

論文91-28B-12-2

한국어 음운 변동 처리를 위한 효율적인 Rule Base System의 구성

(Implementation of an Effective Rule Base System for the Change of Korean Vocal Sound)

李 桂 英* 李 相 範**

(Gye Young Lee and Sang Burm Rhee)

要 約

본 논문에서는 한국어 음운 변동 현상에 우선 순위를 두어 처리하는 기존의 방법에 대한 문제점을 해결할 수 있는 음운 변동 Rule Base를 제안한다.

제안된 Rule Base는 음성 처리에 사용되는 표기 언어(한글)와 음성 언어(한국어) 사이의 문제 해결이 가능하며, 이를 위하여 한글 맞춤법의 표준 발음법을 근거로, 각 규정별 음운 변동 현상을 받침별 Rule Base로 정의하였다.

본 연구에서 제시된 Rule Base는 음운 변동 처리 시스템의 구현을 통한 실험에 의하여, 더욱 간략화될 수 있음을 보였으며, 음절 단위의 한국어 음성 처리를 위한 음가 정보의 생성에 효율적으로 이용될 수 있다.

Abstract

In this Paper, a rule-based method for the phenomenon of Korean vocal sound change is proposed.

This method could be used to solve a problem between symbolic(Hangul) and phonetic language (Korean) for the study of Korean speech processing.

A rule on the phenomenon of vocal sound rearranged for the rule base with a end-consonants on the authority of standard pronunciation rule.

The proposed rule base system is simplified by the implementation for the vocal sound change. Also, it is useful to create the data base with phonetic value for the Korean voice processing by syllable unit.

I. 서 론

인간과 기계간의 정보 교환을 위하여 음성을 이용하고자 하는 음성 정보처리에 대한 연구가 활발히 진행되고 있으며, 이에 따라 음성의 물리학적 특징과 음성·음운·언어학적 정보 사이의 복잡한 관계를 규명하려는 연구가 시도되고 있다.^{1,2} 음성 언어를 이해

*正會員, 東國大學校 電子計算學科
(Dept. of Computer Sci., Dongguk Univ.)

**正會員, 檀國大學校 電子工學科
(Dept. of Elec. Eng., Dankkook Univ.)

接受日字: 1991年 8月 16日

하려는 시스템의 구축은 음성 언어의 물리적 현상의 규명³⁻⁷⁾이 외에도, 언어의 본질을 이해하는 언어학이나 심리학 입장에서의 연구, 자연어 텍스트 처리분야의 연구 및 음성 처리 분야의 연구가 결합된 형태를 갖추어야 할 것이다.

현재, 자연어 텍스트에 대한 형태소, 구문, 의미, 문맥 처리 등에 대해서는 각종 연구⁸⁾가 이루어지고 있으나, 한국어 음성 처리와 자연어 텍스트 처리의 연결 과제인 음운 변동 규칙에 대한 연구는 한국어의 구조적, 음운·음성학적인 특징으로 인해 많은 어려움이 있는 실정이다.^{9,10)} 따라서, 음성 인식시의 인식된 한국어 음가(단어 또는 어절 단위)를 완성된 한글 단어로 표기하여 주거나 입력된 한글 단어 또는 어절을 음가 정보로 변환시켜 음성 합성¹¹⁾을 시도하는 경우에는 많은 어려움이 발생하게 된다.

기존의 음운 변동 처리는 자음 접변, 경음화, 구개음화, 격음화, 중성법칙, 연음 법칙, 절음 법칙, 음운 첨가, 음운 탈락 등을 적용할 때, 각 규칙의 충돌을 방지하기 위하여 변동 규칙에 우선 순위를 두어 변동 현상을 처리하여 왔다.¹²⁾ 이러한 처리 방식은 어떤 단어나 어절에 필요치 않은 변동 현상을 무의미하게 적용시키게 되는 문제점이 있으며, 정확한 변동 결과를 얻기 위해서는 처리의 횟수가 길어진다는 단점이 있다.

본 연구에서는 이러한 문제점을 해결하기 위하여, 한글 맞춤법의 표준 발음법^{12,13)}에 근거한 음운 변동 규칙에 대해서, 음절의 중성을 기준으로 한 27개의 rule base로 설계하고, 어떠한 입력 단어에 대해서도 음절간의 음운 변동 결과는 1회의 처리로 얻을 수 있게 하였다. 그리고, 이에 따라 생성된 음가 정보를 사전에 적재시켜 음성 처리의 자료로 사용할 수 있는 방법을 연구하였다.

II. 표준 발음법에 의한 음운 변동 규칙

1. 표준어 규정의 표준 발음법

한글 맞춤법의 표준 발음법은 종래의 없던 규정으로, 모두 7장 30항으로 구성되어 있다. 제1장 총칙 제1항에 규정된 바와 같이, 한국어의 표준 발음은 현재 사용되고 있는 언어 현실을 수용하여 표준어의 실제 발음을 정하고 있으며, 실제 발음에서 여러가지로 나타날 수 있는 경우에는 국어의 전통성과 합리성을 따르기로 하였다. 총 7장으로 구성된 표준 발음법 중, 본 연구에서 다루고자 하는 내용은 제 4, 5, 6장으로, 음의 길이와 관련된 제3장과 합성어와 파생어에서의 음의 첨가 현상과 관련된 제7장은 논외로

하였다. 음의 길이와 합성어 및 파생어에 대한 음운 변동은 한국어(순수 우리말/한자어)의 단어 및 개념 레벨의 전자 사전이 구성된 후 처리되는 것이 rule base의 간략화를 도모할 수 있으며, 제한된 단어 레벨의 처리시에는 비교적 간단하게 rule을 확장시킬 수 있다.

2. 각 항별 음운 변동 규칙의 설계

음운 변동 rule base는 우선 각 규정별로 규칙을 정형화한 다음, 음운 변동에 있어서 가장 문제가 되는 음절의 중성을 key로 하는 rule base로 설계한다. 이러한 rule base는 모음이 이중적으로 발음됨을 허용하는 제5항의 ‘ㅈ, ㅊ, ㄴ’, 발음 규정과 합성어 및 파생어 사전의 구성을 전제로 하는 제7장의 음의 첨가 현상을 제외하고는 표준어 규정의 표준 발음법을 모두 수용한다. 표1은 rule 작성을 위한 기호의 정의를 나타내었다.

표 1. Rule base의 기호 정의
Table 1. The symbol of rule base.

기호	정 의	기호	정 의	기호	정 의
C	음절의 초성	V	음절의 중성	J	음절의 종성
/	자모 연결 조건	:	조건 만족시 처리	:	계속 처리
,	조건 OR	()	자모 1개의 처리		자모 집합의 정의
{ }	자모 유형 조건	{T-}	THEN 처리	{E-}	ELSE 처리
...>	자모 변경 처리	n	마지막 음절	i	앞 음절
j	뒤 음절	ATBL	조사	BTBL	어미
CTBL	접미사	DTBL	실질 형태소	ETBL	한자어

1) 받침의 발음에 관한 각 항별 rule 정의

한국어의 음절말에 대한 발음 규정으로, 받침소리로 발음되는 자음은 총 7개이다. 따라서, 제 8항에서는 이들 이외의 받침이 중화되어 7개 중(ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅇ)의 하나로 발음됨을 규정하였다. 제 9, 10, 11항은 겹받침의 대표음에 대한 규정으로, 마지막 음절과 초성 자음 앞에서의 겹받침에 대한 발음 규칙을 정하고 있다. 제12항은 받침 ‘ㅎ, ㄴ, ㄷ’에 대한 발음 규칙으로, 뒤음절 초성의 날자 유형에 따라서 어떻게 발음되는 가를 규정하고 있다. 받침 뒤에 모음으로 시작되는 어미, 접미사, 조사가 결합될 때의 발음 규칙은 제 13, 14, 15항에서 정하였다. 제 16항은 앞서 규정한 발음 규칙에 예외적 처리를 하여야 하는 한글 자모의 이름에 대한 발음 규정으로 예외 처리 사전의 구성이 필요함을 보여주고 있다. 표2는 다음과 같은 표준 발음법 제13항에 관한 rule 정의와 그 예이다.

〈제13항〉

홀받침이나 쌍받침이 모음으로 시작된 조사나 어미, 접미사와 결합되는 경우에는, 제 음가대로 뒤 음절 첫소리로 옮겨 발음한다.

표 2. 13항의 Rule 정의
Table 2. The rule definition of section 13.

규정	Rule 정의	예
R130	$J_1 \{ \text{ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅆ, ㅈ, ㅊ, ㅋ, ㆁ, ㅍ, ㅌ} / C_i(\text{o}) \langle \text{ATBL, BTBL, CTBL} \rangle : J_1(\dots) J_1(\dots); C_i(\dots) \dots \rangle$ $C_i \{ \text{ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅆ, ㅈ, ㅊ, ㅋ, ㆁ, ㅍ, ㅌ} /$	작은/작근, 섞어/서겨, 물음/무늘, 굳어/구더, 몸이/모미, 맛이/마시, 낮에/나새, 늘어진/느러진, 있어서/이세서 등

2) 음의 동화에 관한 각 항별 rule 정의

제5장은 음의 동화 현상에 대한 규정으로, 5장에서 규정한 동화현상 이외는 표준 발음으로 인정치 않기로 규정하고 있다(제 21항).

제17항은 구개 음화, 18항은 ‘ㅇ, ㄴ, ㄹ’ 으로 발음되는 비음화에 대해서, 제19항과 20항은 ‘ㄴ’ 과 ‘ㄹ’ 로의 동화 현상에 대해서 규정하고 있다. 제 20항은 발음에 따라 ‘ㄹ’ 동화가 일어나지 않는 예외적인 단어를 규정하고 있다. 표3은 제17항에 관한 rule 정의와 그 예이다.

표 3. 제17항의 Rule 정의
Table 3. The rule definition of section 17.

규정	Rule 정의	예
R170	$J_1 \{ \text{ㄷ, ㅌ, ㅍ, ㅌ} / C_i(\text{o}) / V_i(\text{l}) \langle \text{ATBL, CTBL} \rangle : J_1(\dots) J_1(\dots); C_i(\dots) \dots \rangle C_i \{ \text{ㅅ, ㅆ, ㅈ, ㅊ} /$	굳어/구지, 발이/바치, 훑이다/훑치다 등
R171 (붙임)	$J_1(\text{ㄷ}) / C_i(\text{ㅎ}) / V_i(\text{l}) \langle \text{ATBL, CTBL} \rangle : J_1(\dots) J_1(\dots); C_i(\dots) \dots \rangle C_i(\text{ㅎ})$	굳히다/구치다, 물히다/무치다 등

3) 경음화 현상에 대한 각 항별 rule 정의

제 23항은 ‘ㄱ, ㄷ, ㅂ’ 받침에 연결되는 초성 자음 ‘ㄱ, ㄷ, ㅂ, ㅅ, ㅆ’ 이 예외 없이 된소리로 발음됨을 정하고 있으며, 제24, 25항은 용언의 어간 받침과 어미의 첫소리가 만나 어미의 첫소리가 된소리로 발음됨을 규정하고 있다. 제 26항은 ‘ㄹ’ 받침으로된 한자어의 경우, 뒤에 연결되는 초성 자음 ‘ㄷ, ㅅ, ㅆ’

이 된소리로 남을 규정하고 있다. 제27항은 관형사형 ‘-(으)ㄹ’ 뒤에서 연속하여 발음되는 ‘ㄱ, ㄷ, ㅂ, ㅅ, ㅆ’ 과 어미가 된소리화 되는 규정이며, 28항은 관형격 기능을 갖고 휴지기가 성립하는 합성어의 뒷 단어 첫소리 ‘ㄱ, ㄷ, ㅂ, ㅅ, ㅆ’ 의 된소리 규정이다. 제27항과 28항은 텍스트 분석 기능 및 합성어 사전이 구성된 후, 간략한 rule 수정의 후처리가 행해져야 한다. 표4는 제23항을 rule로 정의한 것이다.

표 4. 제 23항의 Rule 정의
Table 4. The rule definition of section 23.

규정	Rule 정의	예
R230	$J_1 \{ \text{ㄱ, ㄷ, ㅂ, ㅅ, ㅆ, ㄴ, ㄷ, ㅂ, ㅅ, ㅆ, ㅈ, ㅊ, ㅋ, ㆁ, ㅍ, ㅌ, ㅍ, ㅌ, ㅍ, ㅌ} / C_i \{ \text{ㄱ, ㄷ, ㅂ, ㅅ, ㅆ} : C_i(\dots) C_i \{ \text{ㄱ, ㄷ, ㅂ, ㅅ, ㅆ} /$	학교 / 학교, 넓과/넉과 잇다 / 일떠, 늦다/느떠, 설블리 / 설플리, 뽀다 / 뽀따 등

Ⅲ. 음운 변동 Rule Base 시스템

1. 받침을 이용한 rule base의 생성

음운 변동 규칙의 우선 순위에 관계없이 변동 결과의 처리를 일원화하기 위하여, 표준 발음의 각 항별 rule을 받침별로 설계한 음운 변동 rule base를 생성한다. rule base는 각 음절의 중성을 중심으로 다음 음절의 초성 또는 중성/중성의 결합 형태에 따라서 처리가 행해지며, 해당 rule base에 적용되지 않을 때에는 음운 변화가 없는 것으로 간주된다. 그리고 조사, 어미, 접미사의 참조 조건과 제한된 범위 내에 있는 한자어의 참조 조건을 각 rule에 명시하여, 필요시에는 허당되는 데이터 베이스를 참조하게 된다.

예외 조건의 처리는 입력된 단어 또는 어절에 대한 데이터 베이스 검색 후, 예외 tag를 확인하여 예외 사전을 검색하게 된다. 받침에 의한 27개 rule base는 표5에서 보였다. 단 표5에서 표시된 A, B, C, D, E, 는 표1의 ATBL, BTBL, CTBL, DTBL, ETBL 과 각각 대응되며, R_{xy}에서 xx는 표준 발음법의 해당 항 번호를, y는 해당 항에서 규정된 순서를 나타낸다. 그리고, RULE_{zz}에서 zz는 규칙 적용후 다른 RULE로 순환 및 반복 적용됨을 표시하며, ‘+’ 는 두가지 조건의 모든 만족을, ‘=’ 또는 ‘≠’ 은 조건 만족 여부에 따라 각각의 결과가 달라짐을 표시한다.

표6은 받침 ‘ㄴㅎ’을 위한 rule base의 예이다. 받침 ‘ㄴㅎ’에 적용되는 규정은 모두 5개이다.

표 5. 받침의 Rule base (1)

Table 5. The rule base by the end-consonants.

종성Ji 초성Cj	RULE01		RULE 02		RULE03		RULE04		RULE05		RULE06		RULE07		RULE08		RULE09					
	ㄱ		ㄲ		ㄴ		ㄷ		ㄸ		ㄹ		ㄺ		ㄻ		ㄼ					
	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙	Ji 조건	Cj 규칙				
ㄱ	ㄱ	ㄲ R230	ㄱ	ㄱ RULE01	ㄱ	ㄱ R090	ㄱ	ㄱ RULE01	ㄴ	ㄱ R241	ㄴ	ㄲ R240	ㄴ	ㄲ R120	ㄴ	ㄲ R230	ㄴ	ㄲ R270	ㄴ	ㄴ RULE01	ㄴ	ㄴ R110
ㄲ			ㄱ	ㄲ RULE01	ㄱ	ㄲ R090	ㄱ	ㄲ RULE01	ㄴ	ㄲ RULE04	ㄴ	ㄲ R100							ㄱ	ㄲ RULE01	ㄴ	ㄴ R110
ㄴ	ㅇ	ㄴ R180	ㅇ	ㄴ R180	ㅇ	ㄴ R180		ㄲ R240	ㄴ	ㄴ RULE04	ㄴ	ㄴ R100	ㄴ	ㄴ R125	ㄴ	ㄴ R180	ㄴ	ㄴ R200	ㅇ	ㄴ	ㄴ	ㄴ R180
ㄷ	ㄱ	ㄸ R230	ㄱ	ㄷ RULE01	ㄷ	ㄷ R090	ㄴ	ㄸ R240	ㄴ	ㄸ R240	ㄴ	ㄸ R120	ㄷ	ㄸ R230	ㄷ	ㄸ R260	ㄷ	ㄸ R260	ㄱ	ㄷ RULE01	ㄷ	ㄷ R110
ㄸ			ㄱ	ㄸ RULE01	ㄱ	ㄸ R090	ㄴ	ㄸ RULE01	ㄴ	ㄸ RULE04	ㄴ	ㄸ R100							ㄱ	ㄸ RULE01	ㄸ	ㄸ R110
ㄹ	ㅇ	ㄴ R191	ㄱ	ㄹ RULE04	ㄱ	ㄹ R090	ㄴ	ㄹ R100	ㄴ	ㄹ R200	ㄴ	ㄹ R100							ㄴ	ㄹ RULE01	ㄹ	ㄹ R110
ㅁ	ㅇ	ㅁ R180	ㅇ	ㅁ R180	ㅇ	ㅁ R180			ㄴ	ㅁ R100			ㄴ	ㅁ R180					ㅇ	ㅁ	ㅁ	ㅁ R180
ㅂ	ㄱ	ㅃ R230	ㄱ	ㅂ RULE01	ㅂ	ㅂ R090	ㄴ	ㅃ R280	ㄴ	ㅂ R100	ㄴ	ㅂ R100	ㄷ	ㅃ R230	ㄴ	ㅃ R270	ㄴ	ㅃ R270	ㄱ	ㅂ RULE01	ㅂ	ㅂ R110
ㅃ			ㄱ	ㅃ RULE01	ㄱ	ㅃ R090	ㄴ	ㅃ RULE01	ㄴ	ㅃ RULE04	ㄴ	ㅃ R100							ㄱ	ㅃ RULE01	ㅃ	ㅃ R110
ㅅ	ㄱ	ㅆ R230	ㄱ	ㅅ RULE01	ㅅ	ㅅ R090	ㄴ	ㅆ R240	ㄴ	ㅆ R240	ㄴ	ㅆ R123	ㄷ	ㅆ R230	ㄷ	ㅆ R260	ㄷ	ㅆ R260	ㄱ	ㅅ RULE01	ㅅ	ㅅ R110
ㅆ			ㄱ	ㅆ RULE01	ㄱ	ㅆ R090	ㄴ	ㅆ RULE01	ㄴ	ㅆ RULE04	ㄴ	ㅆ R100							ㄱ	ㅆ G 01	ㅆ	ㅆ R110
ㅇ	· (ABC) (D)	ㄱ R180 R150	· (ABC) (D)	ㄲ R130 R150	(ABC) (D)	ㄴ R140 R151	(ABC) (D)	ㄷ R130 R150	(ABC) (D)	ㄸ R140 R151	(BC) (D)	ㄹ R126 R151	(1+AC) (=) (D)	ㅁ R170 R130 R150	(ABC) (D)	ㄺ R130 R150	(ABC) (D)	ㄻ R130 R150	(ABC) (D)	ㄼ R140 R151	ㄽ R140 R151	ㄾ R140 R151
ㅈ	ㄱ	ㅉ R230	ㄱ	ㅈ RULE01	ㅈ	ㅈ R090	ㄴ	ㅉ R240	ㄴ	ㅉ R240	ㄴ	ㅉ R120	ㄷ	ㅉ R230	ㄷ	ㅉ R260	ㄷ	ㅉ R260	ㄱ	ㅈ RULE01	ㅈ	ㅈ R110
ㅉ			ㅈ	ㅉ RULE01	ㅈ	ㅉ R090	ㄴ	ㅉ RULE01	ㄴ	ㅉ RULE04	ㄴ	ㅉ R100							ㄱ	ㅉ RULE01	ㅉ	ㅉ R110
ㅊ			ㄱ	ㅊ RULE01	ㄱ	ㅊ R090	ㄴ	ㅊ RULE01	ㄴ	ㅊ RULE04	ㄴ	ㅊ R100							ㄱ	ㅊ RULE01	ㅊ	ㅊ R110
ㅋ			ㄱ	ㅋ RULE01	ㄱ	ㅋ R090	ㄴ	ㅋ RULE01	ㄴ	ㅋ RULE04	ㄴ	ㅋ R100							ㄱ	ㅋ RULE01	ㅋ	ㅋ R110
ㆁ			ㄱ	ㆁ RULE01	ㄱ	ㆁ R090	ㄴ	ㆁ RULE01	ㄴ	ㆁ RULE04	ㄴ	ㆁ R100							ㄱ	ㆁ RULE01	ㆁ	ㆁ R110
ㆁ			ㄱ	ㆁ RULE01	ㄱ	ㆁ R090	ㄴ	ㆁ RULE01	ㄴ	ㆁ RULE04	ㄴ	ㆁ R100							ㄱ	ㆁ RULE01	ㆁ	ㆁ R110
ㆁ	·	ㅋ R121	ㄱ	ㆁ RULE01	ㄱ	ㆁ R090	ㄴ	ㆁ RULE01	ㄴ	ㆁ R121			· (C) · (≠C)	ㆁ R171 E R121					ㄱ	ㆁ RULE01	ㆁ	ㆁ R121
공백소			ㄱ		ㄱ				ㄴ										ㄱ			R110

표 5. 받침의 Rule base (2)
Table 5. The rule base by the end-consonats.

종성Ji 초성Cj	RULE10		RULE11		RULE12		RULE13		RULE14		RULE15		RULE16		RULE17		RULE18			
	ㄱ		ㄴ		ㄷ		ㄹ		ㄺ		ㄻ		ㅁ		ㅂ		ㅅ			
	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj
변 환	조건	규칙	조건	규칙	조건	규칙	조건	규칙	조건	규칙	조건	규칙	조건	규칙	조건	규칙	조건	규칙	조건	규칙
ㄱ	ㄱ 어간+B	ㄱ R240	ㄴ 어간+B	ㄴ R250	ㄷ RULE08	ㄷ R100	ㄹ 어간+B	ㄹ R250	ㄺ RULE17	ㄺ R110	ㄻ	ㄻ R120	ㅁ 어간+V, (1, ㅏ) ㅁ 어간+(ㅏ B)	ㅁ R241	ㅂ R230	ㅅ RULE17	ㅅ R100			
ㄴ	ㄱ RULE16	ㄴ R110	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110							ㅅ RULE17	ㅅ R100		
ㄷ	ㄱ RULE16	ㄴ R110	ㄴ R180	ㄷ R180	ㄷ RULE08	ㄷ R110	ㄹ R201	ㄹ R180	ㄺ R125	ㄻ R180	ㄻ	ㄻ R125			ㅁ R180	ㅂ R180	ㅅ R180	ㅅ R180		
ㄷ	ㄱ 어간+B	ㄱ R240	ㄴ 어간+B	ㄴ R250	ㄷ RULE08	ㄷ R100	ㄹ 어간+B	ㄹ R250	ㄺ RULE17	ㄺ R110	ㄻ	ㄻ R120	ㅁ 어간+B	ㅁ R240	ㅂ R230	ㅅ RULE17	ㅅ R100			
ㄸ	ㄱ RULE16	ㄴ R110	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110							ㅅ RULE17	ㅅ R100		
ㄹ	ㄱ RULE16	ㄴ R110	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110			ㅁ R190	ㅂ R190	ㅅ R190	ㅅ R190	ㅅ R190	ㅅ R190		
ㅁ	ㄱ RULE16	ㄴ R110		ㄷ R180	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ R180	ㄺ R180					ㅁ R180	ㅂ R180	ㅅ R180	ㅅ R180		
ㅂ	ㄱ RULE16	ㄴ R110	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110			ㅁ (D)	ㅁ R280	ㅂ R230	ㅅ RULE17	ㅅ R100			
ㅃ	ㄱ RULE16	ㄴ R110	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110							ㅅ RULE17	ㅅ R100		
ㅅ	ㄱ 어간+B	ㄱ R240	ㄴ 어간+B	ㄴ R250	ㄷ RULE08	ㄷ R100	ㄹ 어간+B	ㄹ R250	ㄺ RULE17	ㄺ R110	ㄻ	ㄻ R123	ㅁ 어간+B	ㅁ R240	ㅂ R230	ㅅ RULE17	ㅅ R100			
ㅆ	ㄱ RULE16	ㄴ R240	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110							ㅅ RULE17	ㅅ R100		
ㅇ	ㄱ (ABC) · (D)	ㄱ R140 · R151	ㄴ (ABC) · (D)	ㄴ R140 · R151	ㄷ (ABC) · (D)	ㄷ R140 · R151	ㄹ (1, ㅏ + AC) · (D)	ㄹ R170 · R140 · R151	ㄺ (ABC) · (D)	ㄺ R140 · R151	ㄻ	· R126 · R151	· R130 · R150	· R130 · R150	· R130 · R150	· R130 · R150	· R130 · R150	· R140 · R151	· R140 · R151	· R140 · R151
ㅈ	ㄱ 어간+B	ㄱ R240	ㄴ 어간+B	ㄴ R250	ㄷ RULE08	ㄷ R100	ㄹ 어간+B	ㄹ R250	ㄺ RULE17	ㄺ R110	ㄻ	ㄻ R120	ㅁ 어간+B	ㅁ R240	ㅂ R230	ㅅ RULE17	ㅅ R100			
ㅊ	ㄱ RULE16	ㄴ R240	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110							ㅅ RULE17	ㅅ R100		
ㅌ	ㄱ RULE16	ㄴ R240	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110							ㅅ RULE17	ㅅ R100		
ㅍ	ㄱ RULE16	ㄴ R240	ㄴ RULE17	ㄷ R100	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110							ㅅ RULE17	ㅅ R100		
ㅎ	ㄱ RULE16	ㄴ R240		ㄷ R121	ㄷ RULE08	ㄷ R100	ㄹ RULE08	ㄹ R100	ㄺ RULE17	ㄺ R110					ㅁ R121	ㅂ RULE17	ㅅ R100			
공백소	ㄱ	R110	ㄴ	R100	ㄷ	R100	ㄹ	R100	ㄺ	R110							ㅅ	R100		

표 5. 받침의 Rule base (3)
Table 5. The rule base by the end-consonants.

중성Ji 초성Cj	RULE19		RULE20		RULE21		RULE22		RULE23		RULE24		RULE25		RULE26		RULE27			
	入		ㅍ		ㅇ		ㅅ		ㅌ		ㅋ		ㅍ		ㅍ		ㅎ			
	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj	Ji	Cj
ㄱ	ㄱ	ㄱ	ㄱ	ㄱ	ㅇ	ㅍ	ㄱ	ㄱ	ㄱ	ㄱ	ㄱ	ㄱ	ㄱ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	ㄱ	
	RULE07	R090	RULE07	R090	(D)	R280	RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090			R120	
ㄲ	ㄱ	ㄲ	ㄱ	ㄲ			ㄱ	ㄲ	ㄱ	ㄲ	ㄱ	ㄲ	ㄱ	ㄲ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090			RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㄴ	ㄴ	ㄴ	ㄴ	ㄴ			ㄴ	ㄴ	ㄴ	ㄴ	ㅇ	ㄴ	ㄴ	ㅍ	ㄴ	ㅍ	ㄴ	ㅍ	ㄴ	
		R180		R180				R180		R180		R180		R180		R180			R180	R180
ㄷ	ㄱ	ㄷ	ㄱ	ㄷ	ㅇ	ㅍ	ㄱ	ㄷ	ㄱ	ㄷ	ㄱ	ㄷ	ㄱ	ㄷ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090	(D)	R280	RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090			R120	
ㄸ	ㄱ	ㄸ	ㄱ	ㄸ			ㄱ	ㄸ	ㄱ	ㄸ	ㄱ	ㄸ	ㄱ	ㄸ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090			RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㄹ	ㄱ	ㄹ	ㄱ	ㄹ	ㅇ	ㅍ	ㄱ	ㄹ	ㄱ	ㄹ	ㄱ	ㄹ	ㄱ	ㄹ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090		R190	RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㅁ	ㄴ	ㅁ	ㄴ	ㅁ			ㄴ	ㅁ	ㄴ	ㅁ	ㅇ	ㅁ	ㄴ	ㅁ	ㅁ	ㅁ	ㅁ	ㄴ	ㅁ	
		R180		R180				R180		R180		R180		R180		R180			R180	R180
ㅂ	ㄱ	ㅂ	ㄱ	ㅂ	ㅇ	ㅍ	ㄱ	ㅂ	ㄱ	ㅂ	ㄱ	ㅂ	ㄱ	ㅂ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090	(D)	R280	RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㅃ	ㄱ	ㅃ	ㄱ	ㅃ			ㄱ	ㅃ	ㄱ	ㅃ	ㄱ	ㅃ	ㄱ	ㅃ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090			RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㅅ	ㄱ	ㅅ	ㄱ	ㅅ	ㅇ	ㅍ	ㄱ	ㅅ	ㄱ	ㅅ	ㄱ	ㅅ	ㄱ	ㅅ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090	(D)	R280	RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090			R123	
ㅆ	ㄱ	ㅆ	ㄱ	ㅆ			ㄱ	ㅆ	ㄱ	ㅆ	ㄱ	ㅆ	ㄱ	ㅆ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090			RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㅇ	·	ㅅ	·	ㅆ			·	ㅅ	·	ㅆ	·	ㅋ	·	ㅌ	·	ㅍ	·	·	ㅇ	
	(ABC)	R130	(ABC)	R130			(ABC)	R130	(ABC)	R130	(ABC)	R130	(1+AC)	R170	(ABC)	R130	(BC)		R120	
	·	ㄱ	·	ㄱ			·	ㄱ	·	ㄱ	·	ㄱ	·	ㅌ	·	ㅍ	·	·		
	(D)	R150	(D)	R150			(D)	R150	(D)	R150	(D)	R150	(+1)	R130	(D)	R150				
													(D)	R150						
ㅈ	ㄱ	ㅈ	ㄱ	ㅈ	ㅇ	ㅍ	ㄱ	ㅈ	ㄱ	ㅈ	ㄱ	ㅈ	ㄱ	ㅈ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090	(D)	R280	RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090			R120	
ㅊ	ㄱ	ㅊ	ㄱ	ㅊ			ㄱ	ㅊ	ㄱ	ㅊ	ㄱ	ㅊ	ㄱ	ㅊ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090			RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㅌ	ㄱ	ㅌ	ㄱ	ㅌ			ㄱ	ㅌ	ㄱ	ㅌ	ㄱ	ㅌ	ㄱ	ㅌ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090			RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㅍ	ㄱ	ㅍ	ㄱ	ㅍ			ㄱ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	ㄱ	ㅍ	ㅍ	ㅍ	ㅍ	ㄱ	ㅍ	
	RULE07	R090	RULE07	R090			RULE07	R090	RULE07	R090	RULE01	R090	RULE07	R090	RULE17	R090				
ㅎ	·	ㅌ	ㄱ	ㅌ			·	ㅌ	·	ㅌ	·	ㅎ	·	ㅌ	ㅍ	ㅍ	ㅍ	ㄱ	ㅍ	
		R122	RULE07	R090			(1,4)	R121		R122	RULE01	R090		R122	RULE17	R090				
							(≠1,4)	R122												
공백소	ㄱ		ㄱ				ㄱ		ㄱ		ㄱ		ㄱ		ㅍ					
	R090		R090				R090		R090		R090		R090		R090					

즉, 12항 규정의 격음화, 초성 'ㅅ' 과 결합시의 'ㅆ' 발음 규정, 초성 'ㄴ' 과 결합시의 'ㄹ' 탈락 현상, 모음으로 시작되는 어미나 접미사와 결합시의 'ㅎ' 탈

락 현상, 15항의 모음 'ㅏ, ㅑ, ㅓ, ㅕ, ㅗ'로 시작되는 실질 형태소와 결합시의 음가 1개만 초성으로 옮겨져 발음되는 현상 등이다. 예를 들어, '맑음'이라는

표 6. [RULE06] 받침 'ㄹ' Table 6. The rule base for end-consonant 'ㄹ'.

Rule 정의	적용규칙	예
$J_1(\text{ㄹ}) / C_j(\text{ㄱ, ㅋ, ㆁ, ㄷ, ㅌ, ㅍ})$ $J_1(\dots) J_1(\text{ㄹ}); C_j(\dots) \{ \dots \} C_j(\text{ㄱ, ㅋ, ㆁ, ㄷ, ㅌ, ㅍ})$	[R120]	많고/만코 많다/만타
$J_1(\text{ㄹ}) / C_j(\text{ㅅ}): J_1(\dots) J_1(\text{ㄹ}); C_j(\dots) C_j(\text{ㅅ})$	[R123]	얇소/얇쏘
$J_1(\text{ㄹ}) / C_j(\text{ㄴ}): J_1(\dots) J_1(\text{ㄹ})$	[R125]	얇는/얇는
$J_1(\text{ㄹ}) / C_j(\text{ㅇ})(\text{BTBL, CTBL}): J_1(\dots) J_1(\text{ㄹ}); C_j(\dots) C_j(\text{ㄴ})$	[R126]	얇음/마음 얇이/마니
$J_1(\text{ㄹ}) / C_j(\text{ㅇ})(\text{V})\{\text{ㅌ, ㅊ, ㅌ, ㅊ, ㅌ, ㅊ}\}$ (DTBL): $J_1(\dots) J_1(\text{ㄹ}); C_j(\dots) C_j(\text{ㄴ})$	[R151]	

단어를 기존의 방법으로 음운 변동 처리를 하게 되면, 받침 'ㄹ'의 'ㅎ' 탈락 현상을 적용한 후 '만음'으로 만들고, 연음 법칙 현상을 재차 적용하여 '마음'으로 변환하게 된다. 기존 처리 방법의 문제점은 다음과 같다.

- ① 겹받침의 뒤음절에 나타나는 초성 'ㅇ'의 제한 조건인 조사, 어미, 접미사 조건이 검사되지 않는다. (받침 'ㄹ'의 경우는 어미, 접미사)
- ② 초성 'ㅇ과 연결될 때, 실질 형태소 유무의 조건 검사 불능. (예: '넋없다/너겹따', '값어치/가버치' 등)
- ③ 겹받침의 대표 음가 처리의 혼동(예: '흙/흑, '붉은/불근')
- ④ 어떤 입력 자료에 대해서 어떠한 음운 변동 현상이 몇회 적용될 것인가가 확정되지 않기 때문에 처리 시간이 길어진다.

본 연구의 표5에서 제시한 받침에 의한 27개의 rule base는 이상의 문제를 제거할 수 있다. 즉, '얇음'이라는 단어는 표6의 [R126] 규정을 적용하게 되어 어미, 접미사의 조건 처리가 가능하며, 겹받침의 정확한 음가 처리 및 변동 현상의 처리를 몇번 반복시켜야 하는 문제를 해결할 수 있다.

2. 음운 변동 처리 시스템

제시된 27개 rule base의 타당성을 검증하기 위하여 음운 변동 처리 시스템을 구성하였다. 이 시스템의 자료 입력은 키보드를 이용하며, 음운 변동 처리가 끝난 결과는 별도의 음가 데이터 베이스로 구성한 후, 그 결과를 검토한다. 다음 그림1은 음운 변동 처리 시스템의 전체 개요이다.

1) 전 처리부

입력되는 한국어 자료는 분석 작업이 요구되는 완전한 한국어 텍스트를 지양하고, 휴지가 발생하지 않는 한국어 단어나 어절을 대상으로 하였다. 음성 처

단계 1 : START

단계 2 : 한국어 단어의 입력 과정 /*단어 또는 어절 레벨 */

단계 3 : 2바이트 조합형 코드로 변환 /*완성형 코드의 경우 */

① 초, 중, 종성 코드의 분리 /*사전검색기의 key로 사용 */

② 음절수 추출 /*사전검색의 전처리에 사용 */

단계 4 : 초, 중, 종성, 음절수에 의한 단어 사전 검색

단계 5 : IF 음운 변동 처리

THEN GOTO 단계 6 /*변동 처리시 : 음가 정보 생성의 경우 */

ELSE GOTO 단계 9 /* 정보 출력시 : 단어 정보의 출력 경우 */

단계 6 : IF 음운 변동 tag=1, 음가 예외 tag=0

THEN CALL [음운 변동 처리기]

(조사, 어미, 접미사, 어간·한자어[tag, 음가 포인터 사용])

단계 7 : IF 음운 변동 tag=1, 음가 예외 tag=1

THEN CALL [예외 사전]

(음운 변동 예외 사전의 별도 처리)

단계 8 : 음가 정보 생성 /*음가 DB 생성 */

단계 9 : 단어 출력 및 개념 코드 출력

단계 10 : END.

그림 1. 음운 변동 처리 시스템

Fig. 1. The processing system of vocal sound change.

리시, 휴지 또는 띄어쓰기의 유무는 대단히 중요한 요인으로 자립 형태소와 의존 형태소는 부분적으로 결합되어 하나의 단어나 어절을 구성하게 된다. 단어는 뜻을 갖는 최소의 단위이며, 어절은 문장을 구성하는 의미를 갖는 최소의 단위로서 하나 이상의 단어가 모여 어절을 이루게 된다.

어절은 음성 인식시(음성 입력) 반드시 휴지가 발생하며, 음성 합성시(문자 입력)에는 띄어쓰기가 이루어지기 때문에, 본 연구에서는 처리하고자 하는 한국어 입력 자료의 단위를 다음과 같이 하였다.

- ① 체언(명사, 대명사, 수사)+관계언(격조사, 보조사, 접속 조사)
- ② 휴지가 발생하지 않는 단일 품사
- ③ 수식언(관형사, 부사)+관계언
- ④ 어간(용언)+어미+(관계언)
- ⑤ 어근(용언)+접미사+어미+(관계언)

입력된 자료는 2바이트 조합형 코드로 변환된 후, 음운 변동 rule base에 적용키 위하여 각 음절의 초성, 중성, 종성 단위로 분리 처리되며, 음절의 길이 만큼 음운 변동 처리부를 수행한다. 그리고, 마지막 음절의 중성이 홑받침이 아닌 경우에는 대표 음가 처리를 행한다.

2) 사전의 구성

음운 변동 처리를 위해서 단어 사전, 조사·어미·접

미사 사전, 예외 적용 사전, 음가 사전을 구성한다. 입력된 단어, 어절 또는 문장에 대한 조사, 어미, 접미사 및 어간의 분리는 자연어 처리 분야에서 많은 연구가 진행되고 있으나, 본 연구의 주된 관심이 아니기 때문에 조사, 어미, 접미사 사전을 별도로 구성하였으며, 어간은 단어 사전의 항목으로 포함시켜 처리하였다. 단어 사전은 입력된 단어 또는 어절의 음운 변동 처리시에 필요한 정보를 갖고 있다. 즉, J_1 를 포함한 앞의 음절들에 해당 rule base의 조건에 맞는가를 확인해주며, C_j 를 포함한 뒷음절들이 실질 형태소인가를 보여준다. 그리고, 해당 단어가 한자어인가의 유무와 음운 변동의 예외 처리 유무에 대한 tag를 갖는다. 한자어 tag는 J_n 에 해당하는 중성의 음절이 지정된 한자어 조건[R201]인 경우의 예외 처리와 RULE08의 [R260]의 적용시 참고된다. 음운 변동 처리를 위한 단어 사전의 레코드 항목은 다음과 같다.

- ① 단어 코드(체언, 수식언, 독립언, 용언 등의 단어)
- ② 품사 코드(형태소 정보)
- ③ 어간 코드(용언)
- ④ Rule 적용을 위한 한자어 tag
- ⑤ 음가 예외 tag
- ⑥ 조사, 어미, 접미사의 접속 코드
- ⑦ 음운 변동 유무 코드
- ⑧ 음가 정보 포인터
- ⑨ 개념 코드

조사, 어미, 접미사 사전은 초성 C_j 이후의 음절들 ($C_jV_jJ_j-C_nV_nJ_n$)이 각 rule의 조건에 맞는가를 확인하여 주며, 조사, 어미, 접미사 정보는 음운 변동 rule base의 적용시 수시로 참고된다. 예외 적용 사전은 단어 사전을 검색하여 예외 tag가 있으면 등록된 음가 정보를 추출한다. 음가 사전은 단어 사전과는 별도로 생성한다. 그리고, 음성 합성시에는 단어 사전의 검색후 음가 사전을 참조(음가 정보 포인터) 하게 하며, 음성 인식시에는 음가 사전의 검색 후 단어 사전을 참조할 수 있도록 해당 단어에 대한 위치 정보를 음가 사전에 포함시켰다. 다음은 음가 사전의 레코드 항목이다.

- ① 음가 코드
- ② 해당 블록(단어 사전) 내의 위치 정보
- ③ 음운 변동 rule 코드
- ④ 음절별 음성 DB 코드

음가 코드는 입력된 단어나 어절에 대해서 음운 변동 rule을 적용한 후 생성된다. 단어인 경우에는 단어 사전과 연결된 음가 데이터 베이스에 등록하며, 음성 합성을 위한 어절인 경우에는 별도의 텍스트

음가 데이터 베이스에 저장시킨다. 해당 블록 내의 위치 정보는 음가 정보에 대한 해당 단어 사전의 위치를 갖게 되며, 음성인식시 사용될 수 있다. 음운 변동 rule 코드는 음운 변동 현상의 적용 규정을 확인하기 위한 부가적인 정보이다. 음절별 음성 DB코드는 음성 합성에 필요한 음절별 음성 자료를 연결시켜 주는 것으로 음절수 만큼의 음성 코드를 갖게 된다.

3) 음운 변동 적용부

음운 변동 적용부에서는 전 처리부에서 분리된 단어 또는 어절 단위의 초성(C), 중성(V), 종성(J) 코드열을 중성별 음운 변동 rule base에 적용시켜, 음가 코드를 추출하는 작업이 이루어진다. 적용은 인접한 두개의 음절씩 ($C_jV_jJ_j, C_jV_jJ_j$) 차례로 실시하며, 마지막 음절 ($C_nV_nJ_n$)은 홀받침이 아닐 때 대표 음가 처리를 행한다. Rule의 처리시 조사, 어미, 접미사 조건이 발생하면, C_j 이후의 음절에 대해서 각각의 데이터 베이스를 참조하며, 실질 형태소 조건과 한자어 조건인 경우에는 단어 데이터 베이스를 참조한다. 예외 처리인 경우에는 음운 변동 rule의 적용 전에 단어 사전의 음가 예외 tag를 확인하여, 예외 사전 처리를 행하게 된다. 이상의 처리 과정을 통하여 생성된 음가열은 단어 데이터 베이스와 연결된 음가 데이터 베이스와 연결된 음가 데이터 베이스에 등록되며, 음성 합성을 위한 띄어쓰기 단위의 어절인 경우에는 text 음가 데이터 베이스에 저장한다. 그림2는 음운 변동 적용부의 처리 개요이다.

```

단계 1 : START
단계 2 : IF 음절수=1 /* 1음절의 중성 처리 */
      THEN GOTO 단계 4
단계 3 : 음절수 만큼 반복 처리
      IF 반복 횟수=음절수이면, /* 말음의 중성 처리 */
      THEN GOTO 단계 4
      IF 중성i=NOT /* 중성이 없음 */
      THEN GOTO 단계 3
      ELSE 음운 변동 rule 적용 /* 음운 변동 처리 */
단계 4 : 말음의 중성 대표음 처리
단계 5 : END.

```

그림 2. 음운 변동 적용부

Fig. 2. The application of vocal sound change.

IV. 실험 및 검토

Rule 적용시 문제가 된것은 조사, 어미, 접미사의 경우 유형 분류가 문헌에 따라 다양하기 때문에, 본

논문에서는 조사는 격조사와 보조사(특수조사)를, 어미는 선어말 어미와 어말 어미를 구별하지 않고 등록하였으며, 접미사는 사동, 피동의 보조 어간과 단어의 어간에 붙어 본래의 품사와 다른 품사로 바뀌는 것으로 규정하였다. 그리고, 어간의 조건 참조가 많이 일어나기 때문에 처리 속도를 높이기 위하여, 용언인 단어에 대해서는 어간 코드를 생성해 놓았다. 실질 형태소 조건은 뒤음절 이후 단어에 대한 조건이므로, 합성어의 경우 합성 가능한 모든 단어를 단어 사전에 등록하여야 한다. 한자어 조건은 순수한 우리말과 한자어를 구분하여 단어를 등록할 필요성이 있음을 보여 주고 있다. 따라서 본 연구에서는 단어 사전내에 순수 한글과 한자어를 구별할 수 있는 tag를 붙여 주는 방법을 채택하였다.

홀받침과 쌍받침에 적용되는 [R130]은 모음으로 시작되는 조사, 어미, 접미사가 아닌 즉, C,가 'o'로 된 일반적인 단어에도 적용된다. 즉, '국어/구거', '식욕/시곡' 등의 예에서 볼 수 있으며, 예외적인 단어는 예외 tag 및 예외 사전에 등록된다. RULE04의 [R240]은 어간과 어미의 결합시 발생하는 경우화 현상으로 몇몇의 예외(예: 한글, 감기, 반가움 등)를 제외한 일반적인 단어(예: 단가/단까, 문고리/문꼬리 등)에도 역시 적용되었다.

RULE05, RULE13, RULE20의 [R100], RULE05 [R140]의 C, (o) <ATBL, BTBL, CTBL> 조건과 역시 RULE05, RULE06 [R151]의 V, (1) <DTBL> 조건은 생략이 가능하다. RULE 24의 [R180]에서 C, (L) 조건, RULE27의 [R180]에서 C, (口) 조건은 그 예를 찾기 어렵기 때문에 역시 생략이 가능하다. RULE25의 [R170]은 V, (1) <CTBL> 조건이 생략 가능하며, 그리고 각 rule의 실질 형태소 조건 V, (1) <DTBL>과 C, (b) <DTBL>은 합성어에 해당하므로 합성이 가능한 단어간의 관계 설정에 관한 연구가 있어야 한다.

본 연구에서는 IBM-PC AT 호환기와 C언어를 이용하여, 국민학교 교과서의 표제어 약 2,300 단어를 수용한 단어 사전(그림1의 단계 4)과 음가 사전을 처리하는 실험을 하였다. 이 외에도 입력되는 모든 단어 및 어절에 대해서 음가 정보를 생성시켜 주는 음운 변동기가 별도로 제작 설치되었으며(그림1의 단계 5, 6, 7, 8의 확인), 이미 제작된 사전의 확장 및 사전 정보의 확인 및 부수적으로 제작된 각종 사전의 확장, 검색 역시 가능하게 하였다.

음운 변동 처리 시스템의 구현과 입력되는 모든 단어 및 어절 자료를 입력하는 실험을 통하여, 무의미한 규정의 적용과 여러회 규정을 적용하게 되는

기존의 음운 변동 적용 규칙의 단점을 지양한 음절의 받침별 27개 rule base의 타당성을 검증하였다. 그리고, 실험 결과의 검토 과정을 통해, 제시된 rule base가 각 rule별로 더욱 축소될 수 있음을 보였으며, 제시된 음운 변동 rule base는 음성 처리를 위한 음가 데이터 베이스의 생성 연구 등에 사용될 것으로 기대된다.

V. 결 론

본 연구에서는 음성 인식과 합성에 사용될 수 있는 음가 정보의 생성을 위하여, 표준어 규정의 표준 발음법을 이용한 음운 변동 rule base를 생성하였다. Rule base는 각종 음운 변동 규칙의 우선 순위에 따라서 적용할 수도 있으나 관계없는 음운 변동 규정을 여러번 적용하여야 하는 번거로움이 있어 처리 속도가 문제시 된다.

본 연구에서는 한국어 음운 변동의 거의 대부분이 앞음절의 받침과 뒤음절의 첫소리가 합쳐서 일어나므로, 받침을 중심으로 한 27개의 rule로 설계된 받침별 rule base를 제시하였다. 받침별 rule base는 실험 및 검토 과정에서 부분적으로 축약되었으며, 대량의 자료 입력 과정을 통하여 더욱 간략화될 수 있음을 확인하였다.

생성된 음가 데이터 베이스는 음성 인식시의 1차 검색 대상으로 사용될 수 있으며, 1차 검색이 끝나면 인식 내용을 확인하기 위하여 단어 사전을 검색하게 된다. 음성 합성시에는 입력된 텍스트의 자료로써 단어 사전을 검색하게 되며, 단어 사전내의 음가 정보 포인터를 이용하여 음가 데이터 베이스의 음가 정보를 추출해 내게 된다. 본 연구의 실험에서는 한국어의 음절 단위를 처리 대상으로 하였다. 음성 인식의 경우에는 음소 단위에 대한 연구가, 음성 합성의 경우에는 음운 변동 현상 이외에 음울, 음의 장단, 음절간 또는 어절간의 휴지 시간 등에 관한 한국어의 발음상 여러 특질을 대상으로 한 과학적인 연구가 있어야 할 것이다. 그리고, 음가 정보를 문자 정보로 복귀시키는 중의성(ambiguity) 문제가 앞으로 연구된다면, 본 논문에서 제시한 음가 사전의 연결을 해제시켜, 음성 인식시의 처리 속도를 증진시킬 수 있을 것이다.

參 考 文 獻

- [1] I.H. Witten, *Principles of Computer Speech*, Academic Press, 1985.
- [2] F. Fallside & W. A. Woods, *Computer Speech Processing*, Prentice Hall, 1985.

[3] J.D. Markel & A.H. Gray, Jr., *Linear Prediction of Speech*, Springer-verg, 1976.

[4] F. Itakura & Shuzo Saito, "On the optimum quantization parameters in the PARCOR speech synthesizer," *Conf. of IEEE Speech Comm. and Process.*, pp. 434-437, April 1972.

[5] Shuzo Saito & Kazuo Nakata, *Fundamentals of Speech Signal Processing*, Academic Press Japan Inc., 1985.

[6] G.Y. Lee & S.B. Rhee, "A study on korean voice recognition using the neural network," *JTC-CSCC'91*, pp. 415-420, July 1991.

[7] 김성락, 남시병, 이상범, "음절 단위를 이용한 한국어 연속 단어 인식," 대한전자공학회논문집-B, 제28권 제 1 호, Jan. 1991.

[8] 김길창 외, 한국어 처리 시스템 개발 환경의 연구, 한국과학기술원, April 1988.

[9] 허웅, 국어 음운학, 정음사, 1982.

[10] 남광우, 한국어의 발음 연구[I], 일조각, 1989.

[11] 김병수 외, "음절단위를 이용한 한국어 음성합성", 대한전자공학회 논문집, 제27권 제 1 호, pp 143-150, Jan. 1990.

[12] 이희승, 안병희, 한글 맞춤법 강의, 신구 문화사, 1991.

[13] 최종근, 새한글 맞춤법 용례집, 법조각, 1990.

著 者 紹 介



李 桂 英 (正會員)
 1954年 1月 3日生. 1980年 동국대학교 전자계산학과 졸업 (공학사). 1983年 동국대학교 대학원 전자계산학과(공학석사). 1988年 8月~단국대학교 대학원 전자공학과 박사과정. 1985年~현재 동국대학교 전자계산학과(경주) 조교수. 주관심분야는 자연어 처리, 음성 신호처리 등임.

李 相 範 (正會員) 第28卷 B編 第1號 參照
 현재 단국대학교 교수