# 한국어 숫자음에서의 음운변화를 고려한 연결숫자 인식의 성능향상

송명규, 김형순
부산대학교 전자공학과

# Performance Improvement of Connected Digit Recognition by Considering Phoneme Variations in Korean Digit.

Myung Gyu Song, Hyung Soon Kim
Dept. of Electronics Eng., Pusan National Univ.
E-mail: {mgsong, kimhs}@hyowon.pusan.ac.kr

## 요 약

한국어 숫자는 각 숫자가 단음절로 이루어져 있으며, 연속적으로 발음될 때 인접 숫자들의 상호조음현상에 의해 각 숫자의 고유 발음이 변화하고, 또한 그 숫자들의 경계도 모호해지는 문제점이 있다. 한편 연속적인 숫자의 발성을 기대하는 인식시스템에 반하여 일부 사용자는 숫자들을 고립시켜서 발성하기도 한다. 이는 연결숫자의 음운현상만을 고려한 인식 시스템에서는 성능저하의 한 원인이 된다. 본 논문에서는 연결숫자의 인식성능 향상을 위해서 한국어 숫자들의 음운 변화를 고려하여 변이음군을 정하였으며, 사용자의 여러 가지 발성형태에 따른 다양한 음운 현상의 변화를 흡수 할 수 있도록 인식 네트웍을 구성하는 방식을 검토하였다. 전화망 4연숫자음을 이용한 화자독립 인식실험을 통해서 한국어 숫자에서 자주 오인식되는 '이', '오', '일' 인식 성능이 각각 4.2%, 4.2%. 2.9%씩 향상되었으며, 인식속도도 33%의 개선이 있었다

## 1. 서 론

연결 숫자인식은 음성 다이얼링, 증권거래, 은행업무 서비스 등의 다양한 응용분야가 있으나, 한국어 숫자의 특성에 따른 문제 및 배경잡음이나 채널에 의한 왜곡에 따른 문제점들로 인해 서비스에 만족할 정도의 성능을 얻지 못하는 것이 현실이다. 본 논문에서는 배경잡음이나 채널왜곡 등에 의한 환경불일치 문제의 완화 보다는 한국어 숫자의 특성에 기인한 연결 숫자인식의 어려움을 완화시킴으로써 인식성능을 향상시키고자 한다. 한국어 숫자는 인접 숫자에 의한 영향으로 다양한 변이음이 발생하는데, 이를 고려하여 변이음군을 정의하고, 사용자의 발성형태에 따른 다양한 음운 현상의 변화를 흡수 할 수 있도록 연결 숫자인식 네트웍을 구성한다.

본 논문의 구성은 다음과 같다. 2절에서는 한국어 숫자의 특성에 대해 살펴보고, 3절에서 음운변화를 고려한 인식네트웍 구성에 대해 설명한다. 4절에서는 실험 결과를 언급하고, 5절에서 결론을 맺는다.

## 2. 한국어 숫자의 특징 분석

한국어의 숫자는 '영(零),일(一),이(二),삼(三), ...' 처럼 한자어로 된 것과 '하나, 둘, 셋, 넷, 다섯, ... '같이 순수 한국어로 된 것 두 가지가 있다[1]. 번호나 돈의 단위(원), 시간표시(년,월,일, ...) 등에는 한자어 계통이 주로 쓰이므로 본 논문에서는 한자어 계통의 '영, 공, 일, 이, 삼, 사, 오, 육/륙, 칠, 팔, 구'의 11개 숫자에 국한하여 그 특징을 살펴본다.

한자어 계통의 숫자들은 모두 하나의 음절로 이루어져 있으며, 더구나 '일'과 '이', '일'과 '칠', '삼'과 '사' 같이 음절의 일부분의 차이로 인해서 구별되는 숫자가

존재하므로 음성인식에 문제점으로 작용한다. 또한 연결 숫자 인식의 경우처럼 연속적으로 발음되는 숫자열에서는 숫자의 경계가 모호해질 뿐만 아니라, 인접한 숫자들의 상호조음현상에 의해 각 숫자들의 고유한 발음이 변화하므로 음성인식의 혼동가능성을 높인다. 위 숫자들은 'ㄱ, ㄹ, ㅁ, ㅅ, ㅇ, ㅊ, ㅍ'의 7개 자음과 'ㅏ, ㅕ, ㅗ, ㅜ, ㅠ, ㅣ'의 6개 모음으로 구성되지만, 이들이 말소리로 실현이 될 때에는 음운규칙에 따라 다양한 변이음들로 나타난다. 숫자음에서 고려된 음운규칙은 닫음소리 되기, 두들김소리 되기, 울림소리 되기 등이다. 음운규칙이 음소의 변이음을 실현시키는 규칙이라 한다면, 한 형태소의 음소가 그 놓이는 환경에 따라 다른 음소로 바뀌는 현상도 발생하는데 이러한 음소의 바뀜 규칙을 변동 규칙이라 한다[2]. 숫자음에 적용되는 변동 규칙은 소리 이음, 된소리되기, 'ㄹ' 머리소리 규칙, 'ㄴ' 머리소리 규칙, 'ㄹ'의 'ㄴ'되기, 'ㄹ' 겹치기, 콧소리 되기 등이다. 이러한 변동 및 음운규칙을 적용하여 숫자음에 나타나는 변이음들을 표 1에 정리 하였다.

표 1. 숫자음에 나타나는 변이음.

| 변이음/기호/ | 예 | 변이음 | 예 |
|---|---|---|---|
| ㄱ /g/ | 9/구/ | ㅇ /N/ | 0/영/ |
| 닫음소리 ㄱ /gq/ | 6/육/ | ㅊ /c/ | 7/칠/ |
| 울림소리 ㄱ /gg/ | 20/이공/ | ㅍ /p/ | 8/팔/ |
| ㄲ /G/ | 69/육꾸/ | ㅏ /a/ | 4/사/ |
| ㄴ /n/ | 36/삼뉵/ | ㅕ /v/ | 0/영/ |
| 혀옆소리 ㄹ /l/ | 1/일/ | ㅗ /o/ | 5/오/ |
| 두들김소리 ㄹ /r/ | 26/이륙/ | ㅜ /u/ | 9/구/ |
| ㅁ /m/ | 3/삼/ | ㅠ /ju/ | 6/육/ |
| ㅅ /s/ | 4/사/ | ㅣ /i/ | 2/이/ |

한편 한국어 숫자에는 '이'와 '일', '일'과 '칠', '오'와 '구' 등의 오인식이 잘되는 숫자쌍들이 존재하는데, 이들의 변별력을 높이기 위한 한가지 방법으로 변이음 설정과정에서 이를 고려한다. 그림 1 ~ 3에서 이 숫자음들의 음향특징을 관찰할 수 있다. 그림들은 각 각 '5189', '5289', '5789'의 파형과 스펙트로그램 이다. '일'과 '칠'의 'ㅣ'는 '이'의 'ㅣ'에 비해 제2포만트가 비교적 낮음을 알 수 있으며 이는 혀옆소리 'ㄹ'의 영향에 기인한 것으로 판단된다. 또한 하나의 홑소리로 음절을 형성하는 '이'의 'ㅣ'가 '일'과 '칠'의 'ㅣ'보다 지속시간이 비교적 길다는 것을 알 수 있다. 또한 '칠'의 'ㅊ'은 스펙트럼상에 스파이크와 잡음낀 듯한 특성을 나타낸다. 인지적으로도 '일'과 '칠'에서의 'ㅣ'와 '이'의 'ㅣ'는 음소의 시작부분의 세기나 길이, 고저 등에 있어서 차이가 있음을 알 수 있으므로 이를 고려하여 'ㅣ'를 혀옆소리 'ㄹ' 앞의 'ㅣ'(이후 /ii/로 언급)와 그 외의 'ㅣ'(이후 /i/로 언급)로 구분한다.

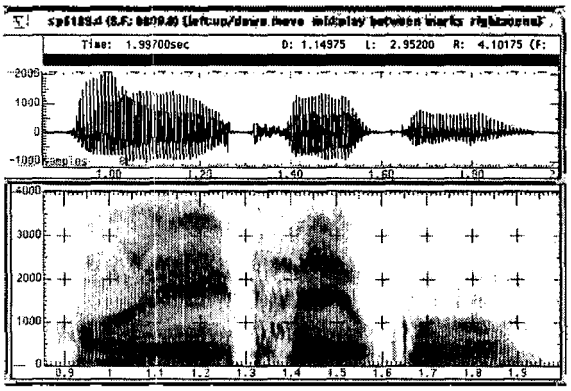

그림 1. '5189'의 파형과 스펙트로그램.

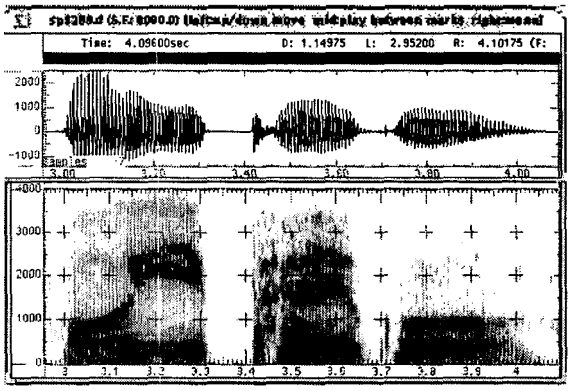

그림 2. '5289'의 파형과 스펙트로그램.

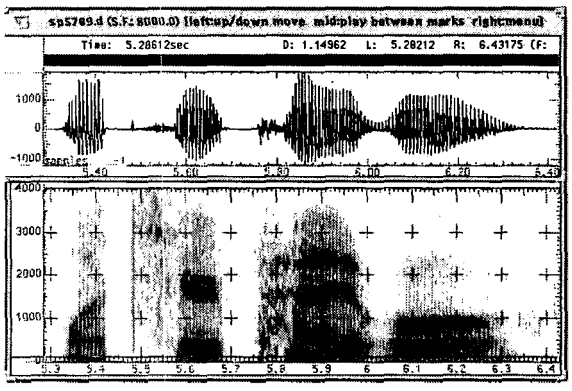

그림 3. '5789'의 파형과 스펙트로그램.

또한 '칠'과 '일'을 비교하면 '칠'의 'ㄹ'은 '71/치릴/, 72/치리/, 75/치로/' 같이 비교적 소리 이음이 자유롭고 이 때의 'ㅣ'는 /i/에 해당한다. 그러나, 일의 'ㄹ'은 그렇지 아니하다. 즉 '일'의 'ㄹ'은 'ㄹ' 겹치기 규칙을 준수하여 '11/일릴/, 12/일리/, 10/일령/' 같이 홑소리 사이에서 겹쳐서 나타난다. 그러나, 특이하게 '일'과 '오'가 연이어 나는 경우에는 '15/일로/' 보다 '15/이로/'가 더 자연스럽다. 그러나 이 때의 'ㅣ'는 /i/라기 보다는 /ii/에 해당한다.

'오'와 '구'의 음향적 특성도 그림 1에서 3을 통해서 관찰할 수 있는데, 'ㅗ'와 'ㅜ' 모두 제 1, 2 포만트가 낮고 그 차이가 크지 않음을 알 수 있다. 그림에서는 'ㅗ'는 포만트 전이가 일어나는 것으로 나타나는데 이는 뒤따라 오는 'ㅣ'음의 영향으로 해석해야 한다. '오'와 '구'가 잘 혼동되는 이유는 모음의 음향특성이 비슷하여 주된 차이가 안울림 터짐소리 'ㄱ'의 유무에 달려 있기 때문이다. 안울림 터짐소리 'ㄱ'이 그림 3과 같이 울림소리가 되면 '오'와 '구'의 분별은 더욱 어려워 진다. 한편 그림 1과 2에서 보는바와 같이 'ㄱ'이 안울림 터짐소리의 제 음가대로 발성이 되는 경우에도 '오' 앞에 약한 헤차기나 입술 부딪히는 소리가 있는 경우 '구'로 오인하기 쉽다. 이러한 문제를 완화하기 위해서 숫자 사이에 garbage 모델을 넣는 방법도 가능하지만, 본 논문에서는 연속음성인식에 기본적으로 사용되는 short pause모델을 보다 정교하게 구성하여 이 문제를 완화하였다. 본 논문에서 사용한 short pause 모델은 그림 4와 같다.

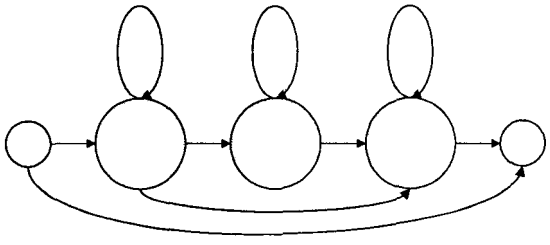

그림 4. short pause 모델

그림에서 음성벡터가 관측되는 상태는 짙은색으로 표시하였다. 더구나 이러한 접근방식은 연속으로 발성한 음성을 가대하는 인식 시스템에 일부 사용자가 각 숫자를 고립시켜 발생하는 경우에도 효과적으로 대처할 수 있는 장점이 있다.

## 3. 음운변화를 고려한 인식네트웍 구성

연결 숫자에서의 음운변화를 고려하여 정의된 변이음군에 기반하여 인식네트웍을 구성하기 위해서 각 숫자가 어떠한 변이음들의 순서열로 발성될 수 있는 가를 정의해야 한다. 즉 하나의 숫자에 대해 복수의 발음을 정의한 발음사전을 만든다. 그리고, 각 숫자의 시작 변이음과 끝 변이음만을 고려하여 phone-pair gram을 형성한다. 즉 임의의 숫자의 끝 변이음 다음에 이어 올 수 있는 임의의 숫자의 시작 변이음을 제한하도록 인식 네트웍을 구성한다. 다음과 같이 발음사전이 정의되어 있다고 하자.

일 /ii/ /l/ /sp/
일 /ii/ /r/
칠 /c/ /ii/ /l/ /sp/
칠 /c/ /i/ /r/
이 /i/ /sp/
오 /o/ /sp/
...

여기서 /sp/는 short pause모델을 의미한다. 위 사전을 이용하여 2연 숫자 인식 네트웍에서 허용되는 변이음 시퀀스는 다음과 같다.

일이 /ii/ /l/ /sp/ /i/
일오 /ii/ /l/ /sp/ /o/ , /ii/ /r/ /o/
일칠 /ii/ /l/ /sp/ /c/ /ii/ /l/
칠일 /c/ /ii/ /l/ /sp/ /ii/ /l/ , /c/ /i/ /r/ /ii/ /l/
칠이 /c/ /ii/ /l/ /sp/ /i/ , /c/ /i/ /r/ /i/
칠오 /c/ /ii/ /l/ /sp/ /o/ , /c/ /i/ /r/ /o/
...

여기서 콤마로 구분된 것은 복수로 허용되는 변이음시퀀스를 나열한 것이다. 이와 같이 인식네트웍에 실제 허용되는 변이음 시퀀스만 발생하도록 제약을 가함으로써 인식속도를 개선할 수 있음은 물론, 각 숫자의 끝에 그림 4와 같은 skip path를 가진 short pause 모델을 둠으로써 화자의 발성특성에 따른 다양한 음운변화를 인식 시스템이 흡수 할 수 있다.

## 4. 실험 및 결과

본 논문에서 사용된 음성데이터는 원광대에서 구축한 전화음성 엔진 평가용 연속음성 DB의 일부를 사용하였다[3]. 사용된 모델은 single mixture의 triphone-based HMM이며, 음성특징 파라메터로는 12차 MFCC와 에너

지 기반의 38 차 파라메터를 사용하였고, 각 모델은 3 개의 상태를 가진다. 결정트리기반의 clustering을 이용하여 전체 상태수를 약 490개로 제한 하였다. 전화망의 채널왜곡을 보상하기 위해 CMS를 적용하였다.

Baseline 인식 시스템은 표 1에서 /n/, /G/, /gg/를 제외하고, '이' 의 'ㅣ'만을 따로 취급하여 16개 변이음을 사용하였으며, 인식 네트웍은 변이음 사이의 연결에 제약을 두지 않았다. 이 시스템의 성능은 개별숫자 인식률이 93.60%, 숫자열 인식률이 78.14%였다. Baseline의 숫자 confusion matrix를 표 2에 나타내었다

표 2. Baseline의 숫자 confusion matrix

| Confusion Matrix | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | j<br>u<br>N | g<br>o<br>N | i<br>l | i | s<br>a<br>m | s<br>a | o | j<br>u<br>g<br>q | c<br>i<br>l | p<br>a<br>l | g<br>u | Del [ %c / %e] |
| juN | 822 | 0 | 1 | 1 | 9 | 3 | 1 | 38 | 0 | 0 | 1 | 2 [94.7/0.5] |
| goN | 0 | 806 | 0 | 0 | 2 | 0 | 6 | 1 | 0 | 0 | 25 | 1 [96.0/0.3] |
| il | 0 | 2 | 806 | 22 | 0 | 1 | 1 | 6 | 51 | 1 | 0 | 8 [90.6/0.8] |
| i | 2 | 14 | 102 | 672 | 3 | 0 | 2 | 4 | 21 | 0 | 0 | 15 [82.0/1.5] |
| sam | 1 | 3 | 1 | 0 | 792 | 5 | 0 | 0 | 1 | 3 | 5 | 2 [97.7/0.2] |
| sa | 0 | 0 | 0 | 0 | 5 | 809 | 2 | 0 | 0 | 14 | 0 | 3 [97.5/0.2] |
| o | 3 | 12 | 2 | 0 | 0 | 1 | 716 | 3 | 0 | 0 | 70 | 5 [88.7/0.9] |
| jugu | 20 | 2 | 24 | 9 | 0 | 0 | 2 | 1629 | 10 | 1 | 7 | 7 [95.6/0.7] |
| cil | 0 | 1 | 17 | 3 | 0 | 1 | 0 | 2 | 775 | 5 | 1 | 1 [96.3/0.3] |
| pal | 0 | 1 | 0 | 0 | 2 | 11 | 0 | 0 | 2 | 777 | 0 | 0 [98.0/0.2] |
| gu | 0 | 0 | 1 | 3 | 0 | 2 | 11 | 1 | 4 | 1 | 801 | 4 [96.3/0.3] |
| Ins | 1 | 5 | 9 | 12 | 3 | 4 | 6 | 1 | 2 | 2 | 3 | |

표에 사용된 기호는 표 1을 참조하면 되며, 위 표에서 알 수 있듯이 '이'와 '오'의 인식률은 각각 82.0%, 88.7%이며, '이'를 '일'로, '일'을 '칠'으로, '오'를 '구'로 오인하는 것이 오인식의 대부분을 차지한다. 숫자음 인식의 성능 향상을 위해 2절에 설명된 것과 같이 'ㅣ'음을 /ii/와 /i/의 두개로 분리하여 19개의 변이음을 정의 하였으며, 그림 4와 같은 short pause 모델을 적용하고, 인식네트웍을 허용 가능한 변이음 시퀀스 만으로 제약하였을 경우 개별 숫자 인식률이 94.85%, 숫자열 인식률이 82.17%였다. 전체적인 성능향상이 확연한 것은 아니다. 그러나 표 3의 숫자 confusion matrix를 보면, '이'와 '오'의 인식률이 각각 86.2%, 92.9%로 향상되었고, 또한 '일'의 인식률도 90.6%에서 93.5%로 향상되었음을 알 수 있다. 즉 한국어 숫자의 특징 분석 결과를 토대로 그 특성을 반영시켜 성능을 개선시킬 수 있는 가능성을 확인할 수 있었다. 한가지 특이한 것은 대부분의 개별 숫자의 인식률은 향상되거나 거의 비슷한 수준인데, '사'의 인식률은 오히려 97.5%에서 96.1%로 1.4%정도나 떨어졌다는 것이다. 이 점에 대해서는 추가적인 오류분석이 요구된다. 인식속도에 있어서는 Pentium III 1GHz dual CPU에 1GB의 메모리를 가진 시스템에서 인식속도의 33%의 개선이 있었다. 이는 네트웍의 제약으로 인한 탐색영역의 감소에 기인한 것으로 판단된다.

표 3. 음운변화를 고려한 경우의 confusion matrix

| Confusion Matrix | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | j<br>u<br>N | g<br>o<br>N | i<br>l | i | s<br>a<br>m | s<br>a | o | j<br>u<br>g<br>q | c<br>i<br>l | p<br>a<br>l | g<br>u | Del [ %c / %e] |
| juN | 828 | 2 | 2 | 0 | 4 | 1 | 2 | 28 | 0 | 0 | 2 | 1 [95.3/0.4] |
| goN | 1 | 810 | 1 | 0 | 2 | 0 | 3 | 0 | 0 | 0 | 23 | 1 [96.4/0.3] |
| il | 0 | 1 | 829 | 25 | 1 | 2 | 0 | 7 | 19 | 2 | 1 | 11 [93.5/0.6] |
| i | 1 | 6 | 86 | 713 | 1 | 0 | 3 | 5 | 12 | 0 | 0 | 9 [86.2/1.1] |
| sam | 0 | 4 | 1 | 0 | 792 | 4 | 0 | 1 | 0 | 4 | 2 | 3 [97.9/0.2] |
| sa | 0 | 0 | 0 | 1 | 5 | 795 | 1 | 0 | 1 | 22 | 2 | 6 [96.1/0.3] |
| o | 3 | 7 | 2 | 0 | 0 | 0 | 751 | 4 | 0 | 0 | 41 | 4 [92.9/0.6] |
| jugu | 10 | 1 | 24 | 4 | 0 | 0 | 2 | 1653 | 7 | 1 | 7 | 2 [96.7/0.6] |
| cil | 0 | 0 | 17 | 2 | 1 | 0 | 0 | 1 | 774 | 9 | 0 | 2 [96.3/0.3] |
| pal | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 784 | 0 | 4 [99.4/0.0] |
| gu | 0 | 7 | 1 | 1 | 1 | 1 | 16 | 0 | 4 | 2 | 801 | 2 [96.0/0.3] |
| Ins | 0 | 3 | 5 | 7 | 2 | 6 | 14 | 1 | 0 | 0 | 3 | |

## 5. 결 론

본 논문에서는 한국어 연결숫자 인식시스템의 성능을 향상시키기 위해 숫자 사이의 다양한 음운변화 및 숫자에서 자주 오인식이 일어나는 쌍들에 대한 특징 분석에 근거하여 변이음군을 정의하고, 인식네트웍에서 사용자의 발성 형태에 따른 다양한 음운변화를 흡수 할 수 있도록 네트웍을 구성하였다. 제안된 방법에 의해 한국어 숫자에서 자주 오인식되는 '이', '오', '일' 인식 성능이 각각 4.2%, 4.2%, 2.9%씩 향상되었으며, 숫자열에 대한 인식률은 약 4% 개선되었다. 숫자열에 대한 인식률 개선이 두드러진 것은 아니지만 약 33%의 인식의 속도의 개선이 있었고, 음운변화의 특성을 고려하여 변이음군을 정의하고 이에 따라 인식네트웍에 제약을 가함으로써 성능이 향상됨을 확인할 수 있었다.

오류분석을 통해 추가적인 숫자의 변별특징을 반영한다면 추가적인 성능 향상이 있을 것으로 기대 된다.

## 참고문헌


[1] 남기심, 고영근, *표준 국어문법론*, 탑출판사, 1996.

[2] 허 웅, *국어음운학 – 우리말 소리의 오늘 어제 –*, 샘문화사, 1999.

[3] S. G. Chon, M. G. Song and H. S. Kim, "Performance comparison of several channel compensation methods in connected digit recognition," in Proc. ICSP, pp. 897-900, 2001.