

# FSN과 반음절쌍 모델을 이용한 연결 숫자음 인식의 성능 향상에 관한 연구

서은경, 최태웅, 김순협

광운대학교 컴퓨터공학과

## A Study on Improvement of the Connected Digit Recognition Using Finite State Network and Demi-Syllable Pair Models

Eun-kyoung Seo, Tae-woong Choi, Soon-hyob Kim

Dept of Computer Engineering, Kwangwoon University

seksek@korea.com, dami\_73@hotmail.com, kimsh@daisy.kw.ac.kr

### 요 약

본 논문에서는 숫자음과 단위음으로 구성된 한국어 연결 단위숫자음 인식의 성능 향상을 위하여 한국어 연결 단위숫자음의 특징을 분석하였다. 한국어의 단위숫자음은 숫자음 한음절과 단위음 한음절로 구성된 두음절의 연속적이고 반복적인 발성으로 나타난다. 숫자음에서의 인식 대상 어휘는 숫자음이라는 제한된 규칙을 갖는 가변 숫자음이다. 따라서 개수, 금액, 단위량, 거래량 등에서 나타날 수 있는 가변 숫자음을 인식하기 위하여 FSN(Finite State Network)을 구성하였다. 음향 모델은 한국어 숫자음과 같이 발성구간이 짧은 어휘의 연결음(connected word)의 인식에서 효과적인 반음절쌍(demi-syllable pair) 모델을 이용하였다. 실험결과, 화자 독립적인 가변 숫자음 60문장의 테스트 데이터에 대해서 문장 인식률 91.0%로 인식 성능을 향상시킬 수 있었다.

### 1. 서론

음성인식 기술의 발전과 함께 다양한 분야에서 이를 이용한 서비스가 이루어지고 있다. 특히 전화망을 통한 증권거래, 쇼핑 등과 같은 거래 형태의 서비스에서는 주민등록번호, 비밀번호, 전화번호 등의 인식에 쓰이는 연결 숫자음의 인식 못지않게 개수, 금액, 단위량, 거래량 등에서 나타나는 연결 단위숫자음의 인식도 중요시 되고 있다.

본 논문에서는 한국어 연결 단위숫자음의 인식

성능향상을 위하여 단위숫자음의 특징을 분석하여 문법적인 제약을 준 FSN을 구성하였다. 또한 연결음(connected word)의 인식에서 해결해야 하는 단어 경계가 어디에서 이루어지는가의 문제를 해결하기 위하여 모음의 안정구간을 기준으로 앞뒤로 이분하여 VCCV (vowel-consonant-consonant-vowel), VCV (vowel-consonant-vowel) 형태를 취하는 반음절쌍(demi-syllable pair) 모델을 이용하였다.

본 논문에서는 한국어 연결 숫자음의 하나인 ‘삼만사천구백오십’, ‘이백구십사’, ‘팔천이백육십삼’ 등의 개수, 금액, 단위량, 거래량 등을 표기하기 위한 연결

숫자음을 연결 단위숫자음이라 칭하였다.

## 2. 한국어 연결 단위숫자음의 특징

이 절에서는 개수, 금액, 단위량, 거래량 등을 표현하기 위한 한국어 연결 단위숫자음의 특징에 대해서 살펴본다.

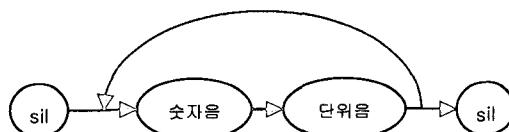
연결 단위숫자음은 숫자음과 단위음으로 이루어진 단어로 구성된 문장이다. 단어는 한음절의 단위음과 한음절의 숫자음이 번갈아가며 나타난다. 즉, 연속된 숫자음 혹은 연속된 단위음으로 나타나지 않는다.

숫자음은 ‘일, 이, 삼, 사, 오, 육, 칠, 팔, 구’로 구성된 9개의 발음이 인식 대상이 된다.

단위음은 ‘십, 백, 천, 만, 억’으로 구성된 5개의 발음이 인식 대상이 된다. 연결 단위숫자음 문장에 있어서 맨 마지막의 단위음은 연결 단위숫자음의 용도가 개수일 경우 ‘개’, 금액일 경우 ‘원’, 넓이 단위량일 경우 ‘평’, 증권거래에서의 주식 거래량일 경우 ‘주’와 같은 단위음으로 구성된다.

[표 1] 증권거래용 연결 단위숫자음의 분석

숫자음	일, 이, 삼, 사, 오, 육, 칠, 팔, 구
단위음	십, 백, 천, 만, 억, 주, 원



[그림 1] 한국어 연결 단위숫자음 문장의 특징

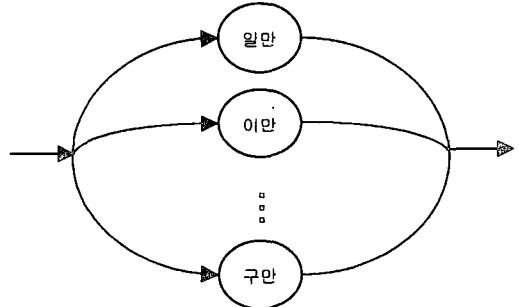
## 3. 제안된 연결 단위숫자음 인식을 위한 방법

앞에서 분석한 연결 단위숫자음의 특징을 이용하여 두음절 기반의 FSN을 제안한다. 또한 발성 구간이 짧은 숫자음의 특성을 반영하여 음소보다 넓은 구간을 모델링하고 단어간 조음현상을 효과적으로

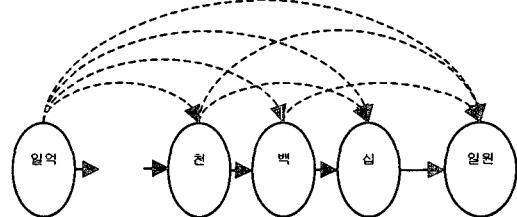
다루며, 연결음(connected word)의 인식에 있어서 단어의 경계 구분의 문제를 보완할 수 있는 반음절쌍(demi-syllable pair) 모델을 사용하였다.

### 3.1 연결 단위숫자음 FSN(Finite State Network)

연결 단위숫자음 문장의 분석 결과 문장을 이루는 단어가 ‘숫자음 한음절 + 단위음 한음절’로 구성된 두음절의 단어이며 문장 전체에서 반복되어 나타나는 것을 바탕으로, [그림 2]와 같이 숫자음 한음절과 단위음 한음절로 구성된 두음절의 단어를 FSN node로 구성하여 인식단위를 두음절의 단어로 하였다. 문장 전체의 FSN은 [그림 3]과 같이 나타난다. [그림 3]의 하나의 node는 [그림 2]와 같은 sub-network로 나타난다.



[그림 2] 제안된 FSN의 sub-network 구조



[그림 3] 제안된 연결 단위숫자음 문장의 FSN

### 3.2 반음절쌍(demi-syllable pair) 모델

연속음성인식에서는 다이폰이나 트라이폰 모델을 사용함으로써 음소의 앞뒤간에 조음현상을 반영하는 방식을 많이 사용한다. 그러나 본 논문에서는 음소보다 넓은 구간을 모델링하고 단어간 조음효과를 효과적으로 반영하며 연결음 인식에 있어서 단어

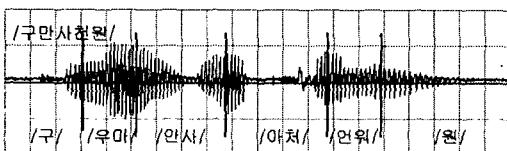
의 경계를 구분하기 어려운 문제를 보완하기 위하여 지속기간이 길게 나타나는 모음 구간을 경계로 음절을 양분하여 앞 음절의 뒷부분과 뒤 음절의 앞부분을 하나의 모델로 하는 반음절쌍(demi-syllable pair) 모델을 사용하였다. 인식가능범위 숫자음 ‘일’부터 ‘9천 9백9십9억9천9백9십9만9천9백9십9’ 까지의 숫자음을 인식하기 위해 반음절쌍은 100개, 반음절은 28개의 모델이 나타난다. [표 2]에서 본 논문에서 사용한 데 이터베이스의 반음절쌍 모델을 만들기 위한 인식단위를 표현하고 있다.

{
 S : 음절의 시작 부분. E : 음절의 끝 부분.  
 D : 숫자발음열 표기. J : 주, W : 원.  
 T : 십, B : 백, C : 천, M : 만, A : 억.

[표 2] 제안된 반음절쌍(demi-syllable pair) 모델

단어	인식단위 모델		
구	S9D	E9D	
구만	S9D	E9SM	EM
만사	SM	EMS4D	E4D
사천	S4D	E4DSC	EC
천원	SC	ECSW	EW

[그림 4]는 본 논문에서 실제 테스트 데이터로 쓰인 ‘구만사천원’의 발성에 대하여 반음절쌍 모델의 예를 보인 것이다.



[그림 4] ‘구만사천원’을 반음절쌍으로 구분한 모습

### 3.3 단위숫자음 FSN과 반음절쌍 모델의 구현

한국어 연속 단위숫자음 인식에서 음향모델을 반음절쌍으로 하였을 경우 두음절의 FSN node와 함께 구현하기 위한 방법으로 [그림 5]와 같은 인식 과

정을 거친다.



[그림 5] 반음절쌍 모델 연결 단위숫자음 인식 과정

음향모델이 VCCV, VCV 형태의 반음절쌍으로 구성되므로, 연결 단위숫자음 문장의 시작부분과 끝부분은 반음절의 형태로 나타난다.

## 4. 실험 및 결과

실험은 증권거래에 이용되는 숫자음인 ‘원’ 단위 발성의 금액과 ‘주’ 단위 발성의 거래량 숫자음에 대해서 실증하였다.

음성 데이터베이스를 구축하기 위해 [표 3]과 같이 두음절로 구성된 데이터베이스를 구성하여 남녀 25명을 대상으로 각각 한 명의 화자가 2번씩 발성하여 녹음하였다.

[표 3] 음성 데이터베이스의 구성

단어형태(음절수)	개 수
단위음(1)	7
숫자음(1)	9
숫자음(1) + 단위음(1)	45
단위음(1) + 숫자음(1)	45
단위음(1) + 단위음(1)	10
숫자음(1) + ‘원’	9
숫자음(1) + ‘주’	9
합 계	154

테스트 데이터는 인식범위인 ‘일’부터 ‘9천 9백9십9억9천9백9십9만9천9백9십9’ 까지의 인식범위 안의 숫자음 중 증권거래에 이용되는 범위 내에서 인식 단

위 단어가 다양하게 분포하도록 하여 금액 단위숫자 음 30문장, 거래량 단위숫자음 30문장, 총 60문장을 표 5와 같이 작성하고 모델에 참여하지 않은 5명의 화자가 자연스럽게 2회 발성하도록 하여 사용하였다.

[표 5] 테스트 데이터 문장

구만사천원 칠만오천오백육십이원 오만칠천오백삼십구원 ⋮ 오십육만구천이백삼십이원 이만삼천이백칠십원	구만사천삼백오십사주 칠만오천주 삼만사천오백이십팔주 ⋮ 삼만칠천오백오십오주 육만팔천팔백오십육주
---	--

실험결과 [표 6]과 같이 인식단위를 두음절로 구성하였을 때가 한음절로 구성하였을 때보다 문장인식 성능이 향상되었음을 볼 수 있었다.

[표 6] FSN과 음향모델에 따른 문장 인식률(%)

FSN 음향모델	한음절		두음절	
	단어	문장	단어	문장
tri-phone	88.4	29.2	89.1	59.7
syllable	99.2	67.7	94.2	75.8
demi-syllable	99.2	85.5	96.4	87.1
demi-syllable pair	84.6	21.0	97.6	91.0*

FSN node를 두음절로 하였을 경우 문장인식률이 최고 30.5%의 인식성능 향상이 있었다. 인식단위의 경계가 한쪽만 모음인 CV(consonant-vowel), VC(vowel-consonant) 형태의 모델인 demi-syllable 모델은 syllable 모델에 비해 11.3%의 인식성능 향상을 보였고, 보다 넓은 구간을 모델링하고 양쪽 경계가 모두 모음인 VCCV, VCV 형태의 모델인 demi-syllable pair 모델은 demi-syllable 모델보다 3.9%의 인식 성능 향상을 보였다. demi-syllable pair 모델의 인식단위를 한음절로 하였을 경우 인식단위와 음향모델간의 부조화로 인

식률이 현저히 떨어지게 나타났다.

인식 결과를 살펴보면, 증권거래에서 이용되는 금액 단위 숫자음인 '원' 단위 연결 단위숫자음 문장에서 '오원' 이 '원'으로 오인식 되거나, 반대로 '원'이 '오원'으로 오인식 되는 경우가 많았다. 또한 '백원'이 '백구원'으로 오인식 되는 경우도 많았다.

## 5. 결론

본 논문에서는 연결 숫자음의 하나인 한국어 연결 단위숫자음 인식의 성능 향상을 위해 두음절의 FSN을 제안하고 반음절쌍(demi-syllable pair) 음향모델을 이용하여 실험하였다. 실험결과 FSN node를 두음절로 구성하였을 경우 tri-phone, syllable, demi-syllable, demi-syllable pair 모델 모두 인식 성능이 향상되었음을 알 수 있었다. 또한 인식단위를 두음절로 한 demi-syllable pair 모델의 경우 단어 인식률 97.6%, 문장 인식률 91.1%의 향상된 결과를 얻었다.

## [참고문헌]

- [1] C. H. Lee, and L. R. Rabiner, " A Frame-Synchronous Network Search Algorithm for Connected Word Recognition", IEEE Trans. on Acoust., Speech, and Signal Processing, vol. 37, no. 11, Nov. 1989
- [2] Yong, Steve, "Large Vocabulary Continuous Speech Recognition : a Review." Technical report, Cambridge University Engineering Department, Cambridge, UK.
- [3] Hiromi FUJII, Masao WATARI, Hiroaki SAKOE, and Seibi CHIBA "Connected Digit Speech Recognition by suing DEMI-WORD pair Reference Pattern(DWPR)", ICASSP 86, TOKYO, p1073~1076
- [4] 이두성, "어휘독립 음성인식 시스템의 구현", 성균관대학교 박사학위논문, 2000
- [5] 윤재선, 퉁광석, "반음절 단위 HMM을 이용한 연속 숫자 음성인식", 한국음향학회지, 제17권 제5호, p73-78, 1998