SCHMM 기반 7연속 숫자음 인식에 관한 연구

김 세 용 *, 정 회 석 *, 강 철 호 *
광운대학교 전자통신공학과

A Study on 7-Connected Digits Speech Recognition using SCHMM

Se Yong Kim *, Hui Seok Jung *, Chul Ho Kang *
Dept. of Electronic Communication Eng., Kwangwoon Univ.
kseyong@explore.gwu.ac.kr

요 약

본 연구에서는 우리말 연속 숫자음 인식에서 본래의 숫자음을 변이 시키는 주된 요인인 연음현상에 대한 인식을 높이기 위해 별도의 연음부분의 레퍼런스를 작성하여 매칭 시키는 방식을 제안한다. 또한 단모음으로 이루어진 /2/와 /5/의 연속된 음에 대하여도 레퍼런스를 작성하였다.

제안한 방식에 의하여 전체적으로 1.4%정도 인식률이 상승됨을 볼 수 있다. 특히 발성 목록중 /82/, /62/, /31/, /15/, /75/ 등의 연음과 /226/, /755/등과 같이 모음의 연속된 발성이 포함된 숫자 열에서 제안된 방식이 인식률에 영향을 미치는 것을 볼 수가 있었다. 이는 연음에서 발생하는 오류가 연속 숫자음에 많은 영향을 미치는 것을 알 수 있다. 그 외에 /22/, /55/등과 같이 단모음으로 이루어진 숫자음의 연속 발성 또한 인식률을 저하시키는데 한 요인으로 작용함으로서 이에 대한 레퍼런스도 작성하여 인식률이 상승되는 것을 볼 수 있었다.

1. 서 론

본 연구의 목적은 현재 맨-머신 인터페이스로 각광 받고 있는 음성 인식 분야에서 대표적인 어플리케이션이라고 할 수 있는 음성 다이얼링 시스템과 본인의 주민등록번호확인과 신용카드 숫자의 인식과 더 나아가서 음성인식의 최첨단 기술인 화자인식기술을 적용하여 그것을 통한 화자인식을 병행하기 위한 선행작업인 7자리의 연속 숫자음 인식의 성능 향상을 목표로 한다.

한국어 숫자 음은 대부분 단음절이고 초성이 'ㅇ'인 단어가 많아 연속 발음시 영어 숫자 음에 비해 연음 현상을 많이 받는다. 이 연음 현상은 조음 현상과 더불어 오인식을 유발하는 대표적 문제점으로 단음절 발성 시 나타나지 않는 변형을 발생시킨다. 이를 극복하기 위하여 본 논문에서는 각 숫자 음끼리 발생할 수 있는 연음부분에 대한 레퍼런스를 따로 작성하여 숫자 음 차례로 인식시마다 이 연음정보를 적용시켜서 인식률의 향상을 꾀하였다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 연속 숫자음 인식 시스템에 발생하는 우리말 음향-음성학적 특성을 서술하였으며, 제 3 장에서는 구성한 연속 숫자 음 인식 시스템에 대하여 그리고 제안한 시스템에 대한 실험 과정을 기술한다. 마지막으로 제 4 장에서는 모든 결과에 대한 결론을 맺는다.

2. 우리말 숫자음의 음향-음성학적 특성

우리말은 숫자음이 가지는 음향학적 특성을 고려해 본다면 이것은 조음단계(articulatory event)로 설명이 가능할 것이다. 일반적으로 조음 현상은 보통 묵음과 모음, 모음과 묵음, 자음과 모음, 모음과 자음, 모음과 모음 사이에서 발생하는데 자음과 자음 사이에서는 거의 음향

학적으로 중요한 변이는 발생하지 않는 것으로 알려져 있다.

이러한 경우의 모든 예를 가정하여 정상적인 발성 이외의 결과를 정리하면 표 1 과 같이 나타낼 수 있다.

표 1. 우리말 숫자음에서의 동시 조음 혼.과

Table 1. Mis-recognition syllable and g·oup in Korean Digits

대상숫자	조음 결합		조음결과
0(공),9	0,1,2,3,4,5,7,9	+ 0(공),9	유성음화('ㄱ')
	6	+ 0(공),9	경음화(꽁,꾸)
0(영)	1,7,8	+ 0(영)	령
	0,3	+ 0(영)	녕
	6	+ 0(영)	경, 넝
1,2	1	+ 1,2	릴, 리
	3	+ 1,2	밀, 미
3,4	1,6,7,8	+ 3,4	경음화(쌈,싸)
7,8	7,8	+ 1,2,5	ㄹ(l→r)
6	6	+ 1,2,5	유성음화('ㄱ')
	2,4,5,9	+ 6	륙(rjug)
	1,7,8	+ 6	륙(ljug)
	6,0,3	+ 6	늑(njug)
	6	+ 6	융(jung)
	2,4,5,9	+ 6 + 6	륭(rjung)
	1,7,8	+ 6 + 6	륭(ljung)
	0,3	+ 6 + 6	늉(njung)

같은 0(공)과 9라고 해도 0(영), 1, 2, 3, 4, 5, 7, 9등과 같은 모음으로 끝나거나 혀옆소리'ㄹ'로 끝나게 되면 0(공)과 9는 10/일공/과 같이 유성음화'ㄱ'으로 발성되고, 6과 같이 닫음소리'ㄱ'뒤에 오면 /꽁/, /꾸/와 같이 경음화되어 발성된다. 1, 7, 8뒤에 0(영)이 오게 되면 'ㄹ'머리소리규칙에 의하여 /령/으로 발성되고, 0(영), 3 뒤에 오면 'ㄴ'머리소리 규칙에 의해 /녕/으로 발성된다. 1뒤에 오는 1과 2는 /릴/과 /리/ 로 발성되는데 이는 'ㄹ'겹치기 규칙에 의한 것이다. 3뒤에 오는 1, 2는 /밀/과 /미/로 발성되고 이는 자연스러운 연음현상이다. 특히, 1, 6, 7, 8 후에 3, 4가 오는 경우 /쌈/과 /싸/로 발성되는데 이는 경음화 현상으로 볼 수 있으며, 7, 3후에 1, 2, 5가 오면 /릴/, /리/, /로/등과 같이 연음 현상이 발생한다. 6의 경우 다양하게 발성이 되는데 2, 4, 5, 9뒤에 오는 6의 두둘김 소리 /륙/과 1, 7, 8, 후에 오는 혀옆소리 /뤽/, 그리고 6뒤에 오는 6/융/을 고려하였다.

특히 연음 현상은 조음 현상과 더불어 오인식을 유발하는 대표적 문제점으로 단음절 발성 시 나타나지 않는 변형음을 발생시킨다.

한국어 숫자음은 영어 숫자음과 달리 모두 단음절이다. 또한, 10개의 숫자음 중에서 6개의 숫자음 (/공/, /일/, /삼/, /육/, /칠/, /팔/)이 종성 받침을 갖고 있고, 4개의 숫자음(/일/, /이/, /오/, /육/)의 초성이 'ㅇ'으로 선행 음절 종성 받침의 영향을 받는다. 따라서, 한국어 숫자음은 연속 발음되었을 때, "35" (/사모/)등과 같은 연음 현상이 매우 많이 발생하고, 발음 과정에서 원래의 10개의 숫자음에는 없는 /모/, /미/, /밀/, /고/, /기/, /길/, /로/, /러/, /릴/ 등과 같은 발음이 추가로 발생한다. 이것은 한국어 연결숫자음 인식을 어렵게 만드는 가장 큰 문제 중에 하나로 이 문제를 고려하지 않고서는 인식률의 개선을 기대하기 어렵다.

3. 실험 및 결과

제안한 연속 숫자음 인식 시스템은 특징 파라미터로 PLP와 동적 PLP 벡터를 사용하였고, 연속 숫자음 인식 알고리즘은 OPDP(One-Pass Dynamic Programming)알고리즘을 사용하였다.

기존의 연속 숫자음 인식 시스템은 추출된 특징 파라미터를 SCHMM의 패턴매칭 방식을 통하여 연속 숫자음 인식 알고리즘인 OPDP를 통하여 인식된 후 후처리를 거쳐 인식되었다.

그러나 제안한 연속 숫자음 인식 시스템은 OPDP알고리즘에 음절의 변형을 쉽게 가져오는 연음현상에 대한 정보를 따로 패턴매칭을 하여 연속 숫자음에 대한 인식률을 향상시킬 수 있었다.

각각 숫자음의 표준패턴은 연속 음성으로부터 연결어 인식을 위한 표준 패턴을 추출하기 위한 것이다. 연음현상(/밀/, /미/, /릴/, /리/, /로/, /경/)과 경음화 현상(/쌈/, /싸/), 'ㄹ'머리소리 규칙(/령/), 그리고 'ㄹ'겹치기 규칙(/릴/, /리/) 등의 조음현상을 고려하여 28가지의 음절 레퍼런스를 작성하였다. 이중 /꽁/, /꾸/와 같은 경음화 현상과 /녕/과 같은 'ㄴ'머리소리 규칙 그리고 6의 유성음화 등은 본 논문에서 제외 시켰다. 학습에 이용된 발성

목록은 한국어 숫자음에서 발생 가능한 조음현상이 고루 포함되어 있는 한국 전자통신 연구소에서 작성한 35종류 4연속 숫자음 리스트를 통하여 작성되었다. 표준패턴은 20대 남성화자 10인으로부터 각각 1회씩 발성되었으며, 이를 수작업을 통하여 음절별로 세그멘테이션을 한 후 각각의 음절단위의 레퍼런스를 작성하였다. 구성한 음절 레퍼런스는 표 2 와 같다.

표 2. 조음현상을 고려해 선정된 단어모델
Table 2. Selected word models with co-articulation

0	영	형	경	
1	길	릴	일	일
2	기	리	이	이
3	삼	쌈		
4	사	싸		
5	고	로	모	오
6	륙	육	용	
7	칠	치		
8	팔	파		
9	구			
0	공			

본 논문에서 다룬 음절단위의 숫자음 중 연음현상이 발생될 수 있는 숫자음은 종성이 있는 /일/, /삼/, /육/, /칠/, /팔/ 이다. 이 숫자음뒤에 오는 'ㅇ'으로 시작되는 /일/, /이/, /오/, /육/ 이 연음현상을 발생시킬 수 있는 결합이다. 그러나 /일/+/일/, /일/+/이/, /칠/+/육/, /팔/+/육/은 'ㄹ'겹치기 규칙에 의하여 /일릴/, /일리/, /칠륙/, /팔륙/으로 발성되고, /삼/+/육/은 'ㄴ'머리소리 규칙에 의하여 /삼뉵/으로 발성된다.

이를 제외한 /일/+/오/, /삼/+/일/, /삼/+/이/, /삼/+/오/, /육/+/일/, /육/+/이/, /육/+/오/, /칠/+/일/, /칠/+/이/, /칠/+/오/, /팔/+/일/, /팔/+/이/, /팔/+/오/에 관한 연음에 대한 레퍼런스를 작성하였다. 이미/로/, /밀/, /미/, /모/, /길/, /기/, /고/, /릴/, /리/ 등이 존재하므로 새로이 작성된 연음 레퍼런스는 연음현상이 일어나는 선행 숫자음의 중성, 종성과 후행 숫자음의 중성의 결합으로 이루어진다. 구성한 연음레퍼런스는 표 3 과 같다.

표 3. 제안한 연음 레퍼런스
Table 3. Proposed word model with prolongation

연음	연음레퍼런스
/일/+/오/	/ㅣ-ㄹ-ㅗ/[i-r-o:]
/삼/+/일/	/ㅏ-ㅁ-ㅣ/[a-m-i:]
/삼/+/이/	
/삼/+/오/	/ㅏ-ㅁ-ㅗ/[a-m-o:]
/육/+/일/	/ㅠ-ㄱ-ㅣ/[yu-k-i:]
/육/+/이/	
/육/+/오/	/ㅠ-ㄱ-ㅗ/[yu-k-o:]
/칠/+/일/	/ㅣ-ㄹ-ㅣ/[i-r-i:]
/칠/+/이/	
/칠/+/오/	/ㅣ-ㄹ-ㅗ/[i-r-o:]
/팔/+/일/	/ㅏ-ㄹ-ㅣ/[a-r-i:]
/팔/+/이/	
/팔/+/오/	/ㅏ-ㄹ-ㅗ/[a-r-o:]

추가로 연음이외에 모음의 연속된 숫자음인 /이이/, /오오/등에 대하여도 별도의 레퍼런스를 작성하였다. 이 숫자음은 하나의 모음으로만 이루어져 있고, 연속으로 2번 이상 발성될 경우 포만트의 변화도 거의 없으므로 /이이/가 /이/로 인식되는 경우가 종종 발생하게 된다.

실험에 사용된 음성 데이터의 샘플링 주파수는 11.025kHz 이고 전처리 과정은 12차 PLP 와 12차 Delta PLP 특징벡터 사용하였고, 음성을 처리하는 한 Frame의 구간은 21.7msec로서 240개의 샘플이 사용되고, 64개의 코드워드로 구성된 코드북 벡터를 사용하였다. 그리고 이 알고리즘을 통해 각각 코드워드의 공분산을 구해 반연속 HMM의 가우시안 분산값으로 사용된다. 본 실험에서 사용된 인식 모델은 반연속 HMM이고 각 음절당 상태 개수를 3개로 하고 Left to right 모델이 적용되었다.

본 논문에서는 학습에 참여한 10인의 남성 화자의 새로운 7연속 숫자음으로 인식 실험을 하였다. 숫자음 인식 시스템의 인식 결과는 81.0% 인식률을 보인다.

주목할 만한 점은, 발성 목록중 '1282262'와 '2315755'에서 볼 수 있듯이 /82/, /62/, /31/, /15/, /75/ 등의 연음과 /226/, /755/등과 같이 모음의 연속된 발성이 포함된 숫자열에서 인식률이 저조함을 알 수 있다. 이는 연음 발생시에 /칠/, /팔/, /일/의 'ㄹ'이 본래 혀옆소리 'ㄹ'[l]에서 /이륙/, /치릴/등의 두들김소리 'ㄹ'[r]로 변화하고, /육/의 닫음소리'ㄱ'[gq]이 /유기/, /유고/등이 되면서 울림소리'ㄱ'[gg]이 되기 때문인 것으로 관찰된다. 그리고 /이이/, 혹은/오오/등의 단모음으로 이루어진 숫자음의 연결은 포만트나 음향적인 특징의 변화가 없으므로 /이

이/를 /이/로 오인식하는 것으로 나타났다.

제안한 방식에서는 전체 인식률은 82.4%이며, 연음에 대한 숫자음이 두 번이상 포함된 발성 목록에서 인식률이 유지되거나 상승됨을 볼 수 있다. 이는 본 논문에서 제안한 연음의 정보가 변화하는 음운에 적절히 작용을 하였다고 보아지나 7자리의 인식률은 그다지 높게 나타나지 않는 것을 볼 수 있다. 그 이유는 인식률이 하락한 발성목록에서 발견할 수 있는데 이는 우리말 특성에서 종종 발견되는 /칠/이 /일/로 인식되고, /구/가 /오/로, /이/가 /일/로 그리고 /삼/이 /사/로 오인식되었기 때문이라고 관찰된다.

제안한 방식에 의하여 전체적으로 1.4%정도 인식률이 상승됨을 볼 수 있다. 특히 발성 목록중 /82/, /62/, /31/, /15/, /75/ 등의 연음과 /226/, /755/등과 같이 모음의 연속된 발성이 포함된 숫자열에서 제안된 방식이 인식률에 영향을 미치는 것을 볼 수가 있었다. 이는 연음에서 발생하는 오류가 연속 숫자음에 많은 영향을 미치는 것을 알 수 있다. 그 외에 /22/, /55/등과 같이 단모음으로 이루어진 숫자음의 연속 발성 또한 인식률을 저하시키는데 한 요인으로 작용함으로서 이에 대한 레퍼런스도 작성하여 인식률이 상승되는 것을 볼 수 있었다. 결국 우리말 숫자음에서 인식률 향상을 위해서 할 수 있는 과제는 조음현상에 대한 깊은 연구와 그것에 대비한 완벽한 레퍼런스 작성에 있음을 알 수 있다.

표 4. 사용자에 따른 인식률 비교

Table 4. Comparison of recognition rate of connected digits for users

User	Baseline	Proposed
1	80.0	84.3
2	84.2	80.7
3	79.2	84.3
4	82.1	80.0
5	80.0	79.3
6	80.7	82.9
7	80.0	82.1
8	81.1	82.1
9	78.6	83.6
10	83.6	83.6
Total	**80.0**	**82.4**

4. 결 론

본 논문에서 제안한 방식은 연속 숫자음 인식에서 중요한 레퍼런스 작성에 우리말 숫자음에서 발생할 수 있는 연음에 대한 정보를 새로이 추가하였다. 기존의 연속 숫자음 연구에서 알 수 있듯이 가장 문제가 되었던 부분은 우리말에서 특히 많이 발생하는 연음현상을 비롯한 각종 조음현상에 따른 인식의 어려움이었다. 이중 특히 연음현상은 본래의 숫자음을 변이시키는 주된 요인으로 작용한다. 여기에 착안하여 본 논문에서는 연음부분의 레퍼런스를 별도로 작성하여 매칭시키는 방식을 제안하였다. 이를 통해 연음이 많이 포함된 연속 숫자음에 대하여는 성능이 향상되거나 유지됨을 확인하였다.

앞으로의 연구과제는 전화나 핸드폰을 이용한 각종 은행의 비밀번호인식과 개인의 신용카드 번호 혹은 주민등록 번호 인식 등에 실용할 수 있는 연구와 음성의 인식을 통한 개인의 화자정보를 탐색할 수 있는 화자인식에 대한 연구이다.

참고문헌

[1] X.D Huang, Y. Ariki and M.A. Jack, Hidden Markov Models for Speech Recognition, Edinburgh University Press, 1990.

[2] 양진우, 김순협 "HMM과 연결 숫자음의 후처리를 이용한 음성 다이얼링에 관한 연구" 한국음향학회지, 14권 5호, 1995

[3] Dynamic Programming Search for Continuous Speech Recognition, IEEE Signal Processing Magazine, September, 1999

[4] 윤재선, 홍광석, "반음절 단위 HMM을 이용한 연속 숫자음 음성인식", 한국음향학회지 17권 5호, 1998

[5] L.R.Rabiner, J. G. Wilpon and F.K. Soong, "High performance connected digit recognition, using hidden Markov models," IEEE Trans. on ASSP, pp1214-1225, Vol37, No.8. Aug. 1989

[6] 남기심, 고영근, 표준 국어문법론, 탑출판사, 88-89, 1996.