

반음절 문맥종속 모델을 이용한 한국어 4 연숫자음 인식에 관한 연구

A Study on Korean 4-connected Digit Recognition Using Demi-syllable Context-dependent Models

이 기 영*, 최 성 호*, 이 호 영**, 배 명 진***
(Ki Young Lee*, Seong-Ho Choi*, Ho-Young Lee**, Myung Jin Bae***)

* 관동대학교 정보통신공학과, ** 서울대학교 언어학과, *** 송실대학교 정보통신공학과
(접수일자: 2002년 10월 17일; 수정일자: 2003년 1월 10일; 채택일자: 2003년 2월 10일)

한국어 숫자음은 단음절이며 연결된 숫자음 사이에 연음현상의 영향 때문에 한국어 연결 숫자음의 인식방법으로 반음절에 기반한 모델들이 제시되어 왔다. 기존에 제안된 반음절이나 반음절+반음절의 인식모델을 이용한 방법에서는 아직까지 우수한 인식성능을 보이지 못하고 있다. 본 논문에서는 확장된 문맥종속 반음절 모델을 이용한 한국어 4 연숫자음 인식방법을 제안한다. 실험에서 연결숫자음은 SiTEC의 4 연숫자음 데이터 베이스를 사용하였으며 학습과 인식방법으로는 HTK 3.0의 C-HMM을 이용하였다. 기존의 방법들과 인식율을 비교해 본 결과, 92%의 비교적 우수한 인식성능을 보였다.

핵심용어: 한국어 4 연숫자음, 연속 HMM, 문맥종속모델, 문맥독립모델

투고분야: 음성처리 분야 (2.5)

Because a word of Korean digits is a syllable and deeply coarticulated in connected digits, some recognition models based on demisyllables have been proposed by researchers. However, they could not show an excellent recognition results yet. This paper proposes a recognition model based on extended and context-dependent demisyllables, such as a tri-demisyllable like a tri-phone, for the Korean 4-connected digits recognition. For experiments, we use a toolkit of HTK 3.0 for building this model of continuous HMMs using training Korean connected digits from SiTEC database and for recognizing unknown ones. The results show that the recognition rate is 92% and this model has an ability to improve the recognition performance of Korean connected digits.

Keywords: Korea 4-connected digits, Continuous HMM, Demisyllable, Context-dependent model, Context independent model

ASK subject classification: Speech signal processing (2.5)

I. 서론

음성은 인간에게 가장 자연스럽게 효과적인 정보교류 수단이며 이에 관하여 1950년대부터 음성의 발전과 이해에 대한 많은 기초적 연구가 수행되어 왔다. 최근에는 개인용 컴퓨터 보급의 가속화와 컴퓨터에 의한 신호처리 기술과 정보처리기술의 급속한 발전과 더불어 음성을 통한 인간과 기계와의 효율적인 통신을 위한 대어휘 연속

음성인식 시스템에 대한 연구가 활발하게 진행되고있다 [1]. 특히 컴퓨터 통신을 이용한 정보 및 금융 서비스에서 주민등록번호, 비밀번호, 통장번호, 회원번호 등의 사용이 확대됨에 따라 키보드 입력뿐만 아니라 음성인식 등의 수단에 의해서 확인할 필요성이 증가하고 있다[2,3].

한국어의 숫자음은 단음절이며, 연결 숫자음은 단음절이라는 짧은 간격의 단일 숫자음들로 구성되었기 때문에 연음 현상이 크다. 이에 따라 음소를 인식모델로 하여 한국어 숫자음을 인식하는 경우, 인식단위 경계가 부정확하게 분할되어 오인식의 원인이 된다. 그러나 단음절로 구성된 숫자음을 전반부와 후반부 반음절로 분배하는

책임저자: 이기영 (kylee@kwandong.ac.kr)
215-802 강원도 양양군 양양읍 임천리 산7
관동대학교 정보통신공학과
(전화: 033-670-3413; 팩스:033-670-3413)

반음절을 인식모델로 하면 인식단위 경계의 부정확성을 감소시켜 인식성능을 향상시킬 수 있다. 이에 따라 한국어 연결 숫자음의 특징을 반영하기 위한 인식모델로 반음절 또는 반음절+반음절을 이용한 연구가 이루어져 왔으며 보다 향상된 인식률을 얻기 위하여 반음절기반의 인식을 수행한 뒤 후처리로 재인식을 수행하는 연구가 시도되었다[4,5].

본 논문에서는 반음절을 확장하여 반음절-반음절+반음절의 단위를 문맥중속 인식모델로 하는 한국어 연결 숫자음 인식방법을 제안하고자 한다. 지금까지의 인식모델에서는 반음절에 기반한 인식모델의 범위를 문맥독립으로 하거나 인접한 한 개의 반음절까지 확장하였다면 큰 연구에서는 연결 숫자음의 연속된 연음현상을 고려하여 한 개의 반음절을 기준으로 인접한 전후 양측의 반음절까지 문맥중속으로 인식모델을 확장하여 사용한다. 또한 그 성능을 평가하기 위하여 기존에 제안되었던 음소와 반음절 기반의 문맥독립과 문맥중속 모델을 이용한 인식 결과들과 비교 검토한다.

본 논문의 구성은 서론에 이어 2 장에서는 확장된 문맥중속 반음절 모델에 관하여 기술하고, 3 장에서는 실험에 사용된 데이터 베이스 및 실험 결과에 대해서 기술하였으며, 마지막으로 4 장에서 결론을 맺는다.

II. 확장된 반음절 인식모델

2.1. 반음절 문맥독립 인식모델

최근 음성인식기에서 가장 널리 사용되는 음향모델은 은닉 마코프 모델(HMM: Hidden Markov Model)에 기반한 것이다. 음향모델에 사용되는 기본 인식단위로는 음소(phoneme), 반음절(demi-syllable), 음절(syllable), 단어(word) 등이 있다. 음소를 인식단위로 사용하면 음향모델의 개수는 적으나 인접한 음소의 변화에 따른 연음현상을 반영하지 못하는 단점이 있다. 이를 개선하기 위

하여 다이폰(diphone)이나 트라이폰(triphone)을 사용함으로써 음소 전후간의 연음현상을 반영하여 대어휘 연속음성인식에 사용하고 있다.

그러나 본 연구에서 인식대상으로 하는 한국어 숫자음의 경우 음소를 인식의 기본단위로 하면 서론에서도 언급되었듯이 한국어 숫자음은 단음절로 이루어져 있고, 연음현상에 의해 각 숫자음들의 고유발음이 변하게 되어 인식률이 저하된다.

한국어 연결 숫자음의 인식성능을 향상시키기 위한 인식단위로는 반음절이나 반음절+반음절의 인식단위가 제시되었다[4,5]. 인식할 단어가 숫자음과 같이 단음절인 경우, 각 음절을 전반부와 후반부 반음절로 분배하는 반음절을 인식모델로 하여 인식단위 경계의 정확성을 향상시키려는 것이다. 즉, 7/칠/과 같이 단음절로 이뤄진 한국어 숫자음을 반음절로 구분하면, 전반부 반음절 '치'와 후반부 반음절 '일'로 나뉘게 된다.

그림 1은 숫자 '칠'을 전반부와 후반부 반음절로 구분한 파형을 보이고 있다. 표 1에는 한국어 숫자음 0/공/부터 9/구/까지의 10개 숫자음들을 인식단위의 모델로 사용하게 될 전반부와 후반부 반음절을 보이고 있다.

표 1. 반음절로 나뉜 인식단위

Table 1. Recognition units for two demisyllables of Korean digits.

digit	forward demisyllable	backward demisyllable
공	go	ong
일	i	il
이	y	yy
삼	sa	am
사	s	a
오	o	oo
육	u	uk
칠	ch	eil
팔	pa	al
구	ku	uoo

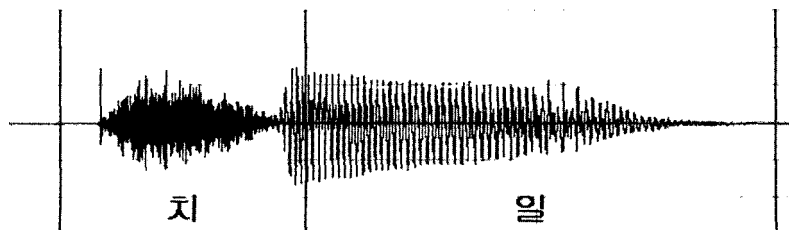


그림 1. '칠'을 반음절로 구분한 모습

Fig. 1. Demisyllable-segmented waveform of '칠'.

표 2. 두 자리 숫자음에 대한 확장된 반음절의 문맥종속 인식모델
 Table 2. Context-dependent model of extended demissyllable for two digits.

digit	0 demissyllable	1 demissyllable	2 demissyllable	3 demissyllable
일어	silence-i+il	i-il+y	il-y+yy	y-yy+silence
삼사	silence-sa+am	sa-am+s	am-s+a	s-a+silence
오육	silence-o+oo	o-oo+u	oo-u+uk	u-uk+silence

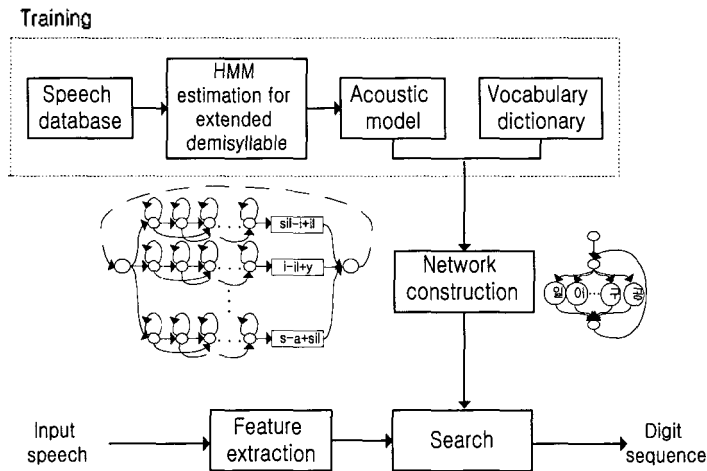


그림 2. 반음절 문맥종속 모델을 이용한 연결숫자음 인식과정
 Fig. 2. Recognition procedure using demissyllable context-dependent model.

2.2. 확장된 반음절 문맥종속 인식모델

기존의 반음절 모델에서 반음절+반음절의 인식단위는 연습현상의 문제를 고려하여 하나의 반음절에서 인접한 한 개의 반음절을 조합적으로 확장하였다면, 본 논문에서 제안하고자 하는 확장된 문맥종속 반음절 모델은 한 개의 반음절을 중심으로 전후 양측의 반음절을 모두 고려하여 반음절-반음절+반음절 인식단위의 문맥종속 모델이다. 구체적으로 연결 숫자음의 시작과 끝 부분은 각각 무음 (silence)-반음절+반음절과 반음절-반음절+무음으로 인식모델을 형성하고, 그 외의 중간 숫자음에 대해서는 인접한 전후 반음절이 추가된 반음절-반음절+반음절의 인식모델을 형성하므로써 음운변화 현상을 반영할 수 있도록 확장된 문맥종속 모델이다. 표 2는 연결 숫자음을 확장된 반음절의 문맥종속 모델로 표현한 예를 보이고 있다.

본 논문에서는 표 2와 같이 반음절-반음절+반음절의 인식단위를 사용하였으며, 연결 숫자음의 인식되는 과정을 그림 2에 보이고 있다. 먼저 음성 데이터를 사용하여 특징을 추출한 다음 학습과정에서 확장된 반음절 문맥종속으로 모델링하고 추정된 음향모델과 발음사전을 이용하여 네트워크를 구성한다. 연결 숫자음 인식과정에서는 음성 신호가 입력됨에 따라 가장 확률이 높은 경로를 찾게 된다.

또한 10개의 숫자음들이 연속 숫자음으로 연결될 때 발생할 수 있는 모든 확장된 반음절의 문맥종속 모델의 개수는 각 숫자음당 22개이다. 1) 10개의 숫자음을 인식대상으로 하므로 모든 인식단위의 문맥종속 모델 개수는 모두 220개이다.

III. 실험 및 결과

본 논문의 연결 숫자음 인식 실험에서는 HTK 3.0의 연속 HMM[7,8]을 이용하여 음소 기반의 문맥독립 모델과 문맥종속 모델, 그리고 반음절 기반의 문맥독립 모델과 제안된 반음절의 문맥종속 모델에 대해서 각각 실험을 하였으며, state수에 따른 인식률의 변화를 비교 고찰하였다.

1) 반음절-반음절+반음절의 모델에서 첫 번째와 두 번째의 반음절이 하나의 숫자음을 이룰 경우, 조합될 수 있는 세 번째 반음절은 모든 숫자음 10개의 전반부 반음절이거나 무음이므로 모델 개수는 11개 이다. 또한 반음절-반음절+반음절의 단위에서 두 번째와 세 번째의 반음절이 하나의 숫자음인 경우, 조합될 수 있는 첫 번째 반음절은 모든 숫자음 10개의 후반부 반음절이거나 무음이므로 모델 개수는 11개 이다. 따라서 각 숫자음에 따라 조합될 수 있는 반음절-반음절+반음절 모델의 개수는 22개이다.

표 3. 발성 데이터 목록

Table 3. List of spoken data.

4-connected digits (35 words)									
1	0287	8	0712	15	8877	22	9176	29	0316
2	5732	9	5267	16	3510	23	3045	30	7083
3	9601	10	6633	17	8065	24	8590	31	8194
4	4156	11	2409	18	2934	25	5500	32	9205
5	1199	12	7954	19	7489	26	6972	33	1427
6	1398	13	1823	20	2244	27	5861	34	2538
7	6843	14	6378	21	4621	28	3649	35	4750

표 4. 사용된 파라미터 계수

Table 4. Parameter coefficients.

Pre-emphasis Constants	0.97
Window type	Hamming
Window Size	25 ms
Frame Rate	10 ms
NiFilterbank Channels	26
Energy + 12 MFCC	13
Delta coefficients	13
Acceleration coefficients	13

3.1. 데이터 베이스

실험에 사용된 음성 데이터는 16 KHz로 샘플링하고 16 Bits로 양자화한 SiTEC의 데이터베이스를 사용하였으며 70명의 남녀화자가 방음실에서 35개의 4연결 숫자음에 각각 4회씩 발성한 데이터 (70명*35개*4회=9800개)로 구성되어있고 연령분포는 20대와 30대가 약 87%를 이루고 있다.

실험에서 사용한 학습데이터는 여성화자 22명과 남성화자 28명으로 이루어진 총 50명의 4연결 숫자음성 데이터 (50명*35개*4회=7000개)를 사용하였으며, 인식실험을 위한 인식 데이터로는 위의 50명을 제외한 20명 (남성화자 10명과 여성화자 10명)의 4연결 숫자음성 데이터 (20명*35개*4회=2800개)를 사용하였다.

4연결 숫자음에 대한 발성 데이터 목록은 각 숫자가 연속해서 '영'부터 '구'까지 나타날 수 있는 모든 경우를 포함한 것으로써 표 3과 같다.

3.2. 음향 모델 생성

가) 특징 벡터 추출

특징 벡터는 16 KHz로 샘플링하고 16 Bits로 양자화

2 본 실험에서 사용한 SiTEC 연결숫자음 데이터 베이스의 발성화자는 여성화자 32명, 남성화자 38명이며 연령은 최하 19세에서 49세 까지 분포되어 있다.

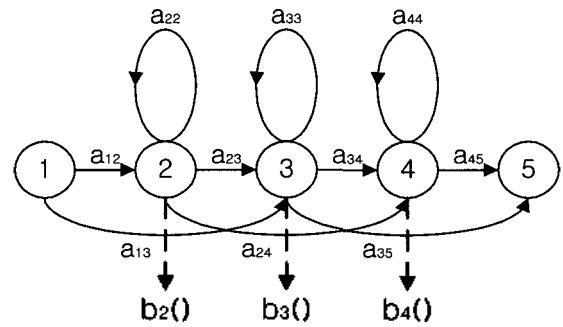


그림 3. 5상태인 Left-Right 모델
Fig. 3. Left-to-right model for 5 states.

된 음성 데이터에서 Mel Frequency Cepstral Coefficients (MFCC)를 추출하였다. 특징 파라미터에 사용된 계수들은 표 4와 같다.

(나) 은닉 마코프 모델

반음절 모델의 확률분포모델로는 연속확률 분포를 사용[7,8]하였고, 모델의 형태는 그림 3과 같은 도약 경로가 존재하는 5 상태에서부터 16 상태에 이르는 총 12가지의 상태수를 가진 left-to-right 모델로 Single Gaussian Mixture를 사용하였다.

그림 3에서 state가 5인 HMM은 1개의 entry state (state 1)와 1개의 exit state (state 5), 그리고 3 개의 emitting state (state 2,3,4)로 이루어져 있다. 그리고 1, 5번의 state는 각각 entry, exit state로 non-emitting state이다. non-emitting state는 output probability distribution을 가지지 않는다. 단지 모델간을 연결할 때 상태 변화에만 관여하는 state들이다.

3.3. 인식 결과

본 논문에서는 state 변화에 따른 인식률을 두 가지로 나누었다. 한가지는 자릿수에 따른 인식률이고, 다른 한가지는 전체 숫자음의 인식률이다. 예를 들어, '일일삼

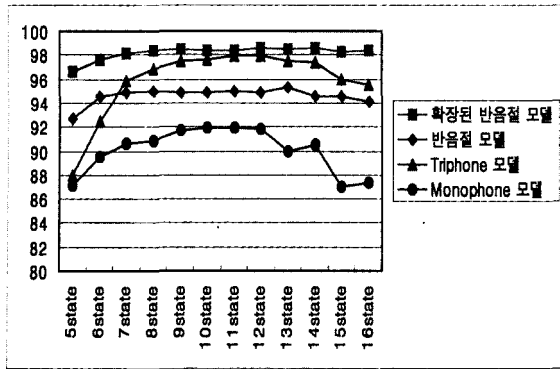


그림 4. 자릿수에 따른 인식률 비교
Fig. 4. Accuracy comparison of each digit.

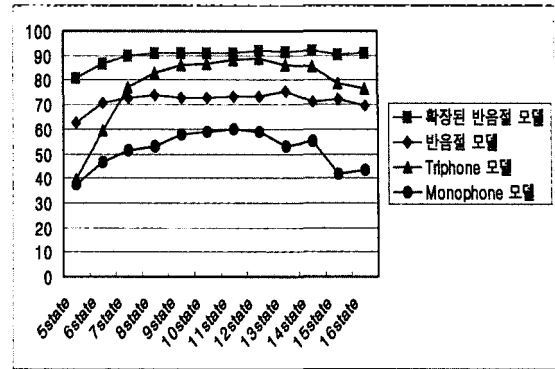


그림 5. 전체 숫자음에 대한 인식률 비교
Fig. 5. Accuracy comparison of whole digits.

사'의 연결 숫자음을 인식하고자 할 때, 인식된 결과가 '일이삼사'였다면 자릿수에 따른 인식률과 전체 숫자음의 인식률은 모두 100%가 된다. 그러나 인식된 결과가 '이이삼사'로 4자리 숫자음중 1자리 숫자가 오인식되었다면 자릿수에 따른 인식률은 75%이고, 전체 숫자음의 인식률은 0%이다.

표 5와 그림 4는 연결 숫자음의 자릿수에 따른 인식률을 나타내었다. 인식모델의 종류는 문맥독립 모델인 monophone 모델에서의 인식률과 반음절 모델에서의 인식률 그리고, 문맥종속 모델인 트라이폰 모델과 본 논문에서 제안한 확장된 반음절 모델이다. 또한, 표 6과 그림 5는 문맥독립과 문맥종속 모델에 따른 전체 숫자음의 인식율을 비교하고 있으며, 본 실험의 인식률은 소수점 셋째 자리에서 반올림하였다.

(가) 문맥독립 모델

monophone 모델과 반음절 모델에 대한 인식 결과를 비교해 보면, 모든 state에 대해 음소를 기반으로 한 monophone 모델보다는 반음절을 기반으로 한 반음절 모델에서의 인식결과가 우수함을 볼 수 있다. 또한 표 6의 전체 숫자음 인식률을 보면 monophone 모델에서는 11 state에서 59.86%의 가장 좋은 인식결과를 얻었으며, 반음절 모델에서는 11state에서 73.22%의 가장 좋은 인식결과를 얻었다.

표 6의 전체 숫자음 인식률에서 monophone 모델과 반음절 모델간의 인식률 차이를 보면 5state에서 24.64%, 6state에서 24.21%, 7state에서 21.14%로 상태수가 증가할수록 두 모델간의 인식률 차이는 점차 감소하다가 11state에서 13.36%로 가장 근소한 인식률의 차이를 보인다. 그리고 다시 상태수가 증가할수록 두 모델간의 인식

표 5. 자릿수에 따른 인식률
Table 5. Recognition accuracy of each digit.

number of state	context-independent model		context-dependent model	
	monophone	segmental	triphone	extended segmental
5	87.08	92.70	88.00	96.66
6	89.53	94.51	92.41	97.62
7	90.55	94.87	95.88	98.19
8	90.80	95.01	96.86	98.41
9	91.73	94.83	97.47	98.42
10	91.91	94.85	97.59	98.41
11	91.90	94.93	97.87	98.38
12	91.83	94.89	97.94	98.60
13	89.98	95.32	97.52	98.52
14	90.48	94.51	97.36	98.54
15	86.97	94.59	95.98	98.27
16	87.29	94.05	95.57	98.40

표 6. 전체 숫자음 인식률

Table 6. Recognition accuracy of whole digits.

number of states	context-independent demisyllable model		context-independent demisyllable model	
	monophone	demisyllable	triphone	extended demisyllable
5	37.68	62.32	39.61	80.97
6	46.29	70.50	59.18	86.61
7	51.36	72.50	77.15	89.82
8	52.68	73.57	82.89	91.00
9	58.00	72.61	86.07	91.11
10	58.68	72.79	86.75	90.90
11	59.86	73.22	88.36	91.00
12	59.07	73.04	88.75	92.04
13	53.00	75.43	86.32	91.68
14	55.07	71.22	85.79	91.79
15	41.75	72.11	78.86	90.29
16	43.14	69.50	76.68	90.96

의 차이는 점차 증가하는 것을 볼 수 있다.

이는 앞서서도 언급되었듯이 한국어 숫자음의 경우 소리의 반음절로 이루어져있기 때문에, 반음절을 기본단위로 사용할 경우 음소보다 모델링하는 구간이 넓으므로 음향적 특성을 잘 모델링할 수 있음을 확인할 수 있었다.

(나) 문맥종속 모델

트라이폰 모델과 확장된 반음절 모델에 대한 인식 결과 비교해 보면, 모든 state에 대해 문맥 독립 모델인 경우와 같이 음소를 기반으로 한 트라이폰모델보다는 반음절 모델을 기반으로 한 확장된 반음절 모델에서의 인식결과가 우수함을 볼 수 있다. 또한 표 6의 전체 숫자음 인식률에서 트라이폰모델은 12state에서 88.75%의 가장 좋은 인식결과를 얻었으며, 확장된 반음절 모델은 12state에서 92.04%의 가장 좋은 인식결과를 얻었다.

표 6의 전체 숫자음 인식률에서 트라이폰모델과 확장된 반음절 모델간의 인식률 차이를 보면 5state에서 4.36%, 6state에서 27.43%, 7state에서 12.67%로 상태수가 증가할수록 두 모델간의 인식률 차이는 점차 감소하다가 11state에서 2.64%로 가장 근소한 인식률의 차이를 보인다. 그리고 다시 상태수가 증가할수록 두 모델간의 인식률의 차이는 점차 증가하는 것을 알 수 있다.

그러나 트라이폰모델에서 5state의 인식률과 12state의 인식률을 보면 49.14%의 현저한 차이가 나는 것을 알 수 있고, 확장된 반음절 모델에서 5state의 인식률과 12state의 인식률을 보면 11.07%의 인식률 차이가 나는 것을 알 수 있었다.

이는 한국어 연결 숫자음에서 문맥종속 모델인 트라이폰

모델과 확장된 반음절 모델은 5개의 state를 설정하여 파라미터 값을 계산하기에는 상태변화가 잘 표현되지 못하여 인식률의 저하를 보인다. 그러나 state를 증가할수록 트라이폰모델과 확장된 반음절 모델에 대한 상태변화가 잘 표현되어 인식률이 향상되는 것을 알 수 있다. 하지만 13state에서부터 state수가 증가할수록 인식률이 낮아지는 것을 알 수 있는데 이것은 지나친 상태변화에 따라 오히려 인식률의 저하를 보이며 13state 이상부터는 훈련용 데이터로 사용한 음성 데이터의 양이 상태수 13개 이상부터를 모델링 하기에는 부족하기 때문으로 판단된다.

3.4. 고찰

이상의 실험결과에서 한국어 연결 숫자음의 인식단위로는 음소단위보다 반음절단위가 더 우수한 것을 확인하였다. 음소단위는 숫자음 중에서 '0', '3', '7', '8' / '공', '삼', '칠', '팔' / 등은 각각 한 음절이면서 3개의 음소로 표현되므로 각 숫자음을 인식하기에 용이한 반면, 숫자음 '2', '5' / '이', '오' / 등은 한 음절이면서 1개의 음소로 표현되므로 한국어 숫자음을 음소단위로 인식하기에 적절하지 못하며 음소단위를 이용한 모델에서 인식율을 저하시키는 요인이 될 수 있다. 그러나 반음절 단위는 한국어 숫자음을 각각 전반음절과 후반음절의 2개로 표현하므로 1개의 음절로 구성되는 한국어 숫자음들의 특징들을 포함하기 위한 인식단위로 적절하며 인식율도 음소단위를 이용한 모델보다 향상시킬 수 있다.

또한 문맥독립 모델과 문맥 종속 모델에 대한 인식 결과에서 전후의 숫자음을 고려하지 않은 문맥독립 모델에 비해 문맥종속 모델의 성능이 우수함을 알 수 있다. 이는

연속적으로 발음되는 숫자들은 전후의 숫자음에 의해 연음영향을 받아 그 특성이 변화하기 때문이며, 전후 숫자에 따라 모델을 다르게 두는 것이 그 특성을 나타내는데 적합함을 나타낸다. 한국어 연결 숫자음 인식에서 음소를 기본단위로 한 모델과 반음절을 기본단위로 한 모델의 인식 결과에서 반음절을 기본단위로 한 모델의 성능이 우수할 뿐만아니라 문맥종속인 경우가 더 우수함을 알 수 있었다. 이는 한국어의 숫자음은 단음절이고 일반적으로 발음길이 짧기 때문에 음소단위보다는 반음절단위가 한국어 연결 숫자음의 음향적인 특징을 잘 반영하는 것으로 사료된다.

IV. 결론

본 논문에서는 기존에 제안된 반음절 모델을 이용하여 연결 숫자음을 인식하는 방법보다 인식성능을 향상시키기 위하여 확장된 반음절의 문맥종속 모델을 이용한 연결 숫자음 인식방법을 제안하였다.

본 논문에서 제안한 확장된 반음절의 문맥종속 모델을 이용하여 인식성능을 개선시킬 수 있음을 확인하기 위하여, SiTEC의 4 연숫자음 데이터 베이스를 대상으로 문맥 독립 모델인 monophone 모델과 반음절 모델, 그리고 문맥 종속 모델인 트라이폰 모델과 본 논문에서 제안한 확장된 문맥종속 반음절 모델을 사용하여 인식실험을 수행하였다. 인식 실험 결과, monophone 모델에서는 59.86%의 인식율을 얻었으며 반음절 모델에서는 75.43%, 트라이폰 모델에서는 88.75%, 그리고 본 논문에서 제안한 확장된 문맥종속 반음절 모델에서는 92.04%의 인식율을 얻을 수 있었다.

한국어 연결 숫자음의 인식률은 다른 인식 대상의 음성에 비하여 상대적으로 저조한 인식률을 보이고 있다. 이에 대해 본 논문에서는 기본인식단위로 반음절 단위를 사용하여 단음절로 이루어진 한국어 숫자음의 음향적인 특징을 잘 반영할 수 있도록 하였으며 확장된 반음절의 문맥종속 모델을 이용하여 각 숫자음을 기준으로 전후에 나타날 숫자음의 전반부 후반부의 반음절 특징과 silence를 모두 포함하고 있기 때문에 숫자음 사이에서 발생하는 연음현상을 비교적 충분히 고려하고자 하였다.

본 논문에서는 연결숫자인식을 위한 효율적인 숫자 모델링 방법에 초점을 맞추었으며, 키보드의 입력이 아닌 음성을 이용한 연결 숫자음 인식 시스템에 적용이 가능하고, 차후 한국어 대어휘 연속 음성인식 시스템에 적용이 가능하리라 생각된다. 앞으로 좀더 다양한 연결 숫자음

에 대해서도 실험을 수행하고 또한 전화음성과 잡음이 섞인 음성 및 다른 특징 파라미터를 이용한 연결 숫자음에 대한 인식 실험을 비교 분석하고자 한다.

감사의 글

본 연구는 한국과학재단 특정기초연구 (과제번호 R01-2002-000-00278-0)의 지원에 의하여 이루어졌음.

참고 문헌

1. P. C. Woodland, C. J. Leggetter, J. Odell, V. Valtchev and S. J. Young, "The 1994 HTK large vocabulary speech recognition system," *Proc. ICASSP*, 1, 73-76, 1995.
2. 양진우, 김순협, "HMM과 연결 숫자음의 후 처리를 이용한 음성 다이얼링에 관한 연구" *한국음향학회지*, 14 (5), 74-82, 1995.
3. J. G. Wilpon and L. R. Rabiner, *Improvement in connected Digit Recognition Using Higher order Spectral and Energy*, 1991.
4. 윤재선, 홍광석, "반음절 단위 HMM을 이용한 연속 숫자 음성 인식" *한국음향학회지*, 17 (5), 73-78, 1998.
5. 정재부, 정훈, 정익주, "반음절기반의 한국어 연속 숫자음 인식과 그 후처리에 대한 연구" *음성과학*, 8 (4), 1-15, 2001.
6. 김기성, 김승희, 김형순, 지민제, "한국어 연결숫자인식을 위한 숫자 모델링에 관한 연구" 제15회 음성통신 및 신호처리 워크샵, 15 (1), 293-297, 1998
7. S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev and P. Woodland, *The HTK Book Version 3.0*, Microsoft, 2000.
8. L. R. Rabiner, "A tutorial on Hidden Markov Models and selected application in speech recognition," *Proc. of the IEEE*, 77 (2), 257-286, 1989.
9. Young, Steve, 1996, "Large Vocabulary Continuous Speech Recognition: a Review," Technical report, Cambridge University Engineering Department, Cambridge, UK.
10. J. C. Junqua and J. P. Haton, *Robustness in Automatic Speech Recognition*, Kluwer Academic Publishers, Boston, 1996.

저자 약력

● 이 기 영 (Ki Young Lee)

한국음성과학회지 제7권 제6호 참조

● 최 성 호 (Seong-Ho Choi)

2001년 2월: 관동대학교 정보통신공학과 (공학사)

2003년 2월: 관동대학교 정보통신공학과 (공학석사)

* 주관심분야: 음성인식, 디지털신호처리

● 이 호 영 (Ho-Young Lee)

현재: 서울대학교 언어학과 교수

● 배 명 진 (Myung Jin Bae)

현재: 숭실대학교 정보통신공학과 교수 (제14권 1E호 참조)