

# 한국인의 외국어 발화오류 검출을 위한 음성인식기의 발음 네트워크 구성\*

이상필(대전대), 권철홍(대전대)

## <차 례>

- |                      |                                   |
|----------------------|-----------------------------------|
| 1. 서 론               | 4.1. 청취 판단 결과                     |
| 2. 발음오류 자동분류 음성인식기   | 4.2. 음성인식기 자동분류와<br>청취판단 과의 비교 분석 |
| 3. 발음 네트워크 구성        | 4.3. 인식 결과와 청취판단과의<br>일치도 비교 분석   |
| 3.1. 유사발음 분류 발음 네트워크 |                                   |
| 3.2. 오류발음 분류 발음 네트워크 |                                   |
| 4. 실험 및 결과           | 5. 결 론                            |

## <Abstract>

### **Pronunciation Network Construction of Speech Recognizer for Mispronunciation Detection of Foreign Language**

**Sang-Pil Lee, Chul-Hong Kwon**

An automatic pronunciation correction system provides learners with correction guidelines for each mispronunciation. In this paper we propose an HMM based speech recognizer which automatically classifies pronunciation errors when Koreans speak Japanese. We also propose two pronunciation networks for automatic detection of mispronunciation. In this paper, we evaluated performances of the networks by computing the correlation between the human ratings and the machine scores obtained from the speech recognizer.

\* Keywords: Automatic detection of mispronunciation, Speech recognition, Pronunciation networks

\* 본 연구는 한국과학재단 목적기초연구(R01-2002-000-00283-0) 지원으로 수행되었음.

## 1. 서 론

음성언어를 이용한 외국어 교육은 지난 날 “읽기, 쓰기” 중심의 문자기반 교육에서 “말하기, 듣기” 라는 소리기반 교육으로의 전환을 맞이하고 있다. 본 논문은 외국어 교육에서 그간 많은 비중을 차지해오지 못했던 발음교육에 관한 연구이다. 원활한 의사소통을 위해서는 최소한 상대방이 알아들을 수 있는 수준 이상의 발음이 요구되는데 외국어를 배우는 데 있어 가장 어렵고 고치기 힘든 것 중의 하나가 발음이다. 실제로 듣기와 말하기 중심의 교육을 받은 이들도 원어인 발음의 모방 수준으로 그치고 있다. 따라서 외국어를 사용하는 사람들이 말을 듣고 바로 이해하는 것, 즉 효과적인 의사소통을 위해서는 정확한 발음을 교육받는 것이 매우 중요하다.

본 논문에서는 외국어 교육에 있어 많은 시간과 비용을 투자하여 배우는 외국어 발음에 대하여 컴퓨터를 이용한 방법론을 제안한다. 발음 교육의 핵심은 학습자의 발음 진단과 평가이다. 외국어 발음에 대한 음성학적 지식을 이용하여 외국어 발화시 나타나는 현상을 진단하여 필요한 부분에 교정정보를 제공할 수 있도록 음성인식기술에 음성학적 지식을 결합한 형태의 시스템 구현 및 성능평가에 본 연구의 목적이 있다.

일본어 발음교정 시스템은 한국어를 모국어로 하는 화자들의 일본어 발음을 교정하기 위한 것이다. 이 시스템을 구현하기 위해서는 먼저 한국어와 일본어의 음성학적인 지식이 필요하다. 이는 각 언어의 음운현상에 대한 처리가 필요할 때 요구되며 실제로 음성 인식시 인식률에 상당한 영향을 끼치는 부분이다. 그리고 공학적으로는 음성언어처리를 위한 신호처리에 관한 관련지식의 습득이 요구되며, 본 연구에서는 음성인식을 이용한 방법을 제안한다.

본 논문에서는 한국인이 일본어 발화시 음소별로 발음오류를 자동으로 분류하는 HMM 음성인식기를 제안한다. 그리고 학습자의 오류 발음을 검출하기 위한 두 가지 발음 네트워크를 제안하고 어느 것이 더 나은 성능을 보이는 가를 실험을 통해 알아본다. 본 논문의 구성은 2장에서 발음오류 자동분류 음성인식기를 기술하고, 3장에서 두 가지 발음 네트워크의 구성에 대하여 설명하고, 4장에서 실험 방법 및 결과를 논하고, 5장에서 결론 및 향후 연구 과제에 대하여 기술한다.

## 2. 발음오류 자동분류 음성인식기

본 논문에서 제안한 발음오류를 자동으로 분류하는 음성인식기는, 한국인의 일본어 발화시 발음의 교정을 목표로 하고 있으므로 그 오류에 정확히 해당하는 교정지침을 제공하기 위하여, 오류 유형별 발화 모델들을 HMM 모델의 훈련 과정을

통해 미리 구축하고, 학습자의 발화에 따른 오류 양상을 자동으로 분류하는 기능을 수행한다.

발화오류 유형의 분류는 음소별로 수행하므로, 한국어 및 일본어 음소 셋이 필요하다. 한국어 음소 셋은 자음은 변이음을 고려하여 29개를, 모음은 음성학적 차이를 보이는 모음만을 고려하여 17개를 선정하여, 총 46개의 음소로 구성하였다[1]. 일본어 음소 셋은 일본음향학회에서 선정한 음소 셋을 참조하여[2], 38개의 음소를 선정하였다[1]. 또한 일본어 음소와 유사하게 발음되는 음소들의 집합에 대한 정의가 필요하여, 본 연구에서 수집한 음성 DB를 일본어 음성학 전문가가 청취하여 한국인의 일본어 발화시 나타나는 음소별 발음오류 유형을 분석하여 유사발음 음소 셋을 구성하였다[1].

한국어 및 일본어 음소 모델은 다음과 같은 방법으로 생성하였다. 음성신호의 분석은 매 10msec 마다 25msec의 Hamming 창함수를 사용하여 MFCC 39차를 추출하였고, HMM의 구조는 3 state left-to-right continuous HMM을 사용하여 언어별 음소별 음향모델을 생성하였다. Mixture 수는 일본어 음소모델인 경우 1로, 한국어인 경우 15를 사용하였다. 이는 같은 Mixture 수를 사용한 경우에 한국어보다 일본어 음소인식기의 인식률이 높으므로, 일본어 음소로 바이어스 되어 인식되는 경향을 배제하기 위함이다[1].

음성 DB는 연구실 수준의 조용한 방에서 PC에서 수집하였고, 사운드카드는 Soundblaster Audigy를, 마이크는 SHURE 565SD를 사용하였다. 한국어 음성 DB는 SiTEC에서[3] 작성한 한국어 PBW (Phonetically Balanced Words) 452개 단어를 대전대학교 대학생 70명을 대상으로 수집하여 한국어 음소 모델을 생성하였다. 일본어 음성 DB는 ATR(일본 자동통역 연구소)에서 작성한 PBW 216개 단어를 고려대학교에서 한국어 연수 중인 일본인 70명을 대상으로 녹음하여 일본어 음소 모델을 만드는데 사용하였다.

### 3. 발음 네트워크 구성

음성인식기에 입력되는 학습자의 음성 데이터가 원어인 일본어와 한국인이 발화한 일본어 사이에서 어느 정도의 유사도를 갖는가를 분석하기 위하여 두 가지 발음 네트워크를 구성하여 음성인식기를 테스트하였다. 학습자의 오류 발음을 검출하기 위하여 본 논문에서 실험한 발음 네트워크는 다음과 같다.

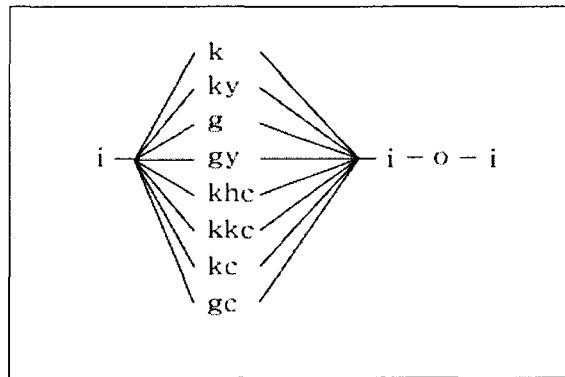
#### 3.1. 유사발음 분류 발음 네트워크

이 방식은 학습자의 발화가 어떤 언어의 어떤 음소로 나타나는가를 분석하기

위하여 일본어와 한국어 음소모델로 발음 네트워크를 정의한다. 자세히 설명하면, 2장에서 설명한 유사발음 음소 셋을 이용하여 <표 1> 및 <그림 1>과 같은 발음 네트워크를 구성하여 학습자의 오류 발음을 검출한다. <그림 1>에서, 일본어 [k]를 일본어 [k, ky, g, gy]로 발화하거나 한국어 [khc/ㅋ/, kkc/ㄱ/, kc/무성 ㄱ/, gc/유성 ㄱ/]로 발화하는 오류가 발생함을 의미한다. 일본어 음소 모델은 일본인이 발화한 일본어 음성 DB로, 한국어 음소 모델은 한국인이 발화한 한국어 음성 DB로 생성하였다. 그리고 발음오류 자동분류 음성인식기를 테스트하기 위한 음성 DB는 일본어 비전공자 30명이 일본어 단어 24개를 발성한 것을 사용하였다.

<표 1> 유사발음 분류 발음 네트워크

\$phone1 = i ;
\$phone2 = k   ky   g   gy   khc   kkc   kc   gc ;
\$phone3 = i ;
\$phone4 = o ;
\$phone5 = i ;
( SENT-START \$phone1 \$phone2 \$phone3 \$phone4 \$phone5 SENT-END )



<그림 1> 유사발음 분류 발음 네트워크  
(일본어 단어 「いきおい」 [i k i o i]인 경우)

### 3.2. 오류발음 분류 발음 네트워크

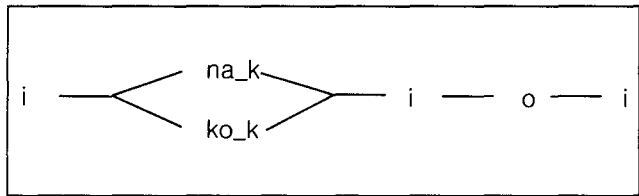
이 방식은 학습자의 발화가 어떤 형태로 나타나는가를 분석하기 위하여, 일본인 발화 일본어 음성 DB로 표준발음 모델을, 비전공자 발화 일본어 음성 DB로 오류발음 모델을 구성하여 1차 발음 네트워크를 정의하여, 1차로 한 단어 내에 오류발음이 존재하는가 여부를 우선 검출한다. 그리고 오류발음으로 인식된 발화

에 대하여 다시 그것이 어떤 양상을 보이는 가를 분석하기 위하여, 2차로 오류발음으로 분류된 음소에 대하여 특정 오류발음을 검출하는 방식이다. 자세히 설명하면, 표준발음 음소 모델(na\_\*)과 부정확하게 발화된 오류발음 음소 모델(ko\_\*)을 만들어 <표 2> 및 <그림 2>와 같은 발음 네트워크를 1차로 구성한다. <그림 1>의 발음 네트워크에서는 일본어 [k] 음에 대하여 [k|ky|g|gy|k|hc|kkc|kc|gc]와 같은 유사발음 음소 셋을 구성하였으나, <그림 2>의 방식은 한국인이 일본어 [k] 발화시 오류로 분류된 음성 DB로 오류발음 음소 모델 (ko\_k) 하나를 대표로 만드는 것이다. 그리고 오류발음으로 분류된 음소에 대하여, 2차로 <표 3> 및 <그림 3>의 발음 네트워크를 이용하여 특정 오류발음을 검출한다. 표준발음 음소 모델(na\_\*)은 일본인이 발화한 일본어 음성 DB로, 오류발음 음소 모델(ko\_\*)은 일본어 비전공자의 일본어 음성 DB로, 한국어 음소 모델은 한국인이 발화한 한국어 음성 DB로 생성하였다. 그리고 발음오류 자동분류 음성인식기를 테스트하기 위한 음성 DB는 일본어 비전공자 30명이 일본어 단어 24개를 발성한 것을 사용하였다.

<표 2> 오류발음 분류 1차 발음 네트워크

```

$phone1 = i ;
$phone2 = na_k | ko_k ;
$phone3 = i ;
$phone4 = o ;
$phone5 = i ;
( SENT-START $phone1 $phone2 $phone3 $phone4 $phone5 SENT-END )
    
```



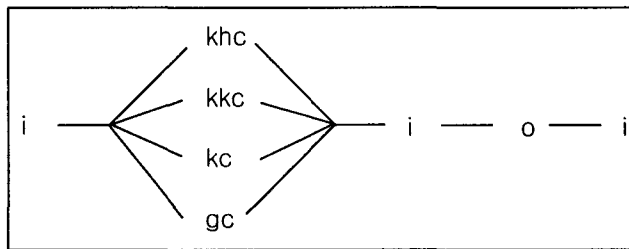
<그림 2> 오류발음 분류 1차 발음 네트워크  
(일본어 단어 「いきおい」 [i k i o i]인 경우)

&lt;표 3&gt; 오류발음 분류 2차 발음 네트워크

```

$phone1 = i ;
$phone2 = khc | kkc | kc | gc ;
$phone3 = i ;
$phone4 = o ;
$phone5 = i ;
( SENT-START $phone1 $phone2 $phone3 $phone4 $phone5 SENT-END )

```



<그림 3> 오류발음 분류 2차 발음 네트워크  
(일본어 단어 「いきおい」 [i k i o i]인 경우)

## 4. 실험 및 결과

### 4.1. 청취 판단 결과

이 절에서는 인식결과의 비교 기준이 되는 음성학 전문가의 청취판단에 대한 실험결과를 요약한다[4]. 실험 대상 자료는 일본어 [k] 음소를 포함하고 있는 24개 단어를 20대 일본어 비전공 여성 15명, 남성 15명 등 30명의 피험자가 발음한 720개의 음성 파일로 하였다. 실험결과 어두 음소와 어중 음소의 차이가 각 음소별로 크게 차이를 보이므로 실험결과를 어두와 어중으로 나누어 분석하였다. <표 4>에 나타난 것처럼 어두에서 일본어 [k] 음소는 주로 한국어의 /ㄱ/를 발음할 때와 같은 청각인상으로 나타났다. 이것은 일반적으로 통용되고 있는 내용인 “일본어 어두의 [k] 음소는 한국어의 /ㄱ/로 실현된다”라는 것과는 다른 결과이다. /ㄱ/가 가장 적고 /ㅋ/인 경우가 더 많았다.

<표 4> 어두 음소 분석 결과

	ㄱ	ㅋ	ㄱʰ
빈도수(개수)	139	61	10

일본어 [k] 음소가 어중에 올 때는 <표 5>에서 보듯이 한국어의 /ㄱʰ/를 발음할 때와 같은 청각인상으로 주로 분석되었다. 이것은 일반적으로 통용되고 있는 내용과 비슷한 결과이다[5].

<표 5> 어중 음소 분석 결과

	ㄱ	ㅋ	ㄱʰ
빈도수(개수)	58	129	323

<표 6>에서는 음성 환경에 관계없이 일본어 음소 [k]를 한국인이 발음했을 때 분석되는 청각 인상을 살펴 본 것으로 /ㄱʰ/가 가장 많았고 다른 두 경우는 비슷한 것으로 분석되었다. 결론적으로, 한국인이 일본어 [k] 음소를 발음할 때 일본어를 배우지 않은 피험자들은 일반적으로 한국어의 /ㄱʰ/로 발음하는 경우가 가장 많은 것으로 나타났음을 알 수 있다.

<표 6> 음성자질에 의한 분석 결과

	ㄱ	ㅋ	ㄱʰ
빈도수(개수)	197	190	333

## 4.2. 음성인식기 자동분류와 청취판단과의 비교 분석

### 4.2.1. 유사발음 분류 발음 네트워크 실험 결과

<표 7>에 유사발음 분류 발음 네트워크를 사용하여 비전공자의 일본어 [k] 발화에 대한 인식결과를 음성 환경에 따라 정리하였다. 일본어 [k]로 발화된 경우를

제외하고 어두환경에서는 /ㄱ/, /ㅋ/, /ㄲ/ 순으로, 어중에서는 /ㄲ/, /ㅋ/, /ㄱ/ 순으로 발화함을 알 수 있다.

<표 7> 유사발음 분류 발음 네트워크를 사용한 [k] 인식 결과(개수)

음성 환경	k	ㄱ	ㅋ	ㄲ
어두	35	98	60	17
어중	33	24	136	317
합	68	122	196	334

#### 4.2.2. 오류발음 분류 발음 네트워크 실험 결과

오류발음 분류 1차 발음 네트워크에서 표준발음(na\_k)과 오류발음(ko\_k)으로 분류하고, 오류발음으로 검출된 데이터에 대하여 오류발음 분류 2차 발음 네트워크를 사용하여 음성 환경에 따른 일본어 [k] 음소의 인식결과를 구했다. <표 8>에 비전공자의 일본어 [k] 발화에 대한 음성 환경에 따른 인식결과를 정리하였다. 어두환경에서는 /ㄱ/, /ㅋ/, /ㄲ/ 순으로, 어중에서는 /ㄲ/, /ㅋ/, /ㄱ/ 순으로 발화함을 알 수 있다.

<표 8> 오류발음 분류 발음 네트워크를 사용한 [k] 인식 결과(개수)

음성 환경	k(na_k)	ㄱ	ㅋ	ㄲ
어두	5	111	69	25
어중	21	25	146	318
합	26	136	215	343

#### 4.2.3. 인식결과와 청취판단과의 비교 분석

이 절에서는 유사발음 분류 및 오류발음 분류 등 두 가지 발음 네트워크를 사용한 음성인식기로 실험한 결과와 음성학 전문가의 청취판단 결과를 비교 분석한



다. <표 9>에 비전공자의 일본어 [k] 발화에 대한 음성인식기의 자동분류 결과와 전문가의 청취판단 결과를 음성 환경에 따라 정리하였다. 음성학 전문가의 청취판단은 비전공자는 일본어 [k]로 발화한 경우는 없다고 보았다.

<표 9> 음성인식기 자동분류와 청취판단 결과 비교(%)

	음성 환경	k	ㄱ	ㅋ	ㄲ
청취 판단	어두	0	66.2	29.0	4.8
	어중	0	11.4	25.3	63.3
유사발음 분류 네트워크	어두	16.6	46.7	28.6	8.1
	어중	6.5	4.7	26.7	62.1
오류발음 분류 네트워크	어두	2.4	52.9	32.8	11.9
	어중	4.1	4.9	28.6	62.4

세 가지 실험결과를 비교해 보면, 일본어 [k] 음이 음성 환경에 따라 어두에서는 한국어 /ㄱ/, /ㅋ/, /ㄲ/ 순으로, 어중에서는 /ㄲ/, /ㅋ/, /ㄱ/ 순으로 발음된다는 사실에, 비율의 차이는 존재하지만 경향이 일치하는 결과를 보여 주었다. 이와 같은 실험결과로부터 음성인식기에 의한 자동 분류가 제대로 수행되었음을 알 수 있다. 그러나 <표 9>의 각 음소별 출현비율 비교로부터 유사발음 네트워크와 오류발음 네트워크 중에서 어느 것이 청취 판단에 보다 가까운 결과를 주는가에 대한 결론을 내릴 수 없다. 따라서 두 네트워크간의 정확한 성능평가를 위하여 다음절에서 청취 판단과의 일치도를 비교 분석한다.

#### 4.3. 인식결과와 청취판단과의 일치도 비교 분석

이 절에서는 유사발음 분류 및 오류발음 분류 등 두 가지 발음 네트워크를 사용한 음성인식기로 실험한 결과와 전문가의 청취판단 결과와 일대일 비교를 통한 일치도를 비교 분석한다.

<표 10>에 음성인식기의 자동분류와 전문가의 청취 판단과의 일치도를 정리해 두었다. 유사발음 네트워크에서의 실험결과는, 음성 환경에 따라 /ㄱ/ 음이 어두에서 74개, 어중에서 8개, /ㅋ/ 음이 어두에서 27개, 어중에서 86개, /ㄲ/ 음이 어두에서 2개, 어중에서는 265개 등이 청취판단과 일치하였다. 오류발음 네트워크에서의

실험결과는, 음성 환경에 따라 /ㄱ/ 음이 어두에서 83개, 어중에서 7개, /ㅋ/ 음이 어두에서 31개, 어중에서 94개, /ㄱ/ 음이 어두에서 3개, 어중에서 265개 등이 청취판단과 일치하였다.

<표 10> 음성인식기 자동분류와 청취판단과의 일치도 비교(개수)

	음성 환경	ㄱ	ㅋ	ㄱ
청취 판단	어두	139	61	10
	어중	58	129	323
유사발음 분류 네트워크	어두	74	27	2
	어중	8	86	265
오류발음 분류 네트워크	어두	83	31	3
	어중	7	94	265

<표 11>에 <표 10>의 일치도를 백분율로 표시하였다. <표 10>, <표 11>을 살펴 보면, 어두인 경우에 /ㄱ/, /ㅋ/, 어중인 경우에 /ㅋ/에서 오류발음 네트워크가 우수한 성능을 보여 주고, 나머지 경우에는 두 발음 네트워크가 비슷한 성능을 보여 준다. 청취 판단과의 비교대상 전체 발화 데이터 720개 중에서, 유사발음 네트워크 방식은 64.2%(462개)의 일치도를 보여 주었고, 오류발음 네트워크 방식은 67.1%(483개)의 일치도를 보여 주었다. 이와 같은 결과로부터 본 논문에서 제안한 발음 네트워크 구성 방법 중에서, 한국어와 일본어 유사음소를 일차원적으로 비교하는 유사발음 네트워크 보다, 원어민과 비전공자 일본어 모델 간에 비교를 통하여 도출한 결과를 다시 한국어 모델로 비교하는 2차원적 오류발음 네트워크 방식이 더 나은 성능을 보이고 있음을 알 수 있다.

<표 11> 음성인식기 자동분류와 청취판단과의 일치도 비교(%)

	음성 환경	ㄱ	ㅋ	ㄷ
유사발음 분류 네트워크	어두	53.2	44.3	20.0
	어중	13.8	66.7	82.0
오류발음 분류 네트워크	어두	59.7	50.8	30.0
	어중	12.1	72.9	82.0

## 5. 결 론

본 논문에서는 외국어 학습자의 발음 품질을 평가하여 표준 및 오류 발음을 자동으로 분류하는 음성인식기를 제안하였다. 학습자의 오류 발음을 검출하기 위하여 유사발음 및 오류발음 분류 발음 네트워크를 사용하여 어느 것이 청취 판단에 보다 가까운 결과를 주는가에 대한 실험을 수행하였다. 발음오류 자동분류 음성인식기의 실험 결과와 음성학 전문가의 청취 판단과의 비교로부터 음성인식기에 의한 자동 분류가 제대로 수행되었음을 알 수 있었다. 또한 유사발음 네트워크보다 오류발음 네트워크 방식이 더 나은 성능을 보이고 있음을 알 수 있었다.

향후 연구에서는 자동 분류의 정확도를 높여 자동 분류 결과와 음성학 전문가의 청취 판단을 좀 더 일치시키는 연구를 수행할 계획이다. 이를 위하여 학습자의 발음에 대한 스코어링(scoring) 방법에 관한 연구를 진행할 계획이다.

## 참 고 문 헌

- [1] 권철홍, 강효원, 이상필, “음성인식기를 이용한 한국인의 외국어 발화오류 자동검출”, *말소리*, 48호, pp.15-23, 2003.
- [2] T. Kawahara, T. Kobayashi et al., “Sharable software repository for Japanese large vocabulary continuous speech recognition”, *Proc. ICSLP 98*, pp.3257-3260, Sydney, 1998.
- [3] SiTEC, 음성정보기술산업지원센터, <http://www.sitec.or.kr>
- [4] 이재강, 권철홍, “청각인상과 음성 파형간의 관계 규명을 위한 일본어 /k/의 기초연구”, *대한음성학회 봄 학술대회 논문집*, pp.52-55, 2003.
- [5] 민광준, *일본어 음성학 입문*, 건국 대학교 출판부, 2002.

접수일자: 2004년 2월 5일

게재일자: 2004년 3월 15일

▶ 이상필(Sang-Pil Lee)

주소: 300-716 대전광역시 동구 용운동 96-3 대전대학교

소속: 대전대학교 정보통신공학과 BMW 연구실

전화: 042) 280-2567

E-mail: nishot@hanmail.net

▶ 권철홍(Chul-Hong Kwon) : 책임저자

주소: 300-716 대전광역시 동구 용운동 96-3 대전대학교

소속: 대전대학교 정보통신공학과 BMW 연구실

전화: 042) 280-2555

E-mail: chkwon@dju.ac.kr