 추출 정확도, 모음열 검출 정확도, 인식률, 인식 시간을 실험하였으며, 인식 시간과 인식률의 비교 대상은 4연속 숫자음의 기준모델 10,000개와 평가하였다.

표 2 음절개수 및 모음열 추출 정확도

화자	음절개수 정확도	모음열추출 정확도	모음열검출 오류 숫자음
1	34/35	33/35	5732,7954
2	35/35	33/35	5732,5267
3	35/35	31/35	6843,5267, 6378,4750
4	34/35	31/35	6843,5861, 5267,4750
5	35/35	33/35	5732,4156
평균	98.8%	92%	

모음열 검출 오류가 발생한 목록을 살펴보면 /칠/, /육/이 포함된 연속숫자음에 나타난 것을 알 수 있다. /칠/인 경우는 /이/와 /일/의 숫자음에 비해 주파수 대역이 자음 /ㅊ/ 때문에 많이 다른 것을 알 수 있었으며, /육/인 경우에는 /이/와 /오/의 연속된 모음열이 검출되어야 하지만, 자연스럽게 발성할 경우 /오/의 발음이 약하게 발성되면서 빨리 변화하기 때문에 검출을 하지 못하여 오류가 발생된 것으로 판단되어진다.

표 3에는 모음열 정보를 이용한 시스템과 10,000개의 후보를 둔 시스템과의 연속숫자음 인식 결과를 나타내었다. 인식 결과를 살펴보면 모음열을 이용한 시스템이 21.8%의 인식률이 향상됨을 알 수 있다. 즉 모음열을 적용하여 인식후보를 줄임으로써 인식 성능이 향상됨을 알 수 있었다.

표 4 인식 결과

화자	모음열 정보	10,000개 후보
1	26/35	20/35
2	26/35	19/35
3	28/35	21/35
4	28/35	18/35
5	28/35	20/35
평균	77.8%	56%

4. 결론

본 논문에서는 모음열을 적용하여 4연속 숫자음 인식 시스템을 구현하였다. 모음열 정보를 이용하여 인식후보의 수를 줄였기 때문에 인식 시간과 인식률 면에서 좋은 성능이 나타남을 알 수 있었다. 인식 성능의 개선을 위해서는 모음열 정보를 추출할 때 좀더 세밀한 규칙을 적용하게 되면 더 나은 인식률을 가져다 줄 수 있을 것이며, 모든 모음열 후보를 두고 인식을 하는 것이 아니라, 모음군에 따라 인식된 음절만 다음 후보로 사용하게 되면 인식 속도의 향상을 기대할 수 있을 것이다.

참고 문헌

[1] 김순협 외 4인, "음소 단위에 의한 한국어 연속 숫자음 인식에 관한 연구,""한국음향학회지 VOL.8 No. 3, 1989.

[2] 윤재선, 홍광석, "반음절 단위 HMM을 이용한 연속 숫자 음성인식," 한국음향학회지 제17권 제5호,1998.

[3] 윤재선, 홍광석, "무제한 단어인식 시스템을 위한 VCCV분할에 관한 연구," 한국음향학회 학술발표대회 논문집 제 19권 제1호, 2000.