Journal of the Korea Industrial Information Systems Research Vol.25 No.3, Jun. 2020 : 99-110
http://dx.doi.org./10.9723/jksiis.2020.25.3.099 ISSN:1229-3741

한글 발음을 이용한 중국어 입력 시스템+

(Chinese Input System using Hangeul Pronunciation)

이 중 화[1)*], 윤 홍 원[2)]
(JungHwa Lee and HongWon Yun)

요 약 중국어를 컴퓨터나 스마트폰에 입력할 때 로마자를 이용한 한어 병음 시스템이 많이 사용되고 있다. 그러나 기존의 한어 병음 시스템은 중국어 발음을 그대로 표기할 수 없기 때문에 우리나라 사람들이 사용하기에 어려운 점이 많다. 본 논문에서는 전 세계 문자들 중에서 발음을 가장 잘 표기 할 수 있는 우리 한글을 이용하여 중국어를 입력할 수 있는 중국어 입력 시스템을 설계한다. 본 연구에서 제안하는 시스템에서는 중국어 발음을 유사하게 표기하는 경우에도 중국어 입력이 가능하게 함으로써 입력에 성공하는 확률을 높였으며, 성능평가를 통해 이를 검증하였다.

핵심주제어: 중국어입력시스템, 한어병음, 중국어, 한글

Abstract The Chinese Pinyin system using Roman characters is used to input Chinese into a computer or a smartphone. However, because the Chinese Pinyin system can not express Chinese pronunciation accurately as it is, it is very difficult for Korean people to input Chinese using this system In this paper, we designed a Chinese input system using Hangeul which can best represent pronunciations among the world's characters. In the system proposed in this study, Chinese input is possible even when Chinese pronunciation is similarly expressed, which increases the probability of successful input and also verified it through performance evaluation.

Keywords: Chinese input system, Chinese Pinyin system, Chinese, Hangeul

1. 서 론

최근 중국과의 무역 확대와 유학, 관광 등의 교류로 인해 중국어에 대한 관심이 날로 증가하고 있다. 중국어를 컴퓨터나 스마트폰에 입력하는 방법에 대해서는 이미 많은 연구들이 진행되어 여러 가지 방법들이 제시되었으나, 현재에는 한어 병음을 이용한 입력방법이 가장 많이 사용된다 (Chen and Lee, 2000; Shen et al., 2007; Wong et al., 2011). 한어 병음이란 중국어 한자의 발음을 로마자로 표시해 놓은 것으로써, 입력하고자 하는 한자에 해당하는 한어 병음을 로마자로 입력하면 입력 소프트웨어가 해당 발음을 가지는 한자어를 리스트 형태로 보여주고 사용자가 원하는 글자를 선택하는 방식이다. 그러나 이 방법은 로마자의 한계 때문에 정확한 발음을 예측하기 어려울 뿐만 아니라, 중국어 발

* Corresponding Author: junghwa@deu.ac.kr
+ 이 논문은 2019학년도 동의대학교 연구년 지원에 의하여 연구되었음.
Manuscript received March 03, 2020 / revised March 16 ,2020 / accepted March 20, 2020
1) 동의대학교 컴퓨터소프트웨어공학과, 제1저자, 교신저자
2) 신라대학교 컴퓨터소프트웨어공학부, 제2저자

음을 듣고 그대로 적기 매우 어렵다 (Tan et al., 2013; Gao et al., 2015; Xiao et al., 2016). 특히 우리나라 사람들을 포함한 외국인에게 한어 병음은 한자의 발음 표기라기보다는 또 다른 한자의 로마자 표기 방법으로 취급될 수 있기 때문에 외국인의 입장에서는 한자와 더불어 병음을 따로 외워야하는 이중적인 어려움을 겪게 된다 (Pasquarella et al., 2015; Gottardo et al., 2017). 실제로 중국인의 경우에도 로마자 병음을 정확하게 입력하지 못해서 여러 번 고쳐 입력을 해야 하는 경우도 많은 실정이다.

따라서 이러한 문제점을 해결하기 위해 본 연구에서는 세계의 여러 문자들 중에서 발음을 가장 잘 표기할 수 있는 한글을 이용하여 중국어 발음을 표기하고 이를 중국어 입력기에 사용할 수 있는 방법에 대해 연구한다. 여기에서는 중국어 발음을 한글로 표기할 때 정해진 한글 발음 표기뿐만 아니라 유사한 발음 표기까지 고려하여 사용자가 정해진 발음 표기를 하지 못한 경우에도 가장 근접한 한자를 찾아 줌으로써 한자의 입력 효율을 높이고자 한다.

본 연구의 구성은 다음과 같다. 제 2 장에서는 중국어 입력 방법에 대한 연구로 한어 병음 방법과 그 이외의 몇 가지 입력 방법에 대해 살펴 보고, 제 3장은 본 연구에서 제안하는 중국어 입력 방법에 대해 이론적 배경과 설계 과정을 살펴보고 제안하는 방법의 입력 성능에 대해 평가한다. 마지막으로 제 4 장은 결론과 향후 연구과제에 대해 살펴본다.

2. 관련연구

2.1 한어 병음

중국어는 글자를 보고 그 발음을 유추할 수 없고 각 글자에 대한 발음을 따로 외워야 한다. 이런 이유로 중국에서는 중국어의 발음을 어떻게 표기할 것인가에 대한 많은 연구가 이루어졌으며, 이를 통해 중국어 발음을 로마자로 표기하는 방식을 연구하여 1958년 중국어 로마자 표기법으로 한어 병음(漢語拼音)이 제정되었다.

한어 병음에서는 한글 음절 초성에 해당하는 b, p, d, t, g, k, m, n, f, s, sh, r, x, h, z, c, zh, ch, j, q, l 등의 21개의 성모와 한글 음절 중성 및 종성에 해당하는 a, o, e, i u, ü 의 6개 단운모와 이들을 조합한 30개의 결합운모 등 총 36개의 운모의 조합으로 음절을 나타낸다. 중국어 음절은 성모와 운모를 조합하여 나타내는데, 예를 들어 한자 '北' 은 /bei/ 와 같이 음절을 표기한다.

한어 병음 시스템은 중국어의 발음을 표기하기 위한 방법으로 고안되었지만 현재에는 컴퓨터에 글자를 입력하는 방법으로도 사용된다. 한어 병음을 이용한 글자 입력 방법은 앞의 서론에서 잠깐 살펴보았는데, 예를 들어 '中國'을 입력하고자 하면 '中國'의 발음에 해당하는 병음 /zhong guo/를 먼저 입력하고 입력 소프트웨어에서 사전을 검색하여 해당 발음을 가지는 단어들을 리스트로 제시하면 사용자가 원하는 단어를 선택하여 입력이 이루어진다. Fig. 1은 스마트폰을 이용하여 '中國' 을 입력하는 예이다.

2.2 중국어 한글 표기에 관한 연구

Kim and Choi (1997)와 Eom (2002)에서는 중국어를 한글로 표기하는 방법에 대해 연구하였다. 이 연구들에서는 중국어 성모와 운모에 대해 가장 근접한 발음을 가지는 한글 자음과 모음을 배치하고 이를 조합하여 중국어 음절을 표기한다. 예를 들어 성모 /b/는 자음 /ㅃ/로, /g/의 경우 /ㄲ/ 로 표기하고 운모 /a/는 /ㅏ/로 표기하도록 하여 중국어 음절 /ba/는 한글 음절

Fig. 1 Example of Input '中國' using Google Pinyin Input System

/빠/로 표기한다.

또한 대한민국 정부 부처인 문화체육관광부에서는 외래어를 한글로 표기할 때 어떻게 표기할지에 대한 방법을 제시하고 있다. 문화체육관광부 외래어 표기법에는 여러 나라의 언어에 대한 내용을 포함하고 있는데, 여기에 중국어도 포함된다. 이 방법은 위에서 언급한 두 방법과 원칙적으로는 같은 방법을 사용한다. 그러나 우리나라 외래어 표기 규정에서는 /ㅃ/ 과 같은 경음을 허용하지 않기 때문에 성모 /b/의 경우 자음 /ㅂ/으로 성모 /g/의 경우 /ㄱ/으로 표기한다.

Jeong (2007a)와 Jeong (2007b)의 연구도 중국어의 한글 표기에 대한 연구인데, 이 연구의 특징은 중국어 발음을 최대한 비슷하게 한글로 표기하면서 중국어 한 음절을 반드시 한글 한 음절로 표기하도록 하고 있다. 이 연구에서는 새로운 한글 표기법인 온누리 한글 표기법을 제안하였는데 온누리 한글 표기법에서는 새로운 한글 자음과 모음을 도입하였다. 예를 들어 자음에는 새로운 겹자음 /ㄻ/, /ㅶ/, /ㅱ/ 등을 도입하고 ,모음으로는 / ㅗㅏ/ / ㅜㅓ/ 등을 도입하여 음절을 표기하는데 사용한다.

Ahn (2007)은 한글을 이용한 중국어 입력기에 대한 연구이다. 이 연구에서는 중국어 음절을 동시에 입력하는 정음 입력기를 제안하고 있다. 여기에서는 중국어를 분석하여 한글로 표기할 때 사용하지 않는 한글 자음과 모음을 글자판에서 제거하고, 자주 사용 빈도가 높은 자음과 모음을 글자판 주요 위치에 배치하는 등 수정된 중국어 입력용 글자판을 제안하고 있다 (Fig. 2 참조). 또한 이 연구에서는 글자판뿐만 아니라 419개의 중국어 음절에 대한 한글 표기법을 제안하고 있다.

3. 한글 발음을 이용한 한어 병음 표기

우리 한글은 개발 단계부터 사람이 발음하는 소리를 잘 표기할 수 있도록 만든 언어이기 때문에 현재 사용되고 있는 어떤 언어표기법 보다 소리 표기를 잘 할 수 있다. 따라서 중국어 발음을 로마자로 표기한 기존의 한어 병음 시스템

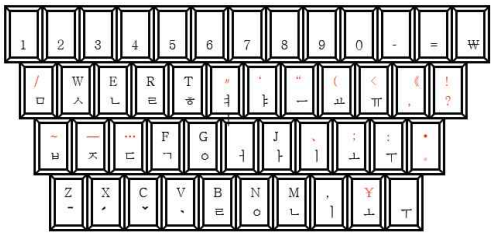

Fig. 2 Jungeum Keyboard Layout

보다 실제 발음에 근접하게 표기할 수 있으며 이를 중국어 입력 시스템에 이용하면 우리나라 사람들이 기존의 한어 병음 시스템으로 중국어를 입력할 때 보다 빠르고 편하게 중국어를 입력할 수 있다. 이러한 장점을 이용하여 본 연구에서는 우리나라 사람들이 중국어를 보다 편리하게 입력할 수 있도록 중국어 한글 발음을 이용한 중국어 입력 시스템을 개발한다.

본 절에서는 먼저, 한어 병음에서 사용되는 음소들에 대해 분석하고 각 음소들을 우리나라 사람들이 어떻게 인식하고 표기 할 수 있는지에 대해 살펴본다. 그리고 이를 이용하여 각 음소에 대한 한글 대표 발음과 유사 발음에 대해 정의하고 실험을 통해 본 연구의 결과에 대해 살펴본다.

3.1 한어 병음에 사용되는 음소와 음절 구성 방법

중국어에서 사용되고 있는 한자는 간체자를 기준으로 약 2,236개가 있는데, 각 한자를 음절 기준으로 다시 분류하면 약 400여 개의 음절로 분류될 수 있으며, 각 음절은 앞에서 언급한 바와 같이 성모와 운모의 조합으로 표기한다.

한어 병음에서 성모는 /b/, /p/, /m/, /f/, /d/, /t/, /n/, /l/, /g/, /k/, /h/, /j/, /q/, /x/, /zh/, /ch/, /sh/, /r/, /z/, /c/, /s/ 등의 총 21개이다. 성모는 한글의 초성에 해당하는데, 예를 들어 /b/는 한글 초성의 /ㅂ/, /p/는 한글 초성 /ㅍ/과 비슷한 음가를 가진다. 각 성모에 대한 자세한 음가는 뒤에서 다시 자세히 살펴보겠다.

운모는 한어 병음에서 소리마디를 이룰 때 모음의 역할을 하는 음소로써 /a/, /o/, /e/, /i/, /u/, /ü/ 등의 6개의 단운모와 이들의 결합으로 만들

어진 30개의 결합운모 등 총 36개로 구성된다. Table 1은 6개의 단운모와 단운모들의 결합으로 만들어 지는 결합운모를 표로 나타낸 것이다.

Table 1 Finals of Pinyin

Simple Finals	Combination Finals			
a	ao	ai	an	ang
o	ou	ong		
e	ei	en	eng	er
i	ia (ya)	ian (yan)	iang (yang)	iao (yao)
	in (yin)	ing (ying)		
	iou (you)(iu)	iong (yong)	ie (ye)	
u	ua (wa)	uan (wan)	uang (wang)	uai (wai)
	uen (wen)	ueng (weng)	uei (wei)	uo (wo)
ü	üe (yue)	üan (yuan)	ün (yun)	

* () : case of using alone without initial

운모 /a/는 한글 중성 /ㅏ/ 와 비슷하며 운모 /o/는 한글 중성 /ㅗ/ 와 비슷한 소리를 낸다. 운모의 자세한 음가 역시 뒤에서 자세히 살펴보겠다.

이제 위에서 언급한 성모와 운모를 조합하면 중국어 음절을 표기할 수 있는데, 중국어 '北京'과 '青島' 를 한어 병음으로 표기하면 다음과 같다.

北京[Bei jing]
上海 [Shang hai]

한어 병음 시스템은 중국어 발음에 사용되는 성모와 운모에 대해 각 발음을 로마자를 이용해 표기하고 있다. 그런데 로마자 표기 방법은 발음에 모호성이 존재한다. 이는 로마자를 사용하는 대표 언어인 영어에서도 쉽게 살펴볼 수 있는데, 예를 들어 영어 모음에 해당하는 u의 경우 어떤 경우에는 /ʌ/ 로 발음하고 어떤 경우에는 /u/ 로 발음한다.

just /dʒʌst/
July /dʒu'laɪ/

위의 예에서 보듯이 로마자의 경우 어떤 글자들은 여러 가지로 발음할 수 있기 때문에 각 단어에 대한 발음의 따로 외워야 하는 경우가 많다. 이러한 문제는 로마자를 차용한 한어 병음의 경우에도 그대로 나타나는데, 예를 들어 上浮 [shang fu] 를 /생퍼/ 라고 발음할지 /상푸/ 라고 발음할지 모르게 된다. 로마자를 이용한 한어 병음 역시 로마자들을 그대로 읽으면 해당 글자를 발음할 수 있는 것이 아니라 각 병음 표기에 해당하는 발음을 따로 외워야 하는 것이다. 따라서 이 경우 글자에 해당하는 문자를 배움과 동시에 발음 표기에 해당하는 글자를 따로 배우는 이중 학습이 필요하다. 특히 우리나라 사람이 중국어를 학습할 경우 한자의 모양과 함께 해당 한자의 발음, 그리고 발음을 표기하는 병음을 따로 학습해야 한다는 뜻이다. 특히 한어 병음을 중국어 표준 입력 방법으로 사용하고 있기 때문에 병음을 모를 경우 중국어를 컴퓨터에 입력할 수 없는 상황이 발생한다. 그러나 한자의 병음을 로마자가 아닌 한글을 사용하여 표기하는 경우에는 발음을 잘 표기할 수 있는 소리글자인 한글이 가지는 특성상 발음에 대한 모호성이 발생하지 않으며, 중국어를 학습하거나 병음을 사용하여 중국어를 입력할 때 보다 쉽게 입력할 수 있다.

3.2 한어 병음에 대한 한글 표기

본 연구에서는 로마자로 표기되어 있는 한어 병음을 우리나라 사람들이 중국어를 들리는 대로 한글 자음과 모음을 사용하여 표기하면 중국어로 변환할 수 있도록 하는 한글 발음을 이용한 중국어 입력 시스템을 개발한다.

일반적으로는 중국어에 대한 한글 음절은 해당 글자의 성모의 한글 발음과 운모의 한글 발음을 조합하면 한글 음절로 만들 수 있다. 따라서 각 성모와 운모에 해당하는 한글 자음, 모음을 미리 정의하고 이를 사용하여 중국어 음절을 표기하면

그 발음에 해당하는 중국어를 찾을 수 있게 된다 (Kim and Choi, 1997; Eom, 2002; Jeong, 2007a; 2007b). 그러나 실제로 중국어 발음을 듣고 이를 한글로 표기해 보면 로마자로 표기된 운모와 성모를 단순히 한글 자음과 모음으로 변환하고 이를 조합하여 음절로 표기하는 경우와 다른 경우가 발생하게 된다. 예를 들어 北京/Bei jing/의 경우 한글로 표기할 경우 '베이징'으로 표기하는 경우가 대부분이지만 듣는 사람에 따라서는 '뻬이징'으로 표기할 수도 있으며 또 모음을 바꾸어 '배이징'으로 표기할 수도 있다는 것이다.

beiging-베이징,뻬이징,빼이찡,배이징,빼이찡...

본 연구의 목적은 사용자가 각 성모와 운모에 해당하는 한글 자음과 모음을 외워서 쓰는 것이 아니라 들리는 대로 적어서 이를 중국어로 바꾸는 데에 있기 때문에 각 성모, 운모의 기준이 되는 한글 자음, 모음뿐만 아니라 대표 자음, 대표 모음과의 유사한 발음을 고려하여야 한다. 즉, [베이징]이라고 입력하는 경우뿐만 아니라 [뻬이징], [배이징]의 경우도 北京으로 변환될 수 있어야 한다는 것이다. 이를 위해 본 연구에서는 한글 자음과 모음의 유사 발음을 고려하여 한어 병음 성모, 운모에 해당하는 한글 자음, 모음을 정의하고 이를 이용하여 한글로 표기된 중국어를 한자로 변환한다.

성모와 운모에 대한 한글 표기는 몇 가지 연구들을 통해 발표된 적이 있으며, 중국어 학습을 위한 한글 교재 등에 이미 한글 표기를 제공하고 있다 (Kim and Choi, 1997; Eom, 2002; Jeong, 2007a; 2007b).

먼저 성모의 한글 표기에 대해 살펴본다. 성모의 한글 표기법에 관한 내용은 정부에서 제정한 한글 외래어 표기법이나 여러 가지 학술 논문에서 살펴볼 수 있는데, 대부분 성모의 발음과 가장 유사한 한글 자음을 배치하는 형태이다. Table 2는 성모에 대한 한글 표기법을 나타낸다.

Table 2의 방법 1은 문화체육관광부 외래어 표기법을 사용하는 경우이며, 방법 2는 Kim and Choi (1997), 방법 3은 Eom (2002)이 각각 제안한 방법을 사용하는 경우이다.

Table 2의 한글 성모 표기법은 각 성모에 대해 한글 자음 중에 가장 근접한 자음을 대응시킨 것이다. 성모 /b/의 경우는 /ㅂ/ 또는 /ㅃ/, /p/는 /ㅍ/, /m/은 /ㅁ/ 으로 표기하고 있는데, 각 성모당 하나의 자음을 대응하는 것을 원칙으로 하고 있으며, Table 2의 표기법은 하나의 방법론으로 실제로 그렇게 발음된다는 뜻은 아니다.

한어 병음에서의 성모와 한글의 자음은 어떤 것들을 상당히 유사하여 하나의 자음으로 대응할 수 있지만 어떤 것들을 한글 자음으로는 정확히 표현할 수 없기 때문에 듣는 사람에 따라 이를 다르게 표기할 수 있다. 예를 들어 성모 /b/ 에 해당하는 한글 자모는 /ㅂ/ 인데, 듣는 사람에 따라서는 /ㅃ/ 또는 심지어 /ㅍ/ 으로 인식할 수 있다는 것이다. 이는 기본적으로 발음이 서로 차이가 나는 경우도 있지만 같은 발음이라 할지라도 우리나라 사람들이 이미 알고 있거나 인지하지는 못하지만 학습을 통해 내제되어 있는 발음

Table 2 Hangeul Notation of Chines Initial

	initial	method1	method2	method3
1	b	ㅂ	ㅃ	ㅃ
2	p	ㅍ	ㅍ	ㅍ
3	m	ㅁ	ㅁ	ㅁ
4	f	ㅍ	ㅎ	ㅍ
5	d	ㄷ	ㄸ	ㄸ
6	t	ㅌ	ㅌ	ㅌ
7	n	ㄴ	ㄴ	ㄴ
8	l	ㄹ	ㄹ	ㄹ
9	g	ㄱ	ㄲ	ㄲ
10	k	ㅋ	ㅋ	ㅋ
11	h	ㅎ	ㅎ	ㅎ
12	j	ㅈ	ㅈ	ㅈ
13	q	ㅊ	ㅊ	ㅊ
14	x	ㅅ	ㅅ	ㅅ
15	zh	ㅈ(즈)	ㅈ	ㅈ
16	ch	ㅊ(츠)	ㅊ	ㅊ
17	sh	ㅅ(스)	ㅅ	ㅅ
18	r	ㄹ(르)	르	ㄹ
19	z	ㅉ(쯔)	ㅉ	ㅉ
20	c	ㅊ(츠)	ㅊ	ㅊ
21	s	ㅆ(쓰)	ㅆ	ㅆ

* () : case of using alone

에 대한 인식이 서로 다르기 때문이다. 이러한 오 인식 문제는 한글의 음운학적 변동 때문에 발생하는데, 한글 자음을 음운학적으로 살펴보면 Table 3과 같이 14개의 기본 자음과 각 기본 자음에 대해 여러 가지 음운 변동에 따라 나타나는 파생 자음들로 나타낼 수 있다.

Table 3 Hangeul Basic and Derived Consonants

consonant	IPA	aspiration	tensification	nasalization	lateralization	palatazation
ㄱ	/k/	/ㅋ/	/ㄲ/	/ㅇ/		
ㄴ	/n/				/ㄹ/	
ㄷ	/t/	/ㅌ/	/ㄸ/	/ㄴ/		/ㅈ/
ㄹ	/ɾ/					
ㅁ	/m/					
ㅂ	/p/	/ㅍ/	/ㅃ/	/ㅁ/		
ㅅ	/sʰ/		/ㅆ/			
ㅇ	무음					
ㅈ	/ʧ/	/ㅊ/	/ㅉ/			
ㅊ	/ʧʰ/					
ㅋ	/kʰ/					
ㅌ	/tʰ/					/ㅊ/
ㅍ	/pʰ/					
ㅎ	/h/					

예를 들어 기본 자음 /ㄱ/은 격음화로 인해 /ㅋ/ 으로 음운이 변화되며 경음화에 의해 /ㄲ/ 으로 변화된다. 또한 비음화에 의해 /ㅇ/ 으로 변화되기도 한다. 이를 역으로 논해보면 화자가 /ㄱ/ 을 발음하였으나 격음이 섞이는 경우나, 실제로 /ㄱ/ 을 제대로 발음하였으나 청취자가 격음으로 들을 경우 /ㅋ/ 으로 인식할 수 있다는 것이다. 예를 들어 '主公'에서 '公'의 경우 병음으로 /gong/ 으로 표기하는데, 발음을 들어보면 한글 /ㄱ/ 발음과 유사하다. 반면 '中國'에서 '國'의 경우 병음을 /guo/ 라고 쓰지만 이때에는 한글 /ㄲ/ 발음과 유사하다. '中國' 발음 전체를 놓고 보면 /중궈/, /쭝궈/, /쭝꿔/ 등으로 쓸 수 있는데, 이유는 청취자의 의식 속에 한글 듣기에서 사용되는 여러 가지 지식들과 한글의 음운변동이 무의식적으로 작용하기 때문이다.

따라서 본 연구에서는 각 기본 자음과 기존 자음의 음운 변동을 통한 파생 자음들을 하나의 그룹으로 묶어 각 그룹들이 같은 발음을 받아적은 것을 수 있을 가능성을 고려한다.
성모 /b/는 한글 /ㅂ/의 발음과 유사하지만, /ㅂ/의 음운 변동에 따라 다르게 들릴 수 있는 /ㅂ/, /ㅍ/, /ㅃ/, /ㅁ/ 등을 하나의 그룹으로 보고 이 네 가지 한글 자음의 경우 성모 /b/에 해당하는 것으로 본다. 이와 같이 그룹화 된 성모의 한글 발음들을 정리한 것이 Table 4이다.

Table 4 Hangeul Consonants corresponding to Chines Initials

	Initial	Representative consonant	Similar consonant
1	b	ㅂ	ㅍ ㅃ ㅁ
2	p	ㅍ	ㅂ ㅃ ㅁ
3	m	ㅁ	
4	f	ㅍ	ㅂ ㅃㅁ
5	d	ㄷ	ㅌ ㄸ ㄴ ㅈ
6	t	ㅌ	ㄷ ㄸ ㄴ ㅈ
7	n	ㄴ	ㄹ
8	l	ㄹ	ㄴ
9	g	ㄱ	ㅋ ㄲ ㅇ
10	k	ㅋ	ㄱ ㄲ ㅇ
11	h	ㅎ	
12	j	ㅈ	ㅊ ㅉ
13	q	ㅊ	ㅈ ㅉ
14	x	ㅅ	ㅆ
15	zh	ㅈ(즈)	츠 쯔
16	ch	ㅊ(츠)	즈 쯔
17	sh	ㅅ(스)	쓰
18	r	ㄹ(르)	느
19	z	ㅉ(쯔)	즈 츠
20	c	ㅊ(츠)	즈 쯔
21	s	ㅆ(쓰)	스

Table 4의 그룹화 된 한글 받아쓰기를 적용해 보면 '主意' /zhu yi/ 를 입력하기 위해 '주이'로 입력해야 하지만 본 연구에서는 사용자가 '쭈이' 또는 '추이' 를 입력하는 경우에도 'ㅉ' 과 'ㅊ' 이 'ㅈ' 의 유사 자음으로 간주하여 '주이'로의 변환을 고려하게 된다.

다음으로 운모의 발음에 대해 살펴본다. 한어병음에서 운모는 6개의 단운모와 이들의 결합으로 만들어진 30개의 결합운모 등 총 36개로 구성된다. 기본 운모는 /a/, /o/, /e/, /i/, /u/, /ü/

등 6개인데 각 운모들은 다시 하나 또는 두개의 /a/ /o/ /e/ /i/ /n/ /ng/ /u/ /r/ 등과 결합하여 30개의 결합 운모를 구성한다.

/i/ + /a/ = ia /i/ + /a/ + /ng/ = /iang/ /i/ + /o/ + /u/ = /iou/

앞에서 언급한 바와 같이 한어 병음의 운모는 한글의 모음 역할을 하는데, /a/ 는 한글 /ㅏ/ 발음에 해당되며 /i/는 한글 모음 /ㅣ/ 에 해당한다고 할 수 있다. 그러나 한글 모음들 중에는 듣기에 따라 /ㅏ/와 유사하게 받아 적을 수 있는 모음들이 있다. 예를 들어 '筲' 의 경우 병음 표기가 /shao/ 인데 발음을 한글로 받아 적으면 '사오' 로 적어야 하겠지만 듣는 사람에 따라서는 '샤오' 로 받아 적을 수 있다. 따라서 각 운모들은 성모와 마찬가지로 한글 모음 여러 가지로 받아 적을 수 있는 것이다.

Table 5에서는 먼저 단운모 10 개의 한글 표기에 대해 살펴본다. Table 5의 방법 1은 문화체육관광부 외래어 표기법을 사용하는 경우이며, 방법 2는 Kim and Choi (1997), 방법 3은 Eom (2002)이 각각 제안한 방법을 사용하는 경우이다.

Table 5 Hangeul Notations of Chines Finals

Basis final	method1	method2	method3
a[ɑ]	ㅏ	ㅏ	ㅏ
o[o]	ㅗ	ㅓ	ㅗ
e[ə][ɤ]	ㅓ	ㅓ	ㅓ
i[i]	ㅣ	ㅣ (ㅡ)	ㅣ(ㅡ)
u[u]	ㅜ	ㅜ	ㅜ
ü [y]	ㅟ	ㅟ	ㅟ

Table 5를 살펴보면 현재 사용되고 있는 주요 한글 표기법 상에서는 /o/에 대한 표기를 제외한 다른 운모들은 모두 동일하게 표기하고 있다. /o/의 경우 /ㅗ/로 표기할 수 도 있지만 실제 중국 사람들의 발음을 들어보면 /ㅓ/에 더 가깝다고 판단한 경우로 보인다. 또한 /i/의 경우는 /z/, /c/, /s/등의 설치음과 /zh/, /ch/, /sh/, /r/ 등의 권설음 다음에 사용할 경우는 /으/ 발음에 가깝기 때문에 /ㅣ/, 와 /ㅡ/ 두 가지 표기법을 모두 사용하는 경우가 있다.

운모 역시 발음을 받아 적을 때 성모와 마찬가지로 위에서 제시하고 있는 모음들과 다른 모음으로 인식할 수 있다. 예를 들어 韶 /shao/ 를 '사오'로 적지 않고 '샤오' 로 적는다던지, 卒 /zu/ 를 '쭈' 대신 '쥬' 로 표기할 수 있다.

이는 우리나라 모음의 기본 구성 원소인 'ㆍ', 'ㅡ', 'ㅣ' 들이 하나씩 결합되어 구성되는 모음을 초출자라고 하는데 초출자인 'ㅗ', 'ㅏ', 'ㅜ', 'ㅓ' 에 'ㆍ' 가 덧붙여진 재출자의 형태로 'ㅛ', 'ㅑ', 'ㅠ', 'ㅕ' 가 만들어졌기 때문에 외국어 발음상에서는 유사하게 들릴 수 있다.

따라서 본 연구에서는 이러한 가능성을 모두 포함하여 단운모에 대응하는 한글 모음 그룹을 Table 6과 같이 정의한다.

Table 6 Hangeul Vowels corresponding to Chines Finals

Basis final	Basic vowel	Similar vowel
a[ɑ]	ㅏ	ㅑ
o[o]	ㅗ(ㅓ)	ㅛ(ㅓ ㅕ)
e[ə][ɤ]	ㅓ	ㅕ (ㅖ,ㅒ:case as a combination Finals)
i[i]	ㅣ (ㅡ)	(ㅟ) (case of combining with zh-ch-sh)
u[u]	ㅜ	ㅠ
ü[y]	ㅟ	ㅣ

이제 결합운모들에 대해 살펴본다. 결합운모는 말 그대로 단운모들의 결합으로 만들어진 운모들이다. 예를 들어 결합운모 /ao/ 는 /a/와 /o/ 가 결합되어 /아오/ 발음이 난다. 따라서 결합운모에 대한 대표 모음은 단운모의 대응 모음의 조합으로 대부분 나타낼 수 있다.

/ai/ = /a/ + /i/ = {아, 야} × {이, 으} = {아이, 아으, 야이, 야으}

그러나, 단운모의 조합으로 결합운모를 만들 경우 단운모 때와는 다르게 발음이 나는 경우가

있다. 단운모 /e/ 의 경우 단독으로 쓰일 경우는 'ㅓ' 발음이 나지만 /i/와 만나 결합운모 /ei/ 로 쓰일 경우는 운모 /e/가 'ㅔ' 발음이 나서 /에이/로 발음된다. /ian/ 도 마찬가지의 경우인데, /ia/는 /이아/로 발음되지만 /n/이 결합될 경우 /이안/이 아닌 /이엔/으로 발음된다. 특히 모음 'ㅔ'의 경우는 'ㅐ' 로 받아 적는 경우를 고려해야 하는데, 실제 두 모음은 다른 발음이지만 최근에는 전설 평순 모음인 'ㅔ' 와 전설 평순 저모음인 'ㅐ'의 구별이 점차 힘들어지고 있어 우리나라 사람들도 발음을 듣고 구별해 내기가 어렵다. 특히 외국어를 듣고 받아 적을 때는 더욱 더 차이를 거의 느끼지 못한다. 외래어 표기 규정에서는 외래어를 적을 때 'ㅐ' 대신 'ㅔ' 만을 사용하도록 규정하고 있지만 외래어 규정을 숙지하지 못하고 있는 경우는 'ㅐ' 로도 적을 수 있기 때문에 본 연구에서는 'ㅔ'와 'ㅐ' 를 같은 것으로 본다.

/ei/ : /어이/ -> /에이/ or /애이/
/ian/ : /이안/ -> /이엔/ or /이앤/

또 다른 경우로 결합운모에서는 모음이 연속적으로 나타나는 경우 연음으로 발음되어 이중모음의 형태로 들리는 것을 고려해야 한다. 한글 음운체계에서는 이를 '이중모음화'라고 한다. 예를 들어 /ia/ 는 /이아/로 세분화하여 발음될 수도 있지만 대부분 연음으로 발음되어 이중 모음 /야/로 들린다. 또 /iao/는 /이아오/ 이지만 /야오/ 로 들릴 수 있다.

한글에서 이중모음은 'ㅑ', 'ㅕ', 'ㅛ', 'ㅠ', 'ㅒ', 'ㅖ', 'ㅘ', 'ㅙ', 'ㅝ', 'ㅞ', 'ㅢ' 등 총 11개 인데 그 중 중국어에서 결합운모가 이중모음으로 발음될 수 있는 경우는 Table 7의 6가지 경우이다. 특히 결합운모가 성모 없이 단독으로 사용되는 경우에는 모두 이중모음으로 발음된다. 이때에는 /i/ 와 결합되는 결합운모는 /i/ 가 /y/ 로 대체되며 /u/ 결합은 /w/로 대체된다. 특히 /ü/와 결합되는 결합운모는 /y/가 추가되고 /ü/ 는 /u/로 바뀌어 나타난다.

Table 7 Combining Finals pronounced as a Diphthong

Diphthong	Combining final
ㅑ(ㅣ + ㅏ)	/ia/ (/ya/)
ㅖ(ㅣ + ㅔ)	/ie/ (/ye/)
ㅛ(ㅣ + ㅗ)	/iou/(/you/)
ㅘ(ㅗ + ㅏ)	/ua/(/wa/)
ㅝ(ㅜ + ㅓ)	/uo/(/wo/)
ㅞ(ㅜ + ㅔ)	/uei/(/wei/)

* () Combining final when used alone without initial

결합운모에 사용되는 운모 중에 /n/, /ng/, /r/은 기본 운모 또는 결합운모와 다시 결합하여 새로운 결합운모로 사용되는데, 한글 음가로 치면 받침 ' ㄴ', ' ㅇ', ' ㄹ' 에 해당한다.

/a/ + /n/ = /an/ = '안' or '아안'
/eng/ = /e/ + /ng/ = '엉' or'으엉'
/er/ = /e/ + /r/ = '얼' or '으얼'

이상에서 중국어 성모 운모를 한글로 받아 적을 때 어떻게 적을 수 있는지 살펴보았다. 이제 중국어 단어를 입력하기 위해서 한글 발음 사용하는 예를 생각해 보자.

미국 도시 샌프란시스코는 중국어로 '三藩市' 인데, 중국어 병음 /San fan shi/ 대신 본 연구의 한글 병음 입력을 쓰면 /산판시(스)/ 뿐만 아니라 /싼판스/, /싼판씨/, /싼반시/ 등도 병음을 통한 입력에 사용할 수 있다. 실제로 '三藩市'를 소리적기 해보면 /산판시/로 적지 않고 /싼판씨/로 적는 경우가 대부분이다.

4. 한글 발음을 이용한 중국어 입력 시스템

본 절에서는 한글 발음을 이용한 중국어 입력 시스템의 설계와 구현에 대해 살펴본다. 기존의 한어 병음 중국어 입력기는 해당 한자와 로마자 병음이 미리 입력된 중국어 사전을 이용하여 단

어를 검색한다. 예를 들어 '判賠'의 경우 로마자 병음인 /pan pei/가 미리 입력되어 있는 한어 병음 사전을 사용한다.

본 연구에서는 병음 사전에 기록된 로마자 대신 한글을 사용한다. 이를 처리하는 가장 간단한 방법은 한글 한어 병음 사전을 만드는 것이다. 그러나 이 방법은 해당 로마자 병음에 대한 한글 병음을 모두 알고 있어야 하며 대응된 한글 병음이외의 유사한 한글 입력에 대해서는 처리할 수 없음을 Yun et el. (2017)에서 살펴볼 수 있다. 물론, 모든 해당 중국어 단어에 대해 가능한 한글 발음을 모두 사전의 구성 요소로 유지할 수 도 있겠지만 이 방법은 사전을 검색하는데 많은 어려움이 따르거나, 새로운 한글 발음이 추가될 때 마다 한글 병음 사전을 새로 만들어야 하는 단점이 있다.

따라서 본 연구에서는 여러 가지 입력이 가능한 한글 병음을 처리하는 전처리기를 설계하고 이를 통해 중국어를 입력하도록 시스템을 설계한다. Fig 3은 본 연구에서 제안하는 한글 발음을 이용한 중국어 입력 시스템 전체적인 구조이다.

먼저 입력하고자 하는 중국어에 대해 한글로 해당 중국어의 발음을 입력하며 한글 병음 전처리기에서는 각 음절의 자음과 모음을 분리한다. 각 한글 자음과 모음에 해당하는 성모와 운모를 찾고 이들을 조합해서 로마자 음절을 재구성한다. 이 때 한글 자음과 모음에 대응되는 성모나 운모가 여러 가지가 될 수 있으며 반대의 경우도 마찬가지이다. 예를 들어 사용자가 '비애'를 입력한 경우 처리되는 과정을 살펴본다.

먼저 자음 'ㅂ' 은 성모 /b/ 에 해당되는데, 성모 /p/, /f/ 의 유사 대응 자음일 수도 있다. 마찬가지로 모음 'ㅣ' 는 운모 /e/ 에 해당하거나 운모 /ü/ 의 유사 대응 자음 일 수 있다. 따라서 결론적으로 한글 음절 '비' 는 /bi/, /pi/, /fi/, /bü/, /pü/, /bü/ 등으로 병음 음절이 생성될 수 있다.

생성된 한어 병음 음절은 다음의 두 가지 규칙에 따라 걸러진다.

① 결합하지 못하는 성모와 운모를 포함하는 병음 음절은 걸러낸다.
② 결합 가능한 음절은 대표 글자 조합을 우선한다.

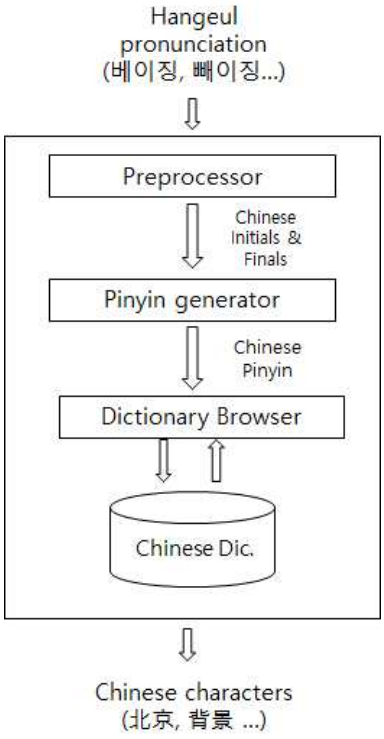

Fig. 3 Proposed System Structure

규칙 1에 의해 위의 병음 음절 중 /fi/,/bü/, /pü/, /bü/ 결합되지 못하는 조합으로 탈락 된다. 남은 /bi/, /pi/, /fi/ 중 ㅂ 은 성모 /b/의 대표 대응 자음이므로 /bi/를 우선적으로 처리한다.

같은 방법으로 두 번째 음절을 처리하여 애는 /e/가 결합 운모로 사용될 때에 해당함으로 최종적으로 /bie/, /pie/, /fie/ 를 해당 로마자 병음으로 선정한다. 이 때 대표 대응 자모음으로 생성된 /bie/ 를 우선하여 단어 검색기를 통해 해당 글자를 찾아 사용자에게 제시한다.

본 연구에서는 중국어 단어 검색을 위해 발음에 대한 한어 병음이 같이 표기되어 있는 공개 중국어 사전인 CC-CEDICT를 사용하였다.

Fig. 4는 한글 발음 '비에'에 대해 중국어 단어를 추천어로 제시한 예이다.

Fig. 5는 본 연구에서 제안하는 한글 발음을 이용한 중국어 입력 시스템을 중국어 사전 검색 시스템에 적용한 예이다.

5. 성능평가

본 절에서는 본 연구에서 개발한 한글을 이용한 한어 병음 입력기의 성능에 대해 평가한다.

한글 한어 병음 입력기에서는 사용자가 입력을 원하는 한자에 대한 한글 음을 입력하였을 때 리스트로 제공되는 추천어에 원하는 단어가 포함되어 있어 입력에 성공할 수 있는지에 대한 여부가 중요하다.

따라서 본 연구에서는 제안하는 한글 한어 병음 입력기를 이용할 경우와 국어연구원의 표준 외래어 표기법을 이용하는 경우를 비교하여 사용자가 입력에 성공하는 경우를 비교한다.

성능 평가는 중국어를 학습하고 있는 100명의 학습자를 대상으로 1음절, 2음절, 3음절, 4음절의 단어 각각 10개 씩 들려주고 이를 입력할 때 입력 성공률을 평균해서 비교한다.

Fig. 6은 두 방법 간의 입력 성공률을 그래프로 나타낸 것이다.

Fig. 6의 두 방법 간의 입력 성공률을 분석해 보면 표준외래어 방법에서는 전체 평균 78.5%의 성공률을 보이고 있으며 본 연구에서 제안하는 방법은 평균 90.75% 의 입력 성공률을 보여 본 연구에서 제안하는 방법의 성능이 더 우수함을 알 수 있다.

음절수 별로 성공률을 분석해 보면 표준외래어 방법에서는 1음절 단어의 경우 약 92%의 성공률을 보이며 음절수가 많아질수록 낮아져서 4음절 단어의 경우 약 65 % 정도로 성공률이 급격히 낮아진다. 반면에 본 연구에서 제안하는 방법은 1음절의 경우 약 95% 성공률을 보이고 있으며 4음절의 경우에도 85% 의 성공률을 보이고 있어 음절수가 많아질수록 두 방법 간의 성공률의 차이가 더 커진다. 이는 음절이 많아지게 될 경우 발음상의 연음이나 동화에 의해 정확하게 발음을 기억하고 입력하기가 더 어려워지게 되는데, 본 연구에서 제안하는 방법은 사용자의 오류에 대한 예측을 통해 유사하게 입력될 음소들을 고려하기 때문에 이 경우에도 상당부분 입력을 성공 시킬 수 있기 때문이다.

6. 결 론

지금까지 중국어를 컴퓨터나 스마트폰에 입력하기 위해서는 한어 병음 시스템을 사용하고 있다. 그러나 한어 병음 시스템은 중국어 발음을 그대로 표기할 수 없기 때문에 우리나라 사람들이 한어 병음 시스템을 이용하여 중국어를 입력하기에 매우 어렵다. 따라서 본 연구에서는 발음

别	别	1	bie	비에
憋	憋	1	bie	비에
瘪	瘪	1	bie	비에
瘪	瘪	1	bie	비에
擎	擎	1	pie	피에
撇	撇	1	pie	피에
撇	撇	1	pie	피에

Fig. 4 Example of Recommendation List

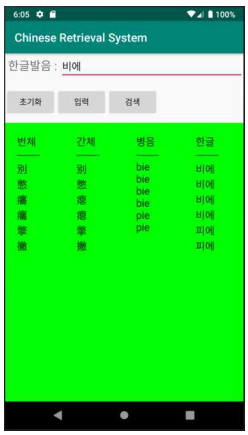

Fig. 5 Chinese Dictionary Retrieval System

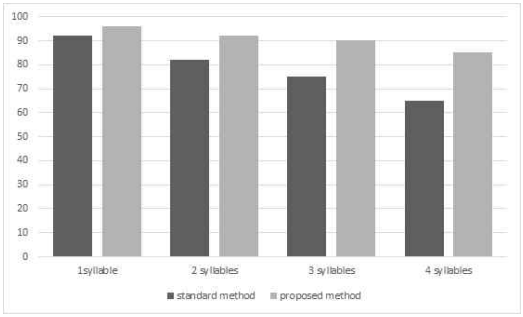

Fig. 6 Input Success Rate Comparison

을 표기하는 문자로 가장 우수한 우리 한글을 이용하여 중국어 발음을 적고 이를 이용하여 중국어를 입력할 수 있는 중국어 입력기를 설계하였다. 본 연구에서 제안하는 시스템에서는 중국어 발음을 유사하게 표기하는 경우에도 중국어 입력이 가능하게 함으로써 입력에 성공하는 확률을 높였으며, 성능평가를 통해 이를 검증하였다.

기존의 많은 연구들을 통해 우리 한글의 우수성을 검증함과 동시에 한글 확산 및 세계화를 위해 많은 노력을 기울이고 있는데, 본 연구도 이러한 측면에서 더욱 의미를 가질 수 있다.

향후 연구를 통해 본 연구에서 제안하는 중국어 입력기의 효율성을 더욱 발전시켜 나갈 계획이며 중국어뿐만 아니라 다른 여러 나라 언어들 중 로마자 병음을 입력 방법으로 사용하고 있는 언어들에 대해 한글을 사용할 수 있는 방안을 연구하여 한글의 세계화에 노력할 계획이다.

References

Ahn, M. T. (2007). ANYIN Ahnmatae ZhengYin Chinese Hanzi Input System Is the First Step for Globalization of ZhengYin, *Korean Language Information Science*, 9(1), 8-16.

Chen, Z., and Lee, K. F. (2000). A New Statistical Approach to Chinese Pinyin Input. *The 38th Annual Meeting of the Association for Computational Linguistics*, pp. 241-247.

Eom, I. S. (2002). Chinese Hangul Notation Revision Plan, *The Journal of Chinese Language and Literature*. 31(1), 111-135.

Gao, W. B., Peng, H. B., Xu, C., and Zhang, Q. W. (2015). Chinese Character Entry via a Pinyin Input Method, *US Patent,* Application No 14. 604, 506.

Gottardo, A., Koh, P. W., Chen, X., and Jia, F. (2017). Models of English and Chinese Word Reading for Adolescent Chinese - English Bilinguals, *Reading and Writing*, 30(7), 1377-1406.

Jeong, W. S. (2007a). A Study on the Hangeul Writing System of Chinese Characters, *Korean Language & Literature,* 62(1), 145-181.

Jeong, W. S. (2007b). A Study on the Hangeul Writing System of Chinese Yunmu Characters, *HAN-GEUL*, 278(1), 67-111.

Kim, Y. O., and Choi, Y. E. (1997), *How to Study Oriental Studies*, Seoul, Tongnamu.

Pasquarella, A., Chen, X., Gottardo, A., and Geva, E. (2015). Cross-language Transfer of Word Reading Accuracy and Word Reading Fluency in Spanish-english and Chinese-english Bilinguals Script-universal and Script-specific Processes, *Journal of Educational Psychology*, 107(1), 1-96.

Shen, H. H., Tsai, C. H., and Zhou, Y. (2007). *Introduction to Standard Chinese Pinyin System*, Beijing, Beijing Language and Culture University Press.

Tan, L. H., Xu, M., Chang, C. Q., and Siok, W. T. (2013). China's Language Input System in the Digital Age Affects Children's Reading Development, *Proceedings of the National Academy of Sciences*, Jan. 1119-1123, San Antonio, USA.

Wong, L. H., Chai, C. S., and Gao, P. (2011). The Chinese Input Challenges for Chinese as Second Language Learners in Computer-Mediated Writing: An Exploratory Study, *Turkish Online Journal of Educational Technology-TOJET*, 10(3), 233-248.

Xiao, G., Guo, J., Gong, Z., and Li, R. (2016). Pinyin-Senses Input Method for Semantic Document Exchange in E-Business. *2016 IEEE 13th International Conference on e-Business Engineering (ICEBE)*, Nov. 311-315. Macau, China

Yun, H. W., Wang Y., Kim, P. K., and Lee, J. H., (2017), Chinese Retrieval system using

Hangeul Pronunciation of Chinese Language, *Information*, 20(9A), 6233-6243.

이 중 화 (JungHwa Lee)

- 정회원
- 부산대학교 전자계산학과 이학사
- 부산대학교 전자계산학과 이학석사
- 부산대학교 전자계산학과 이학박사
- (현재) 동의대학교 컴퓨터소프트웨어공학과 교수
- 관심분야: 데이터베이스, 한글정보처리

윤 홍 원 (HongWon Yun)

- 정회원
- 부산대학교 전자계산학과 이학사
- 한국외국어대학교 전자계산학과 이학석사
- 부산대학교 전자계산학과 이학박사
- (현재) 신라대학교 컴퓨터소프트웨어공학부 교수
- 관심분야: 데이터베이스, 빅데이터, 한글정보처리