

음성 코퍼스 구축을 위한 SiTEC 분절음·운율 레이블링 기준의 검토 및 제안

이숙향(원광대), 신지영(고려대), 김봉완(SiTEC), 이용주(원광대)

<차 례>

- | | |
|-------------------------------|----------------------|
| 1. 서론 | 2.3. 검토 및 제안 |
| 2. 분절음 레이블링 원칙 | 3. 운율레이블링 원칙 |
| 2.1. 레이블링 기호 | 3.1. 한국어의 운율구조 |
| 2.1.1. 음소 기호 | 3.2. K-ToBI 레이블링 기준안 |
| 2.1.2. 음성적 변이형 등을 위한
세부 기호 | 3.3. 검토 및 제안 |
| 2.2. 레이블링 기준안 | 4. 결론 |

<Abstract>

Some considerations on SiTEC segmental and prosodic labeling convention for Korean

Sook-Hyang Lee, Jiyoung Shin, Bong-Wan Kim, Yong-Ju Lee

This paper presents segmental labeling conventions proposed by SiTEC (Speech Information Technology Engineering Center) 2002 and proposes a new directions of a revision for a simpler version. The paper also reviews one of the prosody labelling conventions for Korean, K-ToBI convention(ver. 3.1) and proposes a couple of modifications and suggestions.

* Keywords: spoken language corpus, segmental labeling, prosodic labeling, accentual phrase, intonation phrase

1. 서 론

음성 코퍼스(spoken language corpus)는 대체로 다음과 같이 정의하고 된다[1][2].

(1) 음성 언어 코퍼스의 정의

컴퓨터가 읽을 수 있는 형태로 접근 가능하고, 내부에서 혹은 다른 기관의 과학자들에 의해 재사용될 수 있을 만큼 충분한 양의 주석과 문서화가 되어 있는 모든 발화 녹음의 모음¹⁾

위의 정의를 요약하면, 음성 코퍼스는 단순한 녹음 자료가 아니라 1) 컴퓨터가 읽을 수 있는 형태로 디지털화된 음성 자료이어야 하며, 2) 자료에 대한 정보가 소상히 밝혀져 있어 3) 내부적으로뿐만 아니라 외부적으로도 재사용될 수 있어야 한다.

결국, 위의 정의를 따른다면 단순히 존재하는 음성 녹음 자료가 곧바로 음성 코퍼스가 되는 것이 아니라, '녹음'을 통해 만들어진 음성 언어 자료를 코퍼스화하기 위해서는 우선 컴퓨터가 읽을 수 있는 형태로 디지털화해야 하며, 이렇게 구축된 자료에 대한 상세한 정보를 적어주어야 진정한 의미의 음성 코퍼스가 된다는 것을 알 수 있다[2].

전자화된 음성 자료에 상세한 주석을 다는 작업을 통칭하여 레이블링(labeling)이라고 하는데²⁾, 음성 언어는 문자 언어에 비하여 훨씬 많은 양의 정보를 가지고 있으므로 어떤 종류의 정보에 대한 주석을 얼마나 자세히 달 것인가를 결정하는 작업이 음성 코퍼스 구축에 선행되어야 한다. 물론, 내부뿐만 아니라 외부적으로도 재사용될 수 있어야만 하기 때문에 음성 코퍼스의 주석 작업에 대한 표준화 작업 또한 필수적이다.

음성 코퍼스 구축 및 활용에 있어 분절음 레이블링과 함께 운율 레이블링이 필수적인 작업 과정의 하나이기 때문에, 여러 해에 걸쳐 음성학자들과 음성공학자들이 공동으로 레이블링 기준안 제작과 함께 이에 대한 검토, 수정 보완 작업을 병행해 왔다.

그 한 예가 최근에 이루어진 원광대학교 SiTEC의 분절음, 운율 레이블링 작업이다. 이 논문에서 소개하고 있는 레이블링 원칙은 2002년 1월 SiTEC을 통해 제안되었고, 그 이후 다듬어진 것이다. SiTEC을 통해 제안된 분절음 레이블링 원칙은

1) ... any collection of speech recordings which accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by scientists in other organisations.[1]

2) 전자화된 문자 자료에 주석을 다는 일은 레이블링이라고 하지 않고 태깅(tagging)이라고 한다.

지난 2001년 6월부터 12월까지 SiTEC 관계자와 협력 사이트의 대표들 사이의 몇 차례 회의를 거쳐 다듬어지고 수정된 것임을 밝힌다. 본 논문에서 제안하고 있는 레이블링 원칙을 기초로 전국적 망을 가지고 있는 SiTEC 음성학 분야 7개 협력사이트³⁾가 2년여에 걸쳐 공동 작업으로 음성 코퍼스를 구축한 바 있다. 현재까지 남녀 500명 분량의 전화망 음성과 남녀 성우 각각 한 명이 4,392개의 문장을 발화한 총 8,784개의 녹음된 문장을 대상으로 레이블링하였다. 본 논문에서는 운율 레이블링의 근간이 된 K-ToBI (Korean Tone and Break Indices) 기준안을 간단히 소개하고 이 기준안에 제시되어 있지 않은 여러 다양한 레이블링 환경에 대한 몇 가지 규정과 보다 체계적이고 객관적인 레이블링을 위하여 운율 요소들에 대한 정량화의 시도를 함께 제안하였다.

2. 분절음 레이블링 원칙

2.1. 레이블링 기호

이 절에서는 SiTEC을 통해 제안되었던 분절음 레이블링 원칙을 살펴보려고 한다[3]. 우선 분절음에 대한 기호를 살펴보기로 하자.

2.1.1. 음소 기호

<표 1>과 <표 2>에는 한국어 자음과 모음의 음소를 위한 기호가 제시되어 있다. 표 1에 제시된 자음표에서 주의할 것은 한국어에서 하나의 음소 목록을 가진 유음에 두 개의 음소 기호, 'r'과 'l'이 제공되고 있다는 점이다. 이는 음성적 실현 형에서 많은 차이를 보이는 /r/의 두 변이음에 대한 정보를 제공하기 위한 것이다. <표 2>에 제시된 모음표에서 주의할 것은 제시된 음소 기호가 10모음 체계를 기준으로 하고 있다는 것이다.

비록, 표준발음법은 <표 2>에 제시하고 있는 것과 같이 한국어의 단모음을 10개로 인정하고 있지만[4], 모음 체계는 세대별, 지역별로 차이를 보인다. 표준어 화자들의 많은 부분을 차지하는 청장년층의 모음 체계는 7모음 체계에 가까우며, 경상 방언 화자들의 모음 체계는 6모음 체계이다. 따라서 몇 모음 체계를 기준으로 하고 있는가에 따라서 모음을 위한 음소 기호가 달라질 수 있다.

<표 3>은 7모음 체계와 6모음 체계를 위한 음소 기호와 함께 7모음의 모음 체계를 보이는 경우에 관찰되는 10개의 이중모음을 위한 기호를 보인 것이다.

3) 7개 협력사이트는 경북대(이은영), 고려대(신지영), 부천대(이순향), 서울대(이호영), 아주대(문승재), 연세대(이석재), 원광대(이숙향)임.

<표 1> 한국 자음 음소를 위한 기호

	양순 음	치경 음	치경 경구 개음	연구개 음	성문 음
폐쇄음 (파열 음)	ㅂ b ㅃ p ㅍ B	ㄷ d ㅌ t ㅍ D		ㄱ g ㅋ k ㄲ G	
마찰음		ㅅ s ㅆ S			ㅎ h
파찰음			ㅈ z ㅊ c ㅉ Z		
비음	ㅁ m	ㄴ n		ㅇ N	
유음		ㄹ l/r			

<표 2> 한국어 모음을 위한 기호

	모음의 음소 기호
단모음	ㅣ (i), ㅔ (e), ㅚ (E), ㅜ (u), ㅛ (o), ㅏ (a), ㅓ (U), ㅑ (v), ㅜ(O), ㅟ (Y)
이중모음	ㅟ(wi), ㅞ(we), ㅞ(wE), ㅟ(wv), ㅘ(wa), ㅙ(je), ㅙ(jE), ㅞ(ja), ㅟ(jv), ㅞ(jo), ㅟ(ju), ㅟ(xi)

<표 3> 7모음, 6모음 체계를 위한 음소 기호

7모음 체계	ㅣ (i), ㅔ/ㅚ(A), ㅜ(u), ㅛ(o), ㅏ(a), --(U), ㅑ(v)
6모음 체계	ㅣ (i), ㅔ/ㅚ(A), ㅜ(u), ㅛ(o), ㅏ(a), --/ㅑ(V)
이중모음 (7모음 체계)	ㅟ(wi), ㅞ/ㅙ/ㅞ(wA), ㅟ(wv), ㅘ(wa), ㅙ/ㅙ(jA), ㅞ(ja), ㅟ(jv), ㅞ(jo), ㅟ(ju), ㅟ(xi)

2.1.2. 음성적 변이형 등을 위한 세부 기호

앞 절에 보인 분절음 기호 외에 분절음 이하의 상세한 음성적 국면이나 음성적 변이형을 나타내기 위한 세부 기호들도 함께 제안되었는데 이들을 요약하면 (2)에 보인 바와 같다.

(2) 음성적 국면이나 변이형을 위한 세부 기호

가. H: 개방 혹은 기식 구간

- 1) 폐쇄음의 개방 + 기식 구간
- 2) 파찰음의 기식 구간(마찰 구간이 끝난 이후에 관찰되는 기식 구간)
- 3) 탄설음의 개방
- 4) 마찰음의 기식 구간

나. -: 음가탈락: 탈락된 음소와 - 기호를 선행 혹은 후행 음소에 붙여쓴다.

- 1) 유성음 사이의 /h/ 탈락: 후행 음소 앞에 붙여쓴다.

예) i a 사이의 h탈락: i h-a

- 2) 무성모음화: 선행 음소 뒤에 붙여쓴다.

예) s 뒤에서 무성모음으로 실현된 U: sU-

다. sil: 묵음 기호

라. V: 유성음화

예) 두 모음 사이에서 유성음으로 실현된 폐쇄음: 아기 a gV i sil

R: 공명음화 - 폐쇄음의 기호+R

완전 폐쇄가 일어나지 않은 폐쇄 구간의(b, d, g, r, z)의 공명음화:

예) 할아버지 h a rR a b v z F i sil

F: 파찰음의 마찰 구간

Q: 무성음화

유성음의 무성음 실현: 무성음으로 실현된 탄설음

예) 하더라 h a d v rQ a sil

2.2. 레이블링 기준안

SiTEC에서 제안한 분절음 단위 레이블링의 원칙을 요약하면 (3)과 (4)에 보인 바와 같다.

(3) 레이블링 기본 원칙

- 1) 모든 레이블은 그 끝점을 기준으로 표시한다.
- 2) 발화에서 관찰되는 묵음 구간은 묵음을 표시하는 기호 sil을 달아 표시한다.
 - 2-1) 첫 분절음의 시작점이 바로 묵음의 끝점이므로 묵음 뒤에 나오는 음소의 시작점에 묵음 기호인 sil을 달아준다.
 - 2-2) 발화의 중간에 놓인 묵음 구간은 묵음 구간의 끝점, 즉 다음 음소의 시작점에 sil을 달아 준다.
 - 2-3) 첫 분절음의 시작점은 묵음의 끝점이므로 첫 음소의 시작점에 sil을 달아 준다.
- 3) 음성 총위에는 소리나는 대로 레이블을 달아 준다.
- 4) 첫 음소, 혹은 묵음 뒤에 오는 음소가 폐쇄음, 파찰음과 같이 폐쇄 구간을 가진 경우는 폐쇄기간을 따로 잡아주지 않고, <묵음기호 + 해당 음소기호>를 적어서 묵음 뒤에 오는 첫 음소가 폐쇄 구간을 갖는다는 것을 알려준다.

예) 감 silg a m
- 5) 발화의 마지막 혹은 묵음 앞에 폐쇄음이 오는 경우는 폐쇄 구간을 잡아주지 않고 <폐쇄음에 선행된 음소(모음) + 폐쇄음의 기호>를 달아준다.

예) 악 sil ag
- 6) 마찰음이나 기식성이 큰 자음 뒤에 모음이 안 보이는 경우, 즉 무성모음화가 관찰되는 경우에는 <선행 자음의 음소기호 + 없어진 모음의 기호+ ->를 달아준다.

예) 같습니다 silg a SU- m n i d a
- 7) h 탈락이 있거나 h의 구간이 애매한 경우는 후행하는 모음 앞에 <h->를 달

아준다.

예) 가학적 silg a h-a k Z vg

- 8) 하나의 AP 내에 조음 위치가 같은 폐쇄음과 마찰음의 연쇄는 허용되지 않는다. 즉, d S의 연쇄는 불가능한 것으로 간주한다.

예) 젓소 silz F v S o

- 9) 모음의 시작은 F1과 F2가 관찰되기 시작하는 곳과 끝나는 곳으로 규정하 되, 묵음 앞에 위치한 모음의 끝은 파형에서 모음의 파형(여러 성분들이 복합되어 나타나는)이 관찰되는 곳의 끝을 잡는다.
- 10) 비음이 묵음에 선행하는 경우, 그 끝은 파형에서 규칙적인 형태가 관찰되는 곳까지로 규정한다.
- 11) 폐쇄음은 폐쇄 구간과 나머지 구간(개방 구간과 기식 구간)을 나누어 표시한다. 이때 폐쇄 구간은 표1에 제시된 해당 음소 기호로 표시하고 나머지 구간, 즉 개방과 기식 구간은 H로 표시한다.
- 12) 마찰음에 기식 구간이 존재하는 경우 이를 나눠준다. 이때 마찰 구간은 < 표 1>에 제시된 해당 음소 기호로 표시하고 기식 구간은 H로 표시한다.
- 13) 파찰음은 폐쇄 구간, 마찰 구간, 기식 구간을 나눠주고, 폐쇄 구간은 해당 분절음의 기호로, 마찰 구간은 F로, 기식 구간은 H로 표시한다.

예) 자동차 z F H a d o N c F H a

- 14) 유음의 변이음은 서로 다른 분절음 기호로 나타낸다.

탄설음: r

설측음: l

- 15) 치경 마찰음, /s, S/의 변이음을 위한 기호는 따로 두지 않는다.
- 16) 유성음화는 해당 음소 뒤에 'V'를 붙인다.

예) 유성음으로 실현된 평폐쇄음

bV, dV, gV, zV, hV 등

- 17) 폐쇄 구간의 완전 폐쇄가 이루어지지 않은 음소는 해당 음소 뒤에 'R'을 붙인다.

예) rR, bR, dR, gR, zR

- 18) 완전 폐쇄가 이루어지지 않은 파찰음: 마찰 구간을 표시하는 F를 함께 적어 준다.

예) 아버지 a b v zF i sil

(4) 분절음별 레이블링 세부원칙

가. 모음 관련 원칙

모음의 시작은 해당 모음의 F1과 F2가 관찰되기 시작하는 곳으로, 모음의 끝은 파형에서 모음의 파형(여러 성분들이 복합되어 나타나는)이 관찰되는 곳의 끝을

잡는다.

원칙-1) 단모음: 이웃 자음과 구별되는 포만트 지속시간의 시작점과 끝점을 경계로 분할한다.

원칙-2) 이중모음- 반모음과 단모음 성분을 구별하지 않는다.

원칙-3) 단모음+단모음: 선행 단모음에서 후행 단모음으로의 포만트 전이 (formant transition)의 중간을 경계로 분할한다.

원칙-4) 단모음+이중모음: 이중모음의 반모음 성분의 시작점을 경계로 선행 단모음과 후행 이중모음을 분할한다.

원칙-5) 이중모음+단모음: 단모음 + 단모음의 원칙 적용.

원칙-6) 이중모음+이중모음: 단모음 + 이중모음의 원칙 적용

원칙-7) 무성음화된 모음: /CV/ 환경에서 /V/ 가 무성음화되면 자음과 모음의 기호를 같이 쓰고, 모음 기호뒤에 ‘.’ 를 붙여준다.

원칙-8) 비음화된 모음(후행하는 비음의 성분이 뚜렷하게 나타나지 않는 경우): “모음 + 비음”의 연속체에서 그 경계 분할이 모호한 경우, 모음과 비음의 경계를 별도로 분할하지 않고 모음 기호와 해당 비음 기호를 함께 붙여쓴다.

나. 자음 관련 원칙

원칙 9) 어두 폐쇄음

- 1) IP초 폐쇄음: IP초 폐쇄음의 폐쇄 구간은 IP초 묵음 구간과 구별할 수 없으므로, 이를 별도로 구분하지 않고 폐쇄 구간의 끝점, 즉 기식 구간의 시작점에 묵음을 나타내는 음소 sil뒤에 폐쇄음의 기호를 붙여쓴다.

예) ‘가’ silg H a

- 2) IP말 폐쇄음(불파음): 어절말 폐쇄음은 폐쇄만 일어날 뿐 파열이나 기식 음 구간이 없다. 또한 어절말 폐쇄음의 폐쇄 구간은 IP말 묵음 구간과 구별하기 어려우므로, 폐쇄 구간을 따로 지정해 주지 않고, 앞 음소의 끝점에 폐쇄음 기호를 붙여 써준다.

예) ‘각’ silg H ag

원칙-10) 어중 폐쇄음

- 1) 음절 초 폐쇄음: 폐쇄음의 개방이 뚜렷이 관찰되어 후행 모음 앞에 약간의 기식 구간이 관찰되면 이를 구분하여 표기하되, 뚜렷한 개방이 관찰되지 않는 경우에는 기식 구간을 지정해 주지 않는다.

- 1-1) 음절 초 폐쇄음이 선행 유성음이나 후행 유성음의 영향을 받아 유성음화된 경우 ‘V’를 붙여쓴다.(유성음화 기호를 붙이는 경우는 전체 폐쇄 구간의 50% 이상에서 규칙적인 파형이 관찰될 때로 규정

한다.)

- 1-2) 음절 초 폐쇄음이 완전 폐쇄를 보이지 않고 포먼트 구조를 갖는 접근음으로 실현되면 음소기호 뒤에 ‘R’ 을 붙여쓴다.
- 2) 음절 말 폐쇄음: 폐쇄음의 폐쇄 구간을 별도로 표시하지 않는다. 선행하는 분절음 뒤에 음소 기호를 달아주어 모음의 끝점에 붙과된 폐쇄음이 후행함을 표시한다.

예) 각 silg H ag

- 3) 음절 말 폐쇄음 + 음절 초 폐쇄음: 음절 말 폐쇄음과 음절 초 폐쇄음이 연속하는 경우, 선행하는 음절 말 폐쇄음의 폐쇄 구간과 후행하는 음절 초 폐쇄음의 폐쇄 구간이 중첩되어 경계 분할이 없다. 따라서 임의로 음절 말 폐쇄음의 폐쇄 구간을 절반 잡아주고 나머지를 후행하는 폐쇄음의 폐쇄 구간으로 간주한다.

원칙-11) IP초 파찰음: IP초 파찰음의 폐쇄 구간은 IP초 묵음 구간과 구별할 수 없으므로, 이를 별도로 구분하지 않고 묵음이 끝나는 곳, 즉 첫 음소가 시작되는 곳에 묵음 기호와 음소 기호를 붙여써 준다.

예) 자동차 silz F H d o N c F H a

원칙-12) 어중 음절 초 파찰음: 어중 음절 초 폐쇄음의 폐쇄 구간과 마찰 구간이 명확히 구별되어 실현되는 경우, 이를 별도로 구별하여 표기한다.

- 1) 어중 음절 초에 나타나는 파찰음이 유성음 사이에서 유성음화되는 경우 해당 구간에 ‘V’를 붙여쓴다.(50% 이상에서 규칙적인 파형이 관찰된 경우)
- 2) 어중 음절 초 파찰음이 폐쇄 구간은 나타나지 않고 마찰잡음구간만으로 실현된 경우 폐쇄 구간이 없는 파찰음임을 표기하기 위해 파찰음의 폐쇄 구간기호에 마찰 구간 기호 ‘F’를 바로 붙여쓴다.

예) 가지 silg H a zF i

- 3) 음절 말 폐쇄음 + 음절 초 파찰음: 음절 말 폐쇄음과 음절 초 파찰음이 연속해서 나타나는 경우, 두 자음의 폐쇄 구간의 경계를 구별하기 어렵다. 폐쇄음의 연쇄와 마찬가지로 앞 폐쇄음의 폐쇄 구간을 반으로 나누어 파찰음의 폐쇄 구간으로 간주한다.

원칙 -13) 치경 마찰음

- 1) 치경 마찰음은 기식성을 갖는 경우에는 마찰 구간과 기식 구간을 구분하여 표시한다.

예) 사 sil s H a

- 2) 어중 유성음 사이에 놓인 치경 마찰음의 유성음화는 여타의 분절음과 마찬가지로 분절음 기호 뒤에 ‘V’를 달아준다. (마찰 구간의 50% 이상에서 규칙적인 파형이 관찰되는 경우)

원칙 - 14) 성문 마찰음

- 1) AP 이상의 운율 단위 초에 놓인 성문 마찰음은 마찰 성분의 시작점에서 모음의 시작점까지를 그 경계로 표시한다.
- 2) 유성음화된 성문마찰음: 유성음사이에 나타나는 성문마찰음은 많은 경우에 앞뒤 유성음의 영향을 받아 유성음화 된다. 유성음화된 성문마찰음에는 'V'를 붙여쓴다.
- 3) 유성음 사이에서 탈락이 일어난 경우: 후행하는 모음 음소 앞에 h-를 달아준다.
- 4) 비음, 유음과 모음 사이에 위치한 성문 마찰음: 유음과 비음이 후행 음절의 초성으로 실현되고, 성문 마찰음은 후행하는 모음에 기식 성분으로 없히는 형태로 실현된다. 후행하는 모음 앞에 기식음을 알리는 H를 함께 달아준다.

예) 실현 sil s i r Hjv n

원칙-15) 두 비음의 연속

- 1) 비음의 전 구간을 반으로 나누어 각각 레이블링한다.
- 2) 비음의 연속에서 앞 비음의 구간이 확실히 관찰되는 경우(폐쇄의 개방으로 인하여 스펙트로그램에서 수직의 스파이크가 확실히 관찰되는 경우)에는 임의로 반을 나누지 말고 수직의 스파이크가 시작되는 곳을 앞 음소의 끝점으로 간주하여 레이블링한다.
- 3) 후행하는 비음에 의해 모음이 비음화되고 그 모음 뒤에 비음성분이 뚜렷이 나타나지 않는 경우: 원칙-8)을 참조.

원칙 - 16) 탄설음

- 1) IP초: IP초에 나타나는 탄설음의 폐쇄 구간과 묵음 구간의 경계 분할이 어려우므로, 음절 초에 나타나는 무성 탄설음의 폐쇄 구간은 별도로 분할하지 않고 대신 묵음기호 뒤에 r을 붙여쓴다. 유성음으로 실현된 경우에는 유성음 구간을 탄설음의 폐쇄 구간으로 인정한다.

예) 라면 silrQ H a m jv n (무성음 실현)

라면 sil r H a m jv n (유성음 실현)

2) 어중 음절 초 탄설음

- 1) 유성음으로 실현: r H
- 2) 무성음으로 실현: rQ H

원칙 - 17) 음절 말 유음

- 1) 설측음으로 실현되며 'l'로 표기한다.
- 2) '빨리'와 같이 두 유음이 연결되는 경우는 유음 전체의 길이를 반으로 나누어 레이블링한다.
- 3) 모음에 이어지는 음절 말 유음의 구간이 앞에 오는 모음의 구간과 잘 나

뉘지지 않을 때는 <모음의 음소 + l>을 써 준다.

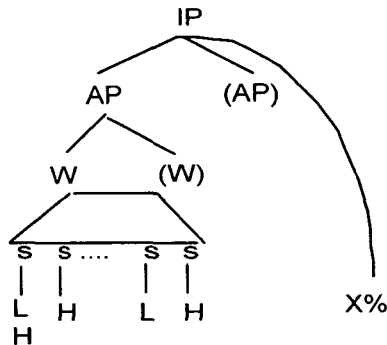
2.3. 검토 및 제안

지금까지 우리는 SiTEC에서 제안한 분절음 레이블링 원칙을 상세히 검토해 보았다. 앞에서 보았듯이 SiTEC의 분절음 레이블링 원칙은 분절음의 특성에 따라서 분절음보다 작은 단위로 레이블링 할 것을 요구하고 있다. 경우에 따라서는 한 분절음이 여러 가지 음성적 국면들을 가지고 있으므로 이를 세분하여 표시하는 것이 유용하기도 하지만, 경우에 따라서는 모든 음성적 국면이 아니라 분절음 단위로만 레이블링을 수행하고자 하기도 한다. 앞으로의 연구에서는 음성적 국면들을 적절히 통합하여 분절음 단위로 묶는 방법을 강구해야 할 것으로 보인다.

3. 운율 레이블링 원칙

3.1. 한국어의 운율 구조

<그림 1>에서와 같이 한국어 발화 문장의 운율 구조는 계층 구조로 이루어져 있으며 하단의 음절층에서 시작하여, 단어/어절층, 악센트구(Accentual Phrase)층, 그리고 최상단에 억양구(Intonation Phrase)층으로 구성된다. 악센트구는 하나 이상의 어절로 이루어지며 억양구는 악센트구의 상위층으로서 하나 이상의 악센트구로 이루어진다.



<그림 1> 서울 방언의 운율 구조. (IP(Intonation Phrase): 억양구, AP(Accentual Phrase): 강세구, W(Prosodic Word): 운율 단어, s(Syllable): 음절, X%: 억양구 경계 성조.

억양구는 구 마지막에 경계 성조로 구분되어지고 구말 장음화 현상(phrase final

lengthening)을 수반하며 그 하위층에 속하는 악센트 구간의 downstep이 일어나는 최대 구간이 된다. 반면에 악센트구는 기본형으로 [L +H ... L+ Ha] 성조 패턴을 가지며 구말 장음화 현상은 보이지 않는다[12][6]. 그리고 경음화와 모음 장음화(전남 방언), 그리고 유성음화가 적용되는 최대 구간이기도 하다[12]. 일반적으로 악센트구 초 [L +H]는 L이 첫 음절에, +H가 둘째 음절에 없으며 악센트구 말 [L+ Ha]는 구말 끝에서 둘째 음절에 L+가 그리고 마지막 음절에 Ha가 없힌다. 또한 기본형 [L +H ... L+ Ha]는 여러 다양한 요인에 의하여 다양한 형태로 실현되어 나타나기도 한다. 첫째, 악센트구의 시작자음이 강자음 (/s/)와 /h/를 포함한 유기음과 경음)이면 [H +H L+ Ha]의 성조 유형이 되며 둘째, 악센트구의 음절수가 4개보다 적을 때는 악센트구 초 둘째 성조 +H 또는 악센트구 말 L+가 실현되지 않거나 두 성조 모두가 미실현되기도 한다. 셋째, 악센트구 말의 Ha 대신 La로 끝나는 경우도 간혹 관찰된다[5][13].

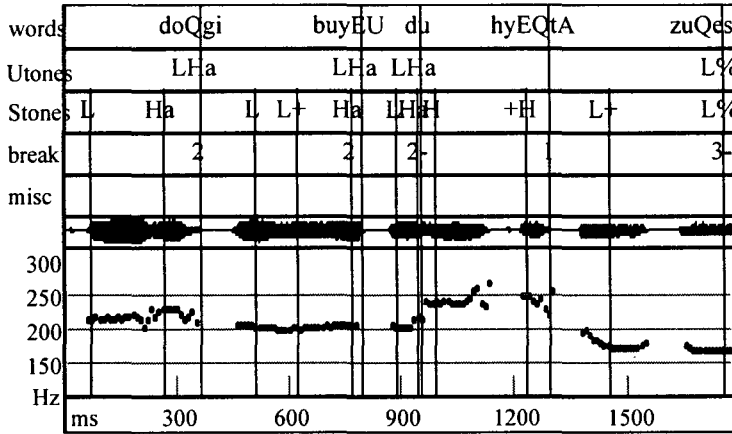
3.2. K-ToBI 레이블링 기준안

한국어 운율 구조 연구는 그 동안 음성, 음운학자들에 의해, 여러 가지 접근 방법으로 이루어져 왔다[8][9][11][14]. Beckman과 Jun은 이러한 연구 결과들을 Pierrehumbert의 운율 이론적 토대[16]에 접목시켜 한국어 운율 구조 기술 체계의 하나인 K-ToBI 기술 체계를 고안하였다[10]. 그 후 1996년에 1차 수정을 거치고(2차 버전), 2000년도에 다시 한 번 수정을 거듭하여 3차 버전이 나왔다[13].

K-ToBI 레이블링 기준안은 그림 2와 같이 5 개의 기술 층을 둔다⁴⁾.

1. 단어층 a word tier
2. 음운[기저]성조층 a phonological tone tier
3. 음성[표면]성조층 a phonetic tone tier
4. 연결층 a break-index tier
5. 기타층 a miscellaneous tier

4) 기준안 3차 버전이 과연 한국어의 운율구조를 기술하기에 적절한 체계인지를 보기 위한 연구의 일환으로 수행한 레이블링 실험에서 실험에 참여한 20명의 레이블러간의 일치도가 비교적 높게 나타났다[14].



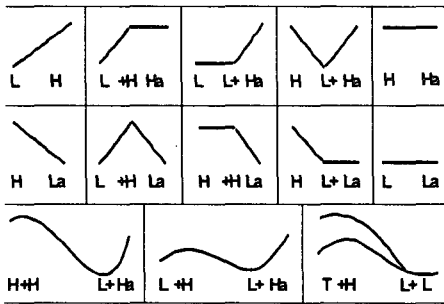
<그림 2> K-ToBI 레이블링 기준안의 운율 레이블링의 한 예.

단어층에서는 한글 맞춤법에 따라 띄어쓰면 일단 하나의 단어로 가정한다. 성조층은 악센트구와 억양구의 존재 또는 경계를 나타내는 기저성조층과 실제 다양한 형태로 나타나는 두 개 구의 억양패턴을 기술하는 표면성조층으로 나뉜다.

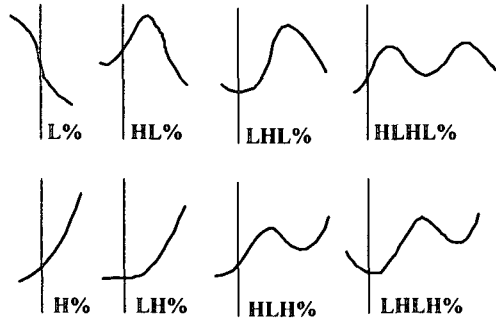
K-ToBI 레이블링 기준안에서 가정하고 있는 한국어의 악센트구는 14개의 유형이며(그림 3), 억양구는 8개의 유형(그림 4)으로서 이들 중 하나가 표면성조층에 각각 억양구와 악센트구의 표면 유형으로 기술된다.

연접층은 발화 문장의 단어와 단어간, 그리고 문장의 마지막 단어와 문장 끝에 오는 휴지간의 연접의 정도를 나타낸다. K-ToBI 레이블링 기준안은 '0'에서 '3'까지 4 단계의 연접 또는 끊김의 정도를 가정한다. 판단은 레이블러의 청각적 판단에 의하며 '0'은 음성적으로 cliticization이 일어나거나, 일음절 불완전 명사에 해당하며, '1'은 단어간의 연접 정도, '2'는 pause의 강한(주관적) 느낌이 없는 약한 구 경계에 해당하는 끊김으로서 전형적으로 악센트구 경계에서 느끼는 끊김 정도를 말한다. '3'은 pause가 실제 waveform이나 spectrogram 상에서 관찰되는 무음 구간이든 또는 무음 구간은 없지만 구말 장음화 현상에 의해 느껴지는 것이든 간에 pause의 강한 느낌을 주는 강한 구경계에서 느껴지는 끊김의 정도를 말한다. 전형적으로 억양구의 경계 성조와 함께 느껴지는 끊김이 이에 해당한다. 실제 성조층과 연접층간의 불일치가 일어나는 경우를 위해 불일치를 나타내는 'm'을 더해준다(2m, 3m). 연접의 정도 판단에 확신이 서지 않는 경우에는 '-'를 더해준다(1-, 2-, 3-).

기타층에서는 성조와 연접 정도 외에 레이블러의 판단에 따라 필요하다면 뭐든 기술해줄 수 있다. 예를 들어, 화자의 문장 발화 중에 기침소리가 났다든지, 말을 더듬는다든지 또는 말하는 중간에 다른 화자가 끼어 들었다든지 등의 것을 기술해줄 수 있다.



<그림 3> 한국어 악센트구의 14가지 유형의 표면형 성조의 도식적 모양.



<그림 4> 한국어의 다양한 억양구 유형.

3.3. 검토 및 제안

K-ToBI 레이블링 기준안 자체에 대한 검토 및 제안과 SiTEC 레이블링 공동작업 시 협력 사이트와의 논의 결과 협의되었던 수정, 보완사항을 제시한다. 그리고 보다 객관적인 기술을 위해 운율 요소들의 음성학적 특성에 대한 몇 가지 정량화 시도를 제안한다.

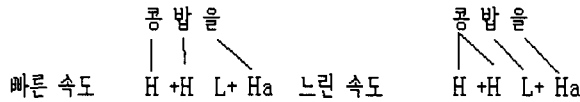
1. K-ToBI 레이블링 기준안 3.1버전에서 억양구는 기저성조층이나 표면성조층 모두에서 실제 표면형 8개 유형 중 하나를 기술해주고 있다. 반면에, 악센트구는 기저성조층에서는 실제 표면형을 기술하지 않고 단순히 기저형만을 기술하고 (LHa), 표면성조층에서 14개의 다양한 성조 중 하나를 기술해주고 있다. 따라서, 악센트구의 기술과 일관되면서 또한 하나이었던 성조층을 기준안 3차 버전에서 들로 나눈 취지에 합당하게, 억양구도 기저성조층에서 8개의 억양구 경계 성조를 대표하는 또는 커버하는 성조를 하나 선정하여 그것으로 억양구의 경계 성조를 나타내는 것이 바람직하다고 생각한다. 가장 빈도수가 높은 L%를 기저 억양구 경계 성조로 제안한다.
2. 실제로 SiTEC 공동 레이블링 작업에서는 기저성조층은 기술하지 않기로 하였다. 이 층은 언어학적으로는 중요한 의미를 가질 수 있으나, 표면성조층으로부터 예측 가능하므로 필요한 경우, 이를 표면성조층에서도 출해내는 것으로 하여 기술에서 제외하였다
3. 에너지가 약하여 억양곡선이 억양구 마지막까지 이어지지 못하고 끊기는 경우가 종종 발생하는데 이 경우, 레이블러의 청각적 판단에 의거하

여 경계 성조 유형을 기술하되, 기호는 보통의 억양구 경계 성조 기술과 마찬가지로 단어층의 경계와 일치하는 곳에 둔다.

4. 또한 피치분석 프로그램의 오류로 인해 억양구 경계에서 억양곡선 끝부분에 소음이 생기는 경우가 종종 발생하는데, 이 경우 또한 운율 판단은 레이블러의 청각적 판단에 의해 기술하며 소음 시작 바로 앞에 성조 기호를 둔다.
5. 모음 무성음화에 의해 악센트구 중간에 억양곡선이 끊기는 경우가 발생하는 경우, 특히 유기음과 고모음의 결합에서 빈번히 발생하는데 이 경우 또한 운율 판단은 레이블러의 청각적 판단에 의해 기술하며 모음 중간 지점에 성조 기호를 둔다.
6. 보다 객관적이고 체계적인 기술과 자동 레이블링을 위해서는 운율적 요소들에 대한 정량화가 필요하다. 예를 들어, 악센트구의 유형별 빈도수나 음절수와 유형간의 관계 등에 관한 여러 연구가 있긴 하나, 화자 수나 출신 성분, 발화 스타일, 발화 속도, 발화 환경, 그리고 자료의 크기에서 볼 때, 아직 미약하다. 다양한 그리고 동시에 대규모의 음성코퍼스 구축과 분석을 필요로 한다. 운율 레이블링 작업 시 많은 어려움을 호소하는 경우가 +H, Ha, H%간의 구분이다. 또한 L+, La, L% 간의 구분 또한 어려움이 많다. 이와 관련한 연구 결과들을 보면, 서울 방언에서 Ha 유형 악센트구 내의 내림(+H L+)의 경사도는 음절수에 따라 경사도가 달라지는 것으로 나타났다(4음절: -0.08 Hz/ms, 6음절: -0.20 Hz/ms)[12]. 그러나 악센트구 경계에 나타나는 내림 (La H)은 악센트구 내에 나타나는 내림보다 경사가 급하게 나타났다 ($-0.5 \sim -0.7$ Hz/ms). 같은 연구에서 서울 방언에서와는 달리 {L +H La} 또는 {H +H La} 패턴을 보이는 전남 방언 악센트구에서는 음절수가 내림(+H La)의 경사도에 큰 영향을 미치지 않는 것으로 나타났다 ($-0.15 \sim -0.18$ Hz/ms). 다른 연구 결과를 보면, 서울 방언의 La 유형 악센트구의 내림의 경사도는 전남 방언의 내림보다 일반적으로 경사가 급하게 나타났으며 서울 방언에서 Ha 유형의 악센트구 내 내림은 음절수가 많아질수록 경사가 급하게 나타난 반면 La 유형의 악센트구 내 내림은 음절수가 많아질수록 경사가 완만하게 나타났다[7]. 이 외에 악센트구의 오름(L +H, L+Ha)의 경사도에 대한 연구 결과들도 있다.

또한 발화속도와 운율실현과의 관계에 대한 연구로는 악센트구의 성조 실현은 음절수의 영향을 받는다고보다는 악센트구의 지속시간의 영

향을 받는 결과를 보여 주는 연구 결과가 있었다[7]. 예를 들어, 보통 발화속도에서 첫째 음절과 둘째 음절에 실현되는 고음 [H +H]가 느린 발화속도에서는 첫째 음절에만 실현되는 경우들이 종종 관찰되었다. 예를 들어 /콩밥을/을 빠른 속도나 보통 속도로 읽었을 때에는 /콩/과 /밥/에 각각 고음 H와 +H가 실현되었으나 느린 속도에서는 /콩/에만 고음이 실현되었다. /밥을/은 [L+ Ha]로 실현되었다. 즉, 빠른 속도에서는 시간이 부족하여 구말 L+가 실현되지 못함으로써 {H Ha}로 실현되었으며 느린 속도에서는 첫째 음절 /콩/이 충분히 길어서 악센트구조 H와 +H 모두 첫째 음절에 연결되고 오히려 둘째 음절 /밥/은 L+에 연결되어 저음으로 실현되는 것으로 볼 수 있겠다.



위와 같은 운율 요소들에 관한 음성학적 특성들에 대한 정량화 또는 운율 요소들에 미치는 여러 다양한 요인들에 대한 연구가 본격적으로 시작되면 보다 객관적이고 따라서 용이한 기술이 이루어질 것이며 또한 보다 정확한 운율 자동 레이블링화에 한 걸음 더 다가가게 될 것이다.

4. 결 론

본 논문은 최근 음성공학과 음성학 분야간의 학제간 연구로서 이루어진 분절음, 운율 레이블링 공동 작업에서 사용한 SiTEC 분절음, 운율 레이블링에 대하여 상세히 살펴보았다. 이를 위하여 레이블링 기호는 물론 원칙에 대한 상세한 소개를 아울렀다. 특히 운율 레이블링의 근간이 되고 있는 K-ToBI 레이블링 기준안을 소개하기도 하였다. 아울러 레이블링의 실제 작업에서 발견된 문제점들을 바탕으로 수정, 보완 사항을 제시하였다. 그리고 보다 체계적이고 객관적인 분절음 운율 레이블링을 위한 음성학적 특성들에 대한 정량화 문제에 대한 사항들도 논의해 보았다.

참 고 문 헌

- [1] D. Gibbon, R. Moore, R. Winski, *Handbook of Standards and Resources for Spoken Language System*, New York: Mouton de Gruyter, 1997.
- [2] 신지영, “음성 언어 정보 처리의 방법과 실제”, 민족문화, 35호, pp.111-129, 2001.
- [3] 신지영, “분절음 레이블링 기준”, 제1회 SiTEC 아카데미 교재.
- [4] 문화관광부, “표준 발음법 규정”, <http://www.korean.or.kr>
- [5] 이숙향, “한국어 ToBI 기준안 및 Segmentation 기준안”, 미출판 원고, 한국전자통신연구소 용역 보고서, 1997.
- [6] 이숙향, “한국어 운율 구조와 관련한 모음 및 음절 길이”, *말소리*, 35-36호, pp.13-24., 1998.
- [7] 이숙향, “한국어 운율구조 기술 체계에 대한 연구: K-ToBI 기술 체계를 중심으로”, *언어학*, 10권, 2호, pp.1-18, 2002.
- [8] 이현복, “서울말의 리듬과 억양”, *어학연구*, 10권, 2호, pp.15-25, 1974.
- [9] 이호영, *국어음성학*, 태학사, 1996.
- [10] M. Beckman, S-A. Jun, “K-ToBI annotation convention”, Unpublished manuscript. Ohio State Universti and UCLA, 1995
- [11] S-A Jun, “The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure”, *Doctoral dissertation*, Ohio State University, 1993.
- [12] S-A. Jun, “The accentual phrase in the Korean prosodic hierarchy”, *Phonology*, Vol. 15, pp.189-226, 1998.
- [13] S-A. Jun, *K-ToBI (Korean ToBI) Labelling Conventions (Version 3.0.)*, Unpublished manuscript, UCLA, 2000.
- [14] S-A. Jun, S-H. Lee, K. H. Kim, Y. J. Lee, “Labeler Agreement in transcribing Korean Intonation with K-ToBI”, *Proceedings of the 2000 International Conference on Spoken Language Processing*, Vol. 3, pp.211-214, 2000.
- [15] S-H. Lee, “Intonational domains of the Seoul dialect of Korean”, *An abstract in Journal of the Acoustical Society of America*, Vol. 85, suppl. Vol. 1, S99, 1989.
- [16] J. B. Pierrehumbert, “The Phonology and Phonetics of English Intonation”, *Doctoral dissertation*, M.I.T., 1980.

접수일자: 2003년 05월 13일

게재결정: 2003년 06월 12일

▶ 이숙향(Sook-Hyang Lee)

주소: 전북 익산시 신용동 344-2 원광대학교 영중어문학부

소속: 원광대학교 영어영문학과

전화: 063) 850-6913

FAX: 063) 850-7454

E-mail: shlee@wonkwang.ac.kr

▶ 신지영(Jiyoung Shin)

주소: 136-701 서울특별시 성북구 안암동 5가 1

소속: 고려대학교 문과대학 국어국문학과

전화: 02) 3290-1973

FAX: 02) 926-8385

E-mail: shinjy@korea.ac.kr

▶ 김봉완(Bong-Wan Kim)

주소: 570-749 전북 익산시 신용동 344-2 원광대학교 음성정보기술산업지원센터

소속: 원광대학교 음성정보기술산업지원센터

전화: 063) 850-7452~3

FAX: 063) 850-7454

E-mail: bwkim@sitec.or.kr

▶ 이용주(Young-Ju Lee)

주소: 570-749 전북 익산시 신용동 344-2 원광대학교 전기전자 및 정보공학부

소속: 원광대학교 전기전자 및 정보공학부, 음성정보기술산업지원센터

전화: 063) 850-7451

FAX: 063) 850-7454

E-mail: yjlee@wonkwang.ac.kr

한국어 공통 음성 DB 구축 및 오류 검증*

이수종(ETRI), 김상훈(ETRI), 이영직(ETRI)

<차 례>

- | | |
|---------------|---------------|
| 1. 서론 | 3. 오류 검증 |
| 2. 1차년도 DB 구축 | 3.1. 오류 검증 대상 |
| 2.1. DB 구성 | 3.2. 인식 실험 |
| 2.2. DB 구조 | 3.3. 오류 유형 분류 |
| 2.3. DB 구축 결과 | 3.4. DB 보완 |
| 2.4. DB 배포 | 4. 향후 계획 |
| | 5. 결론 |

<Abstract>

Common Speech Database Collection and Validation for Communications

Soo-jong Lee, Sanghun Kim, Youngjik Lee

In this paper, we'd like to briefly introduce Korean common speech database, which project has been started to construct a large scaled speech database since 2002. The project aims at supporting the R&D environment of the speech technology for industries. It encourages domestic speech industries and activates speech technology domestic market. In the first year, the resulting common speech database consists of 25 kinds of databases considering various recording conditions such as telephone, PC, VoIP etc. The speech database will be widely used for speech recognition, speech synthesis, and speaker identification. On the other hand, although the database was originally corrected by manual, still it retains unknown errors and human errors. So, in order to minimize the errors in the database, we tried to find the errors based on the recognition errors and classify several kinds of errors. To be more effective than typical recognition technique, we will develop the automatic error detection method. In the future, we will try to construct new databases reflecting the needs of companies and universities.

* Keywords: speech DB, Validation

* 본 연구는 정보통신부 출연 "음성정보처리기반기술" 과제의 일환으로 수행되었습니다.

1. 서 론

음성에 의한 정보기기의 제어를 가능하게 하려는 노력은 20세기 중반 컴퓨터가 개발된 이래 선진국을 중심으로 꾸준히 시도되어 왔다. 음성은 사람들 간의 의사소통을 위해 익숙한 수단으로서 이제 컴퓨터의 활용이 일반화되고 휴먼 인터페이스 개념과 접목되면서 21세기를 선도할 10대 유망 기술로 부상되어 있다. 특히, 각종 정보기기의 소형화와 이동성 추구에 따라 음성 인터페이스의 도입이 필연적인 요건으로 되어 있어서, 음성 기술의 적용은 피할 수 없는 과제가 되어 있다. 음성 처리 기술을 개발하기 위해서는 음성/텍스트 DB가 기반 요소로 요구된다. 이미 미국, 유럽 등의 선진국에서는 산/학/연 공동으로 연구기관을 설립하여 공통 음성 DB를 구축 보급하고 있다[1]. 미국에서는 ARPA와 DARPA를 통하여 수집한 DB를 기반으로 1990년대 초에 LDC를 구성하였고, 유럽에서는 1995년에 비영리기관으로 ELRA를 만들었다. 국내에서도 최근 100여개의 음성 전문업체가 설립되어 다양한 분야에 음성 정보 기술을 응용하고 있다. 그 동안 대다수의 음성 유관 기관과 음성 처리 업체들은 개별적으로 소규모의 DB를 구축하여 연구와 자체 엔진 개발에 활용하여 왔으나, DB 구축에 많은 시간과 비용이 소요되어 외국업체에 비해 음성 기술 경쟁력이 떨어진다는 지적을 받아왔다. 특히 국내 업체간 중복 DB 구축으로 인해 국가적으로 자원이 비효율적으로 활용되고 있어 음성 정보처리 관련 사업자들의 공동 이익을 도모할 수 있는 공통 음성 DB 구축이 시급하였다. 자동차 산업 등 전통 산업 분야에 특화된 대규모 음성 DB의 경우에는 음성정보기술 산업지원센터(SITEC, 원광대)에서 구축하고 있으나, 수요 기반이 확장일로에 있는 통신망 환경에서의 DB 구축은 음성 처리 업계의 요구 사항을 충분히 반영하지 못하고 있었다[2].

국내 음성 업계의 기술 개발과 애로 기술 지원을 목적으로 2001년도 후반에 한국전자통신연구원에 음성/언어정보연구센터를 설립하였으며, “언어 정보 처리 기술 개발(2003.1-2005.12)” 사업을 통해 다양한 통신망 환경에서 대규모 음성 DB를 구축 배포하고 있다[3, 4]. 2002년도부터 통신망에서의 한국어 공통 음성 DB를 대규모로 구축하는 프로젝트가 본격 시작되었다. 이는 한국어 음성처리를 위한 기반 기술로 활용되는 것을 목표로 하고 있다.

본 고에서는 국내 보급을 위해 수행한 한국어 공통 음성 DB의 구축 내용에 대하여 전반적으로 살펴보고, 음성 DB의 무결성을 목표로 오류 검증을 수행한 예를 소개하고자 한다. 서론에 이어 2장에서는 1차년도에 구축한 DB 구축 내용을 집약한 결과를 보여준다. 3장에서는 PC 환경 하에서 중가 마이크를 통하여 수집한 훈련용 단어 음성 DB에 대한 오류 검출과 오류 유형 분류 결과를 서술한다. 4장에서는 향후의 DB 구축 계획을 소개하고, 5장에서 결론을 맺는다.