한국어 발음 교육을 위한 음성 DB 구축 방안

정명숙(고려대)

<차 례>

1. 서론
2. 한국어 학습자의 발음 오류 유형
 2.1. 자음의 오류 유형
 2.2. 모음의 오류 유형
 2.3. 억양의 오류 유형
3. 한국어 교육용 음성 DB 자료
 3.1. 억양 교육용 음성 DB 자료
 3.2. 분절음 교육용 음성 DB 자료
4. 한국어 교육용 음성 DB 구축 설계
 4.1. 음성 자료 수집
 4.2. 음성 정보 주석
5. 결론

<Abstract>

Designing of Speech DB for Korean Pronunciation Education

Myungsook Jung

The purpose of this paper is to design Speech Database for Korean pronunciation education. For this purpose, I investigated types of speech errors of Korean-learners, made texts for recording, which involves all types of speech errors, and showed how to gather speech data and how to tag their informations. It's natural that speech data should include Korean-learners' speech and Korean people's speech, because Speech DB that I try to develop is for teaching Korean pronunciation to foreigners. So this DB should have informations about speakers and phonetic informations, which are about phonetic value of segments and intonation of sentences. The intonation of sentence varies with the type of sentence, the structure of prosodic units, the length of a prosodic unit and so on. For this reason, Speech DB must involve tags about these informations.

* Keywords: speech database, korean language education, pronunciation, speech error, phonetic information

1. 서 론

한국어 학습자들이 한국어를 배울 때 가장 어려워하는 것 중의 하나가 바로 발음이다. 한국어 교육의 다른 하위 분야, 즉 어휘, 문법 등은 개인의 노력에 따라 실력이 향상되는 것이 가시적으로 드러나는 데 반해 발음은 고급 학습자의 경우에도 쉽게 고쳐지지 않는 경우가 대부분이다.

이와 같이 학습자의 발음 오류를 수정하는 것이 쉽지 않은 이유는 여러 가지가 있다. 그 중에서 몇 가지를 제시하면 다음과 같다.

우선 일반적인 한국어 교육은 성인 학습자를 대상으로 이루어지고 있는데, 이들은 이미 모국어의 음운 체계를 완전히 습득한 상태이기 때문에 외국어의 낯선 발음을 새로 익히기가 쉽지 않다. 또한 외국어의 발음 중에는 학습자들의 모국어에 없는 생소한 음이 있는데, 이런 음은 학습자들은 인식조차 할 수 없는 것이기 때문에 외국어의 발음을 익히는 것은 무척 어려운 일이다.

다음으로 실제로 한국어 수업은 일반적으로 다양한 언어권의 학습자들을 대상으로 하여 이루어지는데, 학습자 모국어의 간섭 현상에 따라 실제 학습자들이 보이는 발음 오류 유형이 다르기 때문에 이들 모두를 대상으로 하는 발음 교육은 효율적이지 못하다는 점을 들 수 있다. 따라서 효과적인 발음 교육을 위해서는 학습자들을 언어권별로 묶어서 시행하는 것이 좋다.

그러나 학습자들을 동일 언어권별로 묶어서 함께 교육을 한다고 하더라도 학습자 개개인에 따라 주로 나타내는 발음 오류 유형이 다르며, 개인의 구강 구조, 치열, 주로 사용하는 근육 등에 개인차가 있기 때문에 다수의 학습자들을 상대로 발음 교육을 한다는 것은 현실적으로 매우 힘들다. 따라서 개별적으로 발음 교육을 하는 것이 가장 효과적이라고 할 수 있다.

따라서 학습자를 개별적으로 지도하는 것이 발음 교육의 효과를 극대화할 수 있는 방법이다. 그러나 현실적으로 교사의 수, 특히 한국어 발음 교육을 할 수 있는 교사의 수는 한정적일 뿐만 아니라 모든 학습자들을 개별적으로 지도해야 한다면 경제적인 면에서도 낭비가 아닐 수 없다. 따라서 개별적인 발음 교육을 대신할 수 있는 발음 교육 프로그램이 개발된다면 학습자들은 교사의 도움을 받지 않고 스스로 발음을 교정할 수 있게 될 것이다.

그런데, 한국어 학습자를 위한 자가 발음 교육 프로그램 개발을 위해서는 음성 데이터베이스(이하 음성 DB) 구축이 무엇보다도 필요하다. 최근 한국어 교육을 목적으로 학습자 말뭉치를 구축하는 경우가 늘고 있고, 이렇게 구축된 학습자 말뭉치가 한국어 교육 연구에 유용한 자료로 사용되고 있다[1][2][3][4][5][6]. 그러나 아직까지 한국어 교육용 음성 DB에 대한 연구는 전무한 상태라 할 수 있다. 일반 목적의 음성 DB는 한국전자통신연구원, 한국통신 멀티미디어연구소 등에서 자체적으로 구축하고 있으나, 이 DB들은 각 기관 자체의 연구, 개발을 목적으로 구축

한 것으로서 한국어 발음 교육에 그대로 이용할 수는 없다.

이에 본고에서는 한국어 발음 교육에 이용될 수 있는 음성 DB 구축 방안을 마련하여 실제 한국어 발음 교육용 음성 DB 구축에 도움을 주고자 한다. 신지영(2001)에서 제시한 바와 같이 음성 DB 구축은 설계, 녹음, 주석의 세 단계로 이루어지는데, 이 중 설계가 가장 핵심적인 단계라고 할 수 있다. 녹음과 주석 작업은 설계에 따라서 수행하기만 하면 되는 것이기 때문이다.

한국어 발음 교육용 음성 DB 구축을 위해서는 먼저 한국어 학습자들이 주로 어떤 발음 오류를 보이는지를 살펴보고, 이들을 교정할 수 있는 자료를 선정해야 한다. 이에 2장에서는 한국어 학습자들이 주로 나타내는 발음 오류 유형을 살펴보고자 한다. 3장에서는 이러한 발음 오류 유형을 억양과 분절음의 측면으로 분류하여 이들을 교정할 수 있는 음성 DB 구축용 자료를 선정하게 될 것이다. 4장에서는 실제로 음성 DB를 구축하기 위한 설계를 하고자 한다.

본고의 연구를 토대로 음성 DB가 구축이 되고 이를 이용한 자가 한국어 발음 교육 프로그램이 개발된다면 날로 증가하고 있는 한국어 학습자들에게 효과적인 발음 교육을 할 수 있게 될 것이다. 뿐만 아니라 본고의 음성 DB 구축 방법과 이를 이용해 개발된 발음 교육 프로그램이 한국어 교육뿐만 아니라, 모국어 화자를 대상으로 하는 국어 발음 교육에도 유용하게 사용될 수 있을 것이다.

2. 한국어 학습자의 발음 오류 유형

본고의 연구는 궁극적으로 한국어 발음 교육 프로그램을 개발하는 데 있다. 이를 위해서는 먼저 한국어 학습자들이 어떤 발음을 잘못하는지를 살펴보고, 이런 발음들을 집중적으로 훈련할 수 있도록 해야 할 것이다. 따라서 한국어 발음 교육 프로그램 개발을 위한 음성 DB 역시 이러한 항목들이 포함되어야 할 것이다. 이에 이 장에서는 한국어 학습자들이 주로 나타내는 발음 오류에는 어떤 것들이 있는지 살펴보고자 한다.

한국어 학습자들의 발음 오류에 관한 기존의 연구는 대부분 자음과 모음의 발음에 관한 것이었다[8][9][10][11][12][13][14][15][16][17]. 그러나 억양 교육에 대한 연구는 1990년대에 이르러서야 조금씩 이루어지고 있다[18][19][20].

그러나 개별 분절음에 비해 억양을 익히기가 더욱 어렵다는 점에서도 억양 교육은 강조되어야 한다. 뿐만 아니라 외국인들이 한국어의 자음, 모음 발음을 정확하게 하더라도 억양이 부자연스러우면 한국인은 이들의 발화가 외국인의 발화라는 것을 쉽게 알아차리게 된다. 이와 반대로 개별 음소의 발음이 다소 부정확하더라도 억양이 자연스러우면 매우 유창한 한국어를 구사한다는 인상을 주게 된다. 이러한 이유로 한국어 학습자들이 자연스러운 한국어 억양을 익히게 되면 학습자

들은 자신의 발화에 자신감을 갖게 되며, 이는 한국어 능력을 전반적으로 향상시켜 주는 긍정적인 역할을 하게 된다. 이와 같은 이유로 억양 교육은 개별 분절음의 발음 교육 못지 않게 중요한 부분이라고 할 수 있다. 이에 본고에서는 한국어 학습자들의 발음 오류 유형을 자음과 모음의 발음뿐만 아니라 억양의 오류 유형도 함께 살펴보고자 한다.

학습자들의 발음 오류는 학습자들이 사용하고 있는 모국어에 따라 달리 나타나며 교육 내용도 달라져야 한다. 따라서 본고에서는 학습자들의 언어를 고려하여 대표적인 예로 일본어권 학습자, 중국어권 학습자, 영어권 학습자의 세 언어권별로 학습자들을 분류하여 이들 학습자들에게서 발견되는 발음 오류를 유형별로 제시하고자 한다.

2.1. 자음의 오류 유형

이 절에서는 한국어 학습자들이 잘 하지 못하는 한국어 자음 발음으로 어떤 것이 있는지 살펴보고, 그 오류의 유형이 어떤 것인지를 알아보고자 한다. 기존의 발음 오류에 관한 연구에서 가장 많이 지적되고 있을 뿐만 아니라 선험적으로도 가장 많은 오류를 보일 것으로 예상되는 것이 바로 평음, 격음, 경음의 구별이다. 이 세 계열의 자음을 구별하지 못하는 것은 몽골 학습자를 제외하고는 거의 대부분의 외국인들에게서 발견된다. 이 외에도 중국어권 학습자들의 경우에는 'ㄹ' 발음을, 일본어권 학습자들의 경우에는 받침 발음을 제대로 하지 못한다고 알려져 있다. 이 외에도 다양한 발음 오류들이 존재하며 이들의 발음 오류가 어떻게 나타나는지 학습자들의 언어권별로 제시하면 다음과 같다.[1)]

(1) 일본어권 학습자들의 자음 발음 오류 유형

a. 평음, 격음, 경음 (/k : k^h : k'/, /t : t^h : t'/, /p : p^h : p'/, /tɕ : $tɕ^h$: tɕ'/, /s : s'/)

- 어두의 평음을 격음으로 (/가수/ [kasu] → [k^hasɯ])
- 어두의 평음을 유성음으로 (/바구니/ [paguni] → [bagɯni])
- 공명음 사이의 평음을 무성음으로 (/아기/ [agi] → [aki])
- 어두의 경음을 격음으로 (/딸기/ [t'algi] → [t^harɯgi])
- 어두의 경음을 평음으로 (/빨래/ [p'alle] → [paɾe])
- 어중의 격음을 경음으로 (/같아요/ [kat^{h}ajo] → [kat'ajo])
- 경마찰음 뒤 평마찰음을 경음으로 (/있어서요/ [is'ʌsʌjo] → [is'ʌs'ʌjo])
- 파찰음이 비고모음과 결합할 때 치경경구개파찰음을 경구개음으로

1) 본고에서 제시한 한국어 학습자들의 발음 오류 유형은 한국어 학습자의 발음 오류에 관한 연구에서 제시된 것을 종합하고, 필자가 한국어 발음 교육에 대한 연구를 진행하면서 수집한 자료들을 보충하여 정리한 것이다.

(/짜장면/ [tɕ'adʑaŋmjʌn] → [ʧ'adʒaŋmjʌn])

b. 비음과 폐쇄음 받침 (/m : n : ŋ/, /p : t : k/)

- 어말의 연구개음을 치경음이나 양순음으로
  (/시장/ [ʃidʑaŋ] → [ʃidʑan] [ʃidʑam])
- 어말의 치경음을 연구개음이나 양순음으로
  (/한 잔/ [handʑan] → [handʑaŋ] [handʑam])
- 어중의 연구개음 받침을 뒤의 자음에 위치 동화
  (/장만/ [tɕaŋman] → [tɕamman], /퐁당/ [pʰoŋdaŋ] → [pʰondaŋ])
- 어중의 양순음 받침을 뒤의 자음에 동화 (/밥도/ [papˀt'o] → [pat'o])
- 어중의 양순음 받침을 외파화 (/김치/ [kimtɕʰi] → [kimɯtɕʰi])

c. 유음 (/l/)

- 유음 받침(설측음)을 탄설음으로 (/딸기/ [t'algi] → [t'aɾɯgi])
- 유음 받침 탈락 (/울타리/ [ultʰaɾi] → [ɯtʰaɾi])
- 어중의 유음 겹자음을 탄설음으로 (/빨리/ [p'alli] → [p'aɾi])

d. 성문마찰음 (/h/)

- 성문마찰음이 원순 모음, 활음과 결합할 때 성문마찰음을 양순마찰음으로
  (/후에/ [hue] → [ɸɯe], /화장/ [hwadʑaŋ] → [ɸadʑaŋ])

(2) 중국어권 학습자들의 자음 발음 오류 유형

a. 평음, 격음, 경음 (/k : kʰ : k'/, /t : tʰ : t'/, /p : pʰ : p'/, /tɕ : tɕʰ : tɕ'/, /s : s'/)

- 어두의 평음을 격음으로 (/가수/ [kasu] → [kʰasu])
- 어두의 평음을 경음으로 (/가수/ [kasu] → [k'as'u])
- 공명음 사이의 평음을 경음으로 (/야구/ [jagu] → [jak'u])
- 어두의 격음을 경음으로 (/키/ [kʰi] → [k'i])
- 자음 뒤 격음을 경음으로 (/독특/ [tokˀtʰik] → [tokˀt'ik])
- 어중의 경음을 격음으로 (/바빠요/ [pap'ajo] → [papʰajo])
- 평마찰음을 경음으로 (/사람/ [saɾam] → [s'aɾam], /이사/ [isa] → [is'a])
- 구개음 '씨'를 치경음으로 (/몇 시/ [mjʌʃ'i] → [mjəs'i])

b. 비음과 폐쇄음 받침 (/m : n : ŋ/, /p : t : k/)

- 어중의 연구개음 탈락 (/닦다/ [tak't'a] → [tat'a], /국물/ [kuŋmul] → [kumur])
- 어말의 폐쇄음 받침 외파화 (/입/ [ipˀ] → [ipʰi], /밑/ [mitˀ] → [mitʰi])

c. 유음 (/l/)

- 유음 받침(설측음)을 권설음으로 (/딸기/ [t'algi] → [t'argi])
- 어중의 유음 겹자음을 단자음으로 (/빨리/ [p'alli] → [p'ali])
- 탄설음을 설측음으로 (/우리/ [uɾi] → [uli], /라디오/ [ɾadio] → [ladio])

d. 성문마찰음 (/h/)

- 성문마찰음을 연구개마찰음으로 (/하루/ [haɾu] → [xalu])

(3) 영어권 학습자들의 자음 발음 오류 유형

a. 평음, 격음, 경음 (/k : k^h : k'/, /t : t^h : t'/, /p : p^h : p'/, /tɕ : $tɕ^h$: tɕ'/, /s : s'/)
- 어두의 평음을 격음으로 (/가수/ [kasu] → [k^hasu])
- 어두의 평음을 유성음으로 (/바구니/ [paguni] → [baguni])
- 어중의 평음을 격음으로 (/야구/ [jagu] → [jak^{h}u])
- 공명음 사이의 평음을 무성음으로 (/아기/ [agi] → [aki])
- 어두의 경음을 격음으로 (/빨리/ [p'alli] → [p^halli])
- 어중의 경음을 격음으로 (/바빠요/ [pap'ajo] → [pap^{h}ajo])
- 평마찰음을 경음으로 (/사람/ [saɾam] → [s'aɾam], /이사/ [isa] → [is'a])
- 구개음 '시'나 '씨'를 '씨'를 치경음으로 (/몇 시/ [mjʌʃ'i] → [mjʌs'i])
- 구개음 '시'나 '씨'를 원순 구개음으로 (/시간/ [ʃigan] → [$ʃ^w$igan])
- 치경경구개파찰음을 경구개음으로 (/차요/ [$tɕ^h$ajo] → [$tʃ^h$ajo])

b. 비음과 폐쇄음 받침 (/m : n : ŋ/, /p : t : k/)
- 폐쇄음 받침 외파화 (/입/ [ip˺] → [ip^h], /밑/ [mit˺] → [mit^h])

c. 유음 (/l/)
- 어중의 유음 겹자음을 단자음으로 (/빨리/ [p'alli] → [p'ali])
- 탄설음을 권설음으로 (/우리/ [uɾi] → [uɾi], /라디오/ [ɾadio] → [radio])

위의 (1), (2), (3)에서 제시한 바와 같이 일본어권, 중국어권, 영어권 학습자들이 공통적으로 어려움을 느끼는 한국어 자음 발음은 평음, 격음, 경음의 구별과 유음의 발음이다. 일본어권 학습자들과 중국어권 학습자들의 경우에는 이 외에도 받침 발음을 잘 구별하지 못한다. 일본어권 학습자들은 폐쇄음 받침을 탈락시키기도 하고, 모음을 삽입하여 이 자음을 다른 음절의 초성으로 발음하기도 한다. 중국어권 학습자들도 연구개음 받침을 탈락시키기도 하고, 어말에서는 자음 뒤에 모음을 삽입하여 발음한다는 특성이 있다. 또한 중국어권 학습자들의 경우에는 음절 하나하나를 끊어서 발음하려는 경향이 있어서 선행 음절의 음절 말음을 다음 음절의 초성으로 연음 시켜 발음하지 못하는 오류가 많이 나타난다.

2.2. 모음의 오류 유형

이 절에서는 한국어 모음 발음 중에서 외국인 학습자들이 자주 오류를 나타내는 것들로 어떤 것이 있는지 살펴보고, 그 오류 유형을 알아보고자 한다. 한국어의 모음 중 / ㅓ/와 /ㅡ/는 매우 특징적인 모음이다. 그 결과 외국인들은 이 두 모음의 발음에서 가장 많은 오류를 보인다. 이 외에도 이중모음의 발음 등 다양한 발

음 오류들이 존재하는데 이들의 발음 오류가 어떻게 나타나는지 학습자들의 언어권별로 살펴보기로 하자.

(4) 일본어권 학습자들의 모음 발음 오류 유형

a. 단모음

- / ㅓ/를 원순음으로 (/언니/ [ʌnni] → [onni])
- /ㅡ/를 원순음으로 (/그래서/ [kɨresʌ] → [kuɾeso])
- /ㅜ/를 평순음으로 (/구두/ [kudu] → [kʰɯdɯ])

b. 이중모음

- / ㅕ/를 /ㅛ/로 (/여기/ [jʌgi] → [jogi])
- /ㅢ/를 /ㅜ ㅣ/로 (/의사/ [ɯisa] → [uisa])
- /ㅢ/를 /ㅟ/로 (/의사/ [ɯisa] → [wisa])
- 자음 뒤 w계 이중모음을 단모음으로
  (/도서관/ [tosʌgwan] → [tosʌgan], /쉬워요/ [swiwʌjo] → [ʃiwʌjo])

(5) 중국어권 학습자들의 모음 발음 오류 유형

a. 단모음

- / ㅓ/를 원순음으로 (/언니/ [ʌnni] → [onni])
- / ㅓ/를 전설모음으로 (/선생/ [sʌnseŋ] → [s'ens'əŋ])
- /ㅗ/를 평순음으로 (/가족/ [kadʑok˺] → [katɕək˺])
- /ㅡ/를 중모음으로 (/오늘/ [onɨl] → [onər], /일층/ [ilʧʰiŋ] → [irʧʰəŋ])
- /ㅜ/를 평순 중모음으로 (/문화/ [munhwa] → [mənhua])
- / ㅐ/가 폐음절에 사용된 경우 / ㅓ/나 / ㅏ/로
  (/학생/ [hak˺s'eŋ] → [hak˺s'əŋ], /은행/ [inheŋ] → [əihaŋ], [ihaŋ])

b. 이중모음

- / ㅕ/를 /ㅛ/로 (/여기/ [jʌgi] → [jogi])
- /ㅛ/를 / ㅕ/로 (/요일/ [joil] → [jəir])
- / ㅑ/가 폐음절에 사용된 경우 / ㅕ/로 (/양말/ [jaŋmal] → [jəŋmar])
- / ㅒ, ㅖ/를 단모음으로 (/얘기/ [jegi] → [egi], /옛날/ [jennal] → [ennar])
- 이중모음을 복모음으로
  (/년/ [njʌn] → [nien], /겨울/ [kjʌul] → [kiəur], /과자/ [kwadʑa] → [k'uatɕa], /죄송해요/ [ʧwesoŋhejo] → [ʧues'oŋhjo], [ʧuis'əŋhejə])

(6) 영어권 학습자들의 모음 발음 오류 유형

a. 단모음

- / ㅓ/를 원순음으로 (/언니/ [ʌnni] → [ɔnni])

- /ㅡ/를 원순음으로 (/그래서/ [kɨresʌ] → [kuresɔ])

b. 이중모음

- / ㅖ/를 /ㅛ/로 (/여기/ [jʌgi] → [jogi])
- /ㅢ/를 /ㅜ ㅣ/로 (/의사/ [ɯisa] → [uis'a])
- /ㅢ/를 /ㅟ/로 (/의사/ [ɯisa] → [wis'a])

위의 (4), (5), (6)에서 제시한 바와 같이 일본어권, 중국어권, 영어권 학습자들이 공통적으로 어려움을 느끼는 한국어 모음 발음은 / ㅓ/와 /ㅡ/의 발음이며, 이 두 모음의 발음이 어렵기 때문에 이들 모음이 사용된 이중 모음 / ㅖ/와 /ㅢ/의 발음 역시 많은 오류를 나타내고 있다. 일본어권 학습자들의 경우에는 이 외에도 w계 이중모음을 잘 발음하지 못한다는 특징이 있다. 한편 중국어권 학습자들은 중국어의 이중모음이 활음과 모음의 결합으로 이루어지지 않고, 모음과 모음의 결합으로 이루어진 복모음이기 때문에 한국어의 이중모음을 복모음으로 발음하는 오류를 일으키는 것으로 나타났다. 이 외에도 중국어권 학습자들의 경우에는 중국어의 운모 체계에 없는 모음과 음절말 자음의 결합을 잘 발음하지 못한다는 특징이 있다.[2)]

2.3. 억양의 오류 유형

이 절에서는 외국인들이 한국어를 발화할 때 나타나는 억양의 오류로는 어떤 것들이 있는지 살펴보고, 그 오류 유형을 알아보고자 한다. 한재영 (2001)에서는 발음 교재에서 다루어야 할 내용으로 강세, 장단음, 성조 등을 포함하고 있는데, 한국 표준어에는 강세와 성조가 없으므로 이에 대한 교육 역시 불필요하다. 장단음에 대한 교육 역시 현재 장년층 이하의 세대에서는 장단음의 구별이 모호해졌기 때문에 한국인들도 장단음 구별에 오류를 범하고 있다고 할 수 있다[23]. 따라서 한국인들도 범할 수 오류가 아닌 외국인이기 때문에 범할 수밖에 없는 오류만이 논의의 대상이 되어야 할 것이다. 따라서 본고에서는 억양의 오류만을 중심으로 살펴보고자 한다.

억양이라는 개념은 크게는 '발화 시 음높이(pitch)가 일정한 유형을 나타내면서 만들어내는 말의 가락(speech melody)'이라고 할 수 있다. 그러나 억양은 화자의 감정을 표현하는 감정적, 정서적 기능을 하기도 하며, 의문문과 서술문과 같은 문장

2) 우인혜 (2001), 한재영 (2001) 등 여러 논문에서는 / ㅐ/와 / ㅔ/를 구별하지 못하는 것, 또 /ㅚ/와 /ㅟ/를 이중모음으로 발음하는 것을 발음 오류로 지적하고 있다. 그러나 대부분의 한국인들이 / ㅐ/와 / ㅔ/를 구별하지 못하고[22], /ㅚ/와 /ㅟ/를 이중모음으로 발음하는 것이 일반적이므로[23] 이 두 모음을 동일한 음으로 교육하는 것이 좋다. 왜냐하면 한국어 교육에서 발음 교육의 목표가 한국어 모어 화자들과의 의사소통에 있으므로 한국인들이 실제로 사용하고 있는 현실음을 교육하는 것이 의사소통 능력 향상에 도움이 되기 때문이다.

의 종류를 결정하는 통사적 기능도 담당한다. 이 외에도 억양은 발화의 운율 구조를 나타내는 기능을 담당하기도 한다. 그러나 억양이 표현하는 감정적, 정서적 의미를 정확하게 기술하기 어려운 경우가 많기 때문에 본고에서는 문장의 종류를 결정하는 문미 억양과 발화의 운율 구조를 나타내는 억양에 국한해서 살펴보고자 한다.

(7) 일본어권 학습자들의 억양 오류 유형

a. 문미 억양

- 평서문의 끝음절을 올렸다 내림. (/잤어요./ [H-L-HL])
- 평서문의 마지막 음절까지 음높이를 떨어뜨리지 않고 어느 정도 유지
- 의문사 의문문의 끝에서 두 번째 음절을 낮추고 끝음절을 급격하게 상승 (/뭐 먹었어요?/ [L-H-L-L-H])
- 긍부정 의문문의 끝음절을 내렸다 올려서 발음 (/빵 먹었어요?/ [HL-L-H-L-LH])

b. 문중 억양

- 강세구의 마지막 음절을 내림. (/가게에서/ [H-L-L-L])
- 강세구의 마지막 음절을 올렸다 내림. (/가게에서/ [H-L-L-HL])
- 강세구나 억양구의 마지막 음절을 선행 음절과 같은 길이로 짧게 발음
- 강세구나 억양구의 마지막 음절을 지나치게 길게 발음
- 강세구의 첫음절을 높은 음으로 발음 (/마리아 씨는/ [H-L-L-L-HL])
- 폐음절을 높은 음으로 길게 발음 (/친구가/ [HH-L-L])
- 폐음절과 인접한 개음절을 길게 발음(/친구가/ [LL-HL-L])

(8) 중국어권 학습자들의 억양 오류 유형

a. 문미 억양

- 의문사 의문문의 끝음절을 내려서 발음 (/언제 가요?/ [L-L-H-L])
- 의문사 의문문의 끝에서 두 번째 음절을 낮추고 끝음절을 급격하게 상승 (/뭐 먹었어요?/ [L-H-L-L-H])
- 긍부정 의문문의 끝에서 두 번째 음절의 하강 없이 끝음절만 올려서 발음 (/빵 먹었어요?/ [H-L-H-M-H])

b. 문중 억양

- 강세구의 마지막 음절을 내림. (/가게에서/ [L-L-H-L])
- 강세구의 끝에서 두 번째 음절을 높은 음으로 발음 (/우리들은/ [L-L-H-L])
- 한자어의 경우 중국어의 성조에 따라 발음 (/영국에서/ [H-L-L-L])

(9) 영어권 학습자들의 억양 오류 유형

a. 문미 억양
- 의문사 의문문의 끝음절을 내려서 발음 (/언제 가요?/ [L-H-L-L])
- 의문문의 끝에서 두 번째 음절을 하강시키지 않고 앞에서 마지막 강세구의 두 번째 음절부터 완만하게 상승
(/밥 먹었어요?/ [L-L-H-H-H])

b. 문중 억양
- 강세구의 첫음절을 높은 음으로 강하게 발음 (/가게에서/ [H-L-L-L])
- 강세구의 마지막 음절을 내려서 발음 (/우리들의/ [H-L-L-L])
- 강세구의 마지막 음절을 올렸다 내림. (/우리들의/ [H-L-L-HL])

위의 (7), (8), (9)에서 제시한 바와 같이 일본어권, 중국어권, 영어권 학습자들에게서 공통적으로 발견되는 억양의 오류는 강세구의 끝을 올리지 않고 내려서 발음하는 것과 의문문의 끝에서 실현되는 상승조의 억양을 제대로 실현하지 못한다는 것이다. 특히 영어권 학습자들의 경우에는 끝에서 두 번째 음절에서 하강이 이루어지지 않고 강세를 받는 음절부터 상승이 이루어지는 특징이 발견된다. 일본어권 학습자들의 경우에는 일본어의 모라 언어적 특성으로 인하여 개음절의 경우에는 1박자로, 폐음절의 경우에는 2박자로 길게 발음하는 경향이 있다. 또한 중국어권 학습자들의 한국어 발화도 중국어가 성조 언어라는 특징 때문에 한자어의 경우 중국어의 성조대로 발음하려는 경향이 있는 것으로 나타났다.

3. 한국어 교육용 음성 DB 자료

이 장에서는 한국어 교육용 음성 DB 구축을 위한 녹음용 자료에 대해 살펴보고자 한다. 한국어 교육용 음성 DB 구축은 한국어 학습자들이 자신의 발음을 교정할 수 있는 프로그램 개발의 기초 자료를 제공하는 것을 목적으로 하기 때문에 학습자들의 발화뿐만 아니라 모범이 될 만한 한국인의 한국어 발화를 모두 대상으로 해야 한다.

이를 위해서 한국어 학습자들의 실제 발화를 녹음하고 그 자료를 한국인 화자가 다시 바른 발음으로 읽는 방안도 생각해 볼 수 있으나, 한국어 학습자들이 사용하는 문장이 정확하지 않은 문장인 경우가 많다는 문제가 있다. 또한 이렇게 할 경우 한국어 발음 교육의 대상이 되지 않는 내용까지 포함하게 되어 경제적인 면에서의 문제점이 있다. 이 외에도 학습자들의 실제 발화는 자신이 즐겨 사용하는 단어와 문장만을 사용하여 그 구조가 매우 단순하며, 또한 자신에게 어려운 발음이나 어려운 단어나 표현의 사용을 전략적으로 회피하여 말하는 경향이 있다. 따라서 한국어 발음 교육의 대상이 되는 다양한 예를 확보하기 어렵다는 문제가 있

다.

따라서 본고에서는 2장에서 제시한 한국어 학습자들의 발음 오류 유형을 고려하여 한국어 발음 교육의 내용을 선정하고, 이런 내용들을 포함하여 음성 DB 구축용 자료를 선정하고자 한다. 이를 위해서 먼저 억양 교육용 음성 DB와 분절음 교육용 음성 DB의 내용을 선정할 때 고려해야 하는 사항을 먼저 살펴보고 이를 토대로 음성 DB 구축용 녹음 자료를 구성해 보고자 한다.[3)]

3.1. 억양 교육용 음성 DB 자료

억양 교육용 음성 DB를 구축하기 위해서는 한국어 학습자들이 주로 나타내는 억양의 오류를 집중적으로 훈련할 수 있는 내용이 포함되어야 한다. 따라서 학습자들이 주로 오류를 나타내는 문미 억양, 특히 의문문의 문미 억양을 반드시 포함하여야 한다. 이에 본고에서는 억양 교육용 음성 DB 구축을 위한 녹음 자료를 대화문으로 구성하여 그 안에 평서문, 의문문, 청유문, 명령문 등 다양한 종류의 문장을 포함하여 구성하고자 한다. 의문문의 경우에도 의문사 의문문과 긍부정 의문문, '-지요?' 형태가 사용된 확인 의문문, 반어적 의미를 나타내는 수사 의문문, 의문사가 사용된 긍부정 의문문 등 다양한 형태의 의문문을 모두 포함하고자 한다.

또한 외국인 한국어 학습자들은 강세구나 억양구의 억양을 제대로 실현하지 못하는 것으로 나타났는데, 대화문을 구성할 때 강세구의 억양까지 측정할 수 있도록 문장의 내부 구조를 다양하게 구성해야 한다. 또한 강세구를 구성하고 있는 음절의 수에 따른 억양의 실현을 알아보기 위해서는 하나의 강세구를 구성하고 있는 음절의 수를 다양하게 해야 한다.

이 외에도 자연스러운 억양을 유도하기 위해서는 구성된 대화문이 유의미한 맥락을 지니고 있어야 하며, 억양의 변화를 잘 관찰할 수 있도록 하기 위해서는 대화문에 사용된 분절음을 공명음 중심으로 해야 한다. 이와 같은 점을 고려하여 억양 교육용 음성 DB 구축을 위한 녹음 자료를 다음과 같이 구성할 수 있다.

(10) 억양 교육용 음성 DB 구축을 위한 녹음 자료
A1: 언니, 오늘 미영이 만나지요?
B1: 아니, 요즘 미영이가 일이 좀 많은가 봐.
오늘은 안 만나기로 했는데, 왜 그래? 미영이에게 뭐 전할 말이라도 있니?
A2: 그런 건 아니에요. 그럼, 언니는 오늘 뭐 해요?

---

3) 본고에서는 억양 교육용 음성 DB의 내용과 분절음 교육용 음성 DB의 내용을 구분하고자 한다. 왜냐 하면 억양 곡선을 살피기 위해서는 가급적 격음과 경음, 마찰음을 배제한 문장을 사용해야 하는데, 분절음 발음 교육에서 중점적으로 다루어야 하는 것이 바로 세계열의 파열음, 파찰음의 구별, 치경 마찰음의 발음이기 때문이다.

B2: 오전에는 영진이랑 공부하고, 오후엔 엄마랑 냉면 먹으러 가려고.
A3: 언니네 엄마는 냉면을 좋아해요?
B3: 우리 엄마는 냉면광이야.
A4: 언니 엄마만 냉면광이에요? 언니도 냉면 좋아잖아요.
B4: 맞아. 나도 냉면을 안 먹으면 좀 우울해져.
A5: 나도 냉면 아주 좋아하는데.
B5: 너도 냉면 먹으러 갈래?
A6: 아니에요. 오늘은 일이 많아요. 다음에 가요.
B6: 그래, 다음에 물냉면 잘 하는 데 있는데, 우리 먹으러 가자.
A7: 좋아요. 다음에는 나도 불러 줘요.

(10)에서 제시한 대화문은 문장의 종류에 따른 문미 억양을 고려하여 평서문(B1, B2, B3, A4, B4, A5, A6), 의문문(A1, A2, A3, A4, B5), 청유문(A6, B6), 명령문(A7)을 모두 포함하도록 구성했으며, 의문문의 종류도 확인 의문문(A1), 의문사 의문문(B1, A2), 긍부정 의문문(A3, B5), 의문사가 사용된 긍부정 의문문(B1), 수사 의문문(A4)로 다양하게 구성하였다.

한편, 강세구를 구성하고 있는 음절의 수에 따른 강세구의 억양을 살펴보기 위해서 강세구의 음절수를 1음절(B4)부터 8음절(B1)까지 다양하게 구성하였다. 또한 억양 곡선을 잘 관찰하기 위해서 격음과 경음, 마찰음(/ㅎ/ 제외)을 전혀 사용하지 않았다. /ㅎ/의 경우에는 공명음 사이에 사용된 경우만을 포함했는데, 이런 환경에서는 /ㅎ/가 일반적으로 약화되거나 탈락되기 때문에[24] 억양 곡선을 살피는 데 문제가 되지 않기 때문이다.

3.2. 분절음 교육용 음성 DB 자료

한국어 분절음 교육용 음성 DB는 모든 한국어의 분절음을 균형 있게 포함하고 있을 필요는 없다. 학습자들이 주로 나타내는 오류를 중심으로 자료를 선정하면 된다. 따라서 자료 선정 과정에서 앞의 2.1.과 2.2.에서 살펴본 한국어 학습자들의 자음, 모음의 발음 오류 유형이 고려되어야 한다. 본고에서는 자음의 발음 교육용 음성 DB와 모음 발음 교육을 위한 음성 DB를 구분하지 않고 이들을 통합하여 음성 DB 내용을 살펴보고자 하는데, 그 이유는 자음이 모음에 의존하지 않고 독립적으로 소리를 이룰 수 없기 때문이다.

먼저 자음의 발음과 관련해서는 격음, 경음의 구별을 관찰할 수 있는 단어가 반드시 포함되어야 한다. 이 때 이 세 계열의 자음이 어두와 어중에 사용된 경우를 모두 포함해야 한다. 또한 받침의 발음 중에서는 비음과 폐쇄음 받침의 조음 위치를 구별하는지를 살필 수 있는 분절음 연쇄를 반드시 포함해야 하며, 이 경우 역시 이 자음들이 어중에 오는 경우와 어말에 위치하는 경우를 모두 포함한다. 유

음의 경우에는 이 음이 음절 초성인 경우와 음절 말음인 경우 모두 오류가 나타나므로 이 두 가지 경우를 모두 살필 수 있는 단어를 사용할 필요가 있다. 또한 일본어권 화자들과 중국어권 화자들의 경우 성문 마찰음의 발음에 오류를 나타내므로 /ㅎ/의 발음도 음성 DB 구축용 녹음 자료에 포함되어야 한다. 이 외에도 특정한 모음과 인접할 때만 자음 발음의 오류가 나타나기도 하는데, 이와 같은 현상을 보이는 분절음 연쇄를 가진 단어를 포함하여 자료를 구성해야 한다.

다음으로 모음의 발음과 관련해서는 /ㅓ/와 /ㅗ/, /ㅡ/와 /ㅜ/의 구별을 모든 언어권 학습자들이 잘 하지 못하므로 이들 모음의 구별을 살필 수 있는 단어를 반드시 포함해야 할 것이다. 또한 이들 모음이 사용된 이중모음, 즉 /ㅕ/, /ㅛ/, /ㅟ/의 발음도 함께 살펴볼 수 있어야 한다. 이 외에도 외국인 학습자들이 자주 범하는 오류 중의 하나가 이중모음을 단모음으로 발음하는 것이다. 또한 중국어권 학습자들은 이중모음을 모음과 모음이 대등하게 결합된 복모음으로 발음하고 있기 때문에 j계 이중모음과 w계 이중모음을 모두 포함하여 음성 DB 구축용 녹음 자료를 선정해야 한다. 또한 중국어권 학습자들의 경우 /ㅖ(ㅒ)/ 모음의 발음도 잘 하지 못하므로 이들 모음의 발음을 살필 수 있는 자료도 포함되어야 할 것이다.

지금까지의 내용을 토대로 한국어 발음 교육의 대상이 되는 분절음을 정리하면 다음과 같다.

(11) 한국어 발음 교육의 대상이 되는 분절음

a. 자음

- 음절 초성: /ㄱ, ㅋ, ㄲ, ㄷ, ㅌ, ㄸ, ㅂ, ㅍ, ㅃ, ㅈ, ㅊ, ㅉ, ㅅ, ㅆ, ㄹ, ㅎ/
- 음절 말음: /ㄴ, ㅁ, ㅇ, ㄱ, ㅂ, ㄷ, ㄹ/

b. 모음

- 단 모 음: /ㅓ, ㅗ, ㅡ, ㅜ, ㅖ(ㅒ)/
- 이중모음: /ㅑ, ㅒ(ㅖ), ㅕ, ㅛ, ㅠ, ㅘ, ㅙ(ㅚ, ㅞ), ㅝ, ㅟ, ㅢ/

위의 (11)에서 제시한 음절 초성 16개, 모음 15개, 음절 말음 7개의 발음이 모두 들어가도록 녹음 자료를 구성해야 할 것이다. 만약 위에서 제시한 분절음들을 조합하여 2음절 단어를 구성한다고 가정해 보자. 각 음절이 개음절인지 폐음절인지, 음절 초성이 있는지 없는지 등을 고려해서 가능한 분절음 연쇄는 다음과 같다.

(12) 한국어 발음 교육 대상 분절음의 조합으로 만든 2음절 단어

a. 모음으로 시작하는 경우: VV, VCV, VVC, VCVC, VCCVC

b. 자음으로 시작하는 경우: CVV, CVCV, CVVC, CVCVC, CVCCV, CVCCVC

(12)에서 제시한 각각의 경우에 분절음의 개수를 산술적으로 계산하면 경우의 수는 3,918,840개가 된다.[4] 그러나 이들 중에는 국어의 형태소 구조 조건이나 음소 배열 제약에 의해(예: 파찰음과 j계 이중모음의 연쇄는 불가능함), 혹은 음운 규칙의 적용에 따라(예: /국밥/ [국빱]의 경우와 같이 폐쇄음 받침 뒤의 평음은 경음화하므로 실제로 두 자음이 인접하는 경우 폐쇄음 받침과 평음의 연쇄는 불가능함), 또 우연한 어휘적 공백(예: */으고/)이 다수 포함되므로 이런 것들을 제외하면 경우의 수는 더욱 줄어들 것이다.

그러나 이러한 경우의 수를 모두 고려하여 자료를 구성할 필요는 없을 것이다. 예를 들어 평음, 격음, 경음을 제대로 구별하고 있는지를 살펴보고자 할 때, '거/커/꺼, 고/코/꼬, 그/크/끄, 구/쿠/꾸, ...' 등과 같이 후행 모음의 종류에 따른 모든 포함시킬 필요는 없다. 따라서 한국어의 분절음 교육용 음성 DB에는 학습자들의 발음 오류 유형을 살펴볼 수 있는 자료들이 포함되면 된다.

분절음 교육용 음성 DB 구축을 위한 녹음 자료는 개별 분절음의 발음을 살필 수 있는 것으로 구성해야 하기 때문에 단어를 단독으로 낭독하도록 하는 것이 경제적이라고 생각할 수 있다. 그러나 단어를 단독으로 읽을 때는 단어가 하나의 발화 단위가 되기 때문에 모든 단어의 끝을 내려서 발음하게 된다. 그러나 이런 식으로 단어를 개별적으로 읽은 것을 샘플로 하여 한국어 학습자들을 교육시킬 경우 학습자들은 그 단어가 강세를 지니고 있는 것으로 오해를 하게 된다. 그 결과 학습자들은 그 단어의 억양을 의식하지 못하는 사이에 익히게 되어 실제 문장 내에서 그 단어를 발화할 때 문제가 된다.

따라서 분절음 교육용 음성 DB의 구축을 위해서는 단어를 단독으로 읽히지 말고 이들을 문장으로 구성하는 것이 좋다. 그렇다고 해서 억양의 경우와 같이 녹음 자료를 굳이 대화문으로 구성할 필요는 없다. 대화문의 경우에는 대화를 이끌어 나가기 위한 담화 표지들이 반복적으로 사용되는데, 그 결과 살펴보고자 하는 분절음의 발음과 관계없는 내용들이 불필요하게 포함될 수 있으므로 경제성이 떨어진다고 할 수 있다.

이러한 점들을 고려하여 본고에서는 실제 사용할 수 있는 녹음 자료를 다음과 같이 구성해 보았다.

---

4) 3,918,840이라는 수는 한국어 발음 교육의 대상이 되는 분절음의 수를 음절 구조상의 위치에 따라 가능한 모든 경우의 수를 계산해서 얻은 것이다. 즉, VV (15×15=225), VCV (15×16=240), VVC (15×15×7=1,575), VCVC (15×16×15×7=25,200), VCCVC (15×7×16×15×7=176,400), CVV (16×15×15=3,600), CVCV (16×15×16×15=57,600), CVVC (16×15×15×7=25,200), CVCVC (16×15×16×15×7=403,200), CVCCV (16×15×7×16×15=403,200), CVCCVC (16×15×7×16×15×7=2,822,400)와 같이 분절음 연쇄에 따라 가능한 모든 조합을 합해서 얻어진 값이다.

(13) 분절음 교육용 음성 DB 구축을 위한 녹음 자료
a. 우리는 뿔뿔이 불이 난 풀밭으로 갔어요.
b. 끼륵끼륵 키가 큰 기러기가 날고 있어요.
c. 토끼집 뒤에 독특한 도끼와 토기가 있어요.
d. 방에 누워서 빵빵한 배를 팡팡 치고 있어요.
e. 아침 일찍부터 참새가 창가에서 짹짹 노래해요.
f. 짜장면하고 짬뽕이 모두 짜요.
g. 시원한 바람이 살랑살랑 불더니 요즘은 날씨가 꽤 쌀쌀해졌어요.
h. 오늘 모임에 못 가요. 중요한 일이 있어서요.
i. 하루는 후암동에서 화장을 한 사람을 봤어요.
j. 여기에 간장 공장과 김치 공장이 있어요.
k. 책상 다리가 밥상 다리보다 길어요.
l. 방울 소리가 딸랑딸랑 들려요.
m. 얼굴을 그릴 때 눈, 코, 입, 턱, 귀, 이런 순서로 그리세요.
n. 모리는 머리 하러 멀리 '모리머리방'에 가요.
o. 신촌에서 버스를 타고 신천에 가요.
p. 구름 그림이란 구름을 그린 그림을 말한다.
q. 의빈 씨는 영희 씨의 의자에 앉아 있어요.
r. 현대 백화점에 가면 양말과 생선 좀 사다 주세요.
s. 이 도서관은 옛날에는 매일 문을 열었는데, 요즘은 일요일에 쉬어요.
t. 외국어를 배우는 게 굉장히 어려워요.

(13a)~(13i)까지는 자음 중 평음, 격음, 경음의 발음과 관련된 것이며, (13a)~(13d)까지는 파찰음, (13e, f)는 파찰음, (13g, h)는 치경 마찰음, (13i)는 성문 마찰음의 발음과 관련된 것이다. 특히 (13c)에 '독특한'이란 단어를 추가한 것은 폐쇄음 받침 뒤에 오는 격음을 경음화하는 경향이 있는데, 이를 고려한 것이다. 또 (13f)는 일본인 화자들이 경파찰음을 경구개치경음이 아니라 경구개음으로 발음하는 경향과 관련된 것이다. (13g)는 치경 마찰음이 /i/ 모음 앞에서 구개음화하는 현상과 관련된 것이고, (13h)는 경마찰음 뒤에 평마찰음이 뒤따라 오는 경우 평마찰음을 경마찰음으로 발음하는 경향이 있는데, 이를 고려하여 구성한 문장이다.

(13j)~(13m)까지는 받침 발음과 관련된 것으로서, (13j)는 비음 받침, (13k)는 폐쇄음 받침의 조음 위치 구별과 관련된 것이다. (13l)은 유음의 발음과 관련된 것으로 받침 [l]과 겹자음 [ll] 음절 초성 [ɾ] 발음을 모두 포함하도록 문장을 구성하였다. (13m)은 한국어에서 자음이 음절 말음 위치에 올 때 이 자음들을 파열하지 않고 불파음으로 발음하는 현상과 관련된 것이다.

(13n)~(13t)까지는 모음의 발음을 살펴보기 위해 구성된 문장들이다. 이 중에서

(13n)~(13p)는 단모음의 발음, (13q)~(13t)는 이중모음의 발음을 살펴보기 위한 것이다.

이상 한국어 분절음 발음 교육을 위한 음성 DB 구축에 사용될 녹음 자료를 제시해 보았다. 위의 (13)에서 제시한 문장들은 맥락은 없지만 모두 유의미한 문장으로 구성하였다. 한 문장 내에 살펴보고자 하는 분절음을 가능한 많이 포함하여 문장의 수를 줄이는 방향으로 녹음 자료를 선정하였다. 정보 처리를 위해서 더 많은 문장이 필요하다면 (13)에서 문장을 구성한 원리에 따라 새로운 문장을 추가하면 될 것이다.

4. 한국어 교육용 음성 DB 구축 설계

앞에서 우리는 한국어 발음 교육용 음성 DB 구축을 위해 필요한 녹음 자료를 외국인들이 주로 발음 오류를 나타내는 현상을 고려하여 구성해 보았다. 이 장에서는 이러한 자료를 사용해 실제 음성 DB를 어떤 방법으로 구축할 것인지 논의해 보고, 실제 음성 DB 구축을 위한 설계를 해 보고자 한다.

4.1. 음성 자료 수집

이 절에서는 외국인들의 한국어 발음 교육에 필요한 항목들이 균형 있게 포함된 음성 자료들을 실제로 어떻게 수집할 것인지에 대해 논의해 보고자 한다. 본고에서 다루고 있는 음성 DB는 한국어 발음 교육을 위한 것이기 때문에 당연히 한국어를 배우는 학습자들의 발화가 대상이 될 것이다. 또한 이들의 발음 교육에 이용될 수 있는 바른 발음 샘플을 제시하기 위해서는 한국인의 발화 역시 대상이 되어야 한다.

한국어 학습자들의 발음에는 모국어의 간섭 현상이 강하게 작용하므로 학습자들을 언어권별로 구분해야 하며, 또한 학습자들의 학습 기간에 따라서도 발음 오류가 달리 나타날 수 있으므로 학습자의 학습 단계도 함께 고려해야 한다. 이 외에도 학습자들의 발음에 영향을 미칠 수 있는 요소로는 학습자의 성별, 연령 등이 있으며, 학습자의 한국 체류 경험, 기간, 부모의 국적 등이 있다. 따라서 이러한 기준에 따라 한국어 학습자를 선정해야 한다.

(14) 한국어 학습자 분류

a. 언어권(영어권/일본어권/중국어권)[5]

5) 필요에 따라 한국어 학습자의 언어권을 보다 다양하게 할 수 있다. 일본어권, 중국어권, 영어권 학습자들이 가장 많은 수를 차지하고 있기 때문에 자료 확보의 용이함을 위해

b. 학습 단계(1급/2급/3급/4급/5급/6급)[6]
c. 성별(남/여)
d. 연령(10대/2-30대/4-50대/6-70대)
e. 한국 체류 기간(없음/6개월 미만/6개월-1년/1년-2년/2년-3년/3년 이상)
f. 부모의 국적(한국인/외국인)

위와 같은 분류 기준에 따라 한국어 학습자를 균형 있게 선정하여 학습자들의 발화를 녹취해야 할 것이다. 그런데, 학습자들의 말뭉치를 (14)에서 제시한 분류에 따라 균형 있게 확보하기 위해서는 상당한 시간이 소요된다. (14a)~(14f)의 6가지 분류 기준에 따라 학습자들을 분류하면 학습자 집단이 1,728개에 달하기 때문이다. 따라서 일차적으로는 (14a)~(14c)의 세 분류 기준만을 고려해 자료를 확보하는 것이 바람직하겠다. (14a)~(14c)의 세 분류 기준에 따라 학습자들을 분류하면 36개로 나뉜다. 이 36개의 학습자 집단을 대상으로 발화 자료를 녹취하면 된다.

집단별로 발화자를 최대한 많이 확보하는 것이 좋겠으나, 통계적으로 유의미한 자료가 되기 위해서는 그룹별로 적어도 30명 이상의 학습자를 대상으로 음성 자료를 수집해야 한다. 발화자는 국내의 여러 한국어 교육 기관에서 한국어를 공부하고 있는 학습자 중에서 선정할 수 있다.

한편 외국인들에게 한국어 발음을 교육하기 위해서는 한국인의 발화 자료도 필요하다. 한국인 화자의 경우에도 발음에 영향을 미칠 수 있는 변수가 존재한다. 출신 지역이 가장 중요한 변수가 되겠으나 한국어 교육을 위해서는 발화자를 표준어 화자로 한정해야 할 것이다. 이 때, 표준어 화자라고 해서 아나운서의 발화를 대상으로 할 필요는 없다. 아나운서들의 발음은 실제 한국인들이 두루 사용하고 있는 발음이라기보다는 규정상의 표준 발음이기 때문이다. 따라서 외국인들이 실제로 들을 가능성이 가장 높은 발음, 즉 일반인들의 발음이 외국인들에게는 표준 발음이 되어야 한다.

이 외에도 발화자의 성별, 연령 등을 고려할 수 있다. 다양한 용도로 음성 자료로 사용될 것을 고려하면 가능한 다양한 집단의 음성 자료를 확보하는 것이 바람직할 것이다. 그러나 비교 대상이 되는 한국어 학습자 발음이 연령을 고려하지 않았다면 한국인의 발화 역시 연령을 고려할 필요는 없고, 남녀 두 집단의 음성 자료를 수집하면 될 것이다.

이런 기준에 따라 선정된 화자들을 대상으로 실험 문장들을 발화하게 하여 녹음을 실시한다. 이 때, 깨끗한 음질을 유지하기 위해서는 방음 부스에서 디지털 녹음기를 사용하여야 한다.

일단 본고에서는 세 언어권으로만 구분해 보았다.

6) 국내 교육 기관과 한국어 능력 시험에서 한국어 학습 단계를 6단계로 보고 있기 때문에 본고에서도 이에 따라 학습자의 한국어 능력을 6단계로 구분하였다.

4.2. 음성 정보 주석

위에서 제시한 방법에 따라 한국인 화자와 한국어 학습자들의 음성 자료가 확보된 다음에는 이들 자료가 담고 있는 정보들을 이용할 수 있도록 하기 위해서는 이들 정보를 기호화하여 표시해야 한다. 이들 자료에는 반드시 화자 정보가 표시되어야 하며, 문장의 억양 구조, 분절음의 음향적 정보가 포함되어야 한다.

먼저, 화자 정보는 다음과 같은 기준에 따라 표시할 수 있다.

(15) 화자 정보 표시 체계
a. 언어권: 한국인(KOR) / 영어권(ENG) / 일본어권(JAP) / 중국어권(CHI)
b. 학습 단계: 1급(L1) / 2급(L2) / 3급(L3) / 4급(L4) / 5급(L5) / 6급(L6)
c. 성별: 남(M) / 여(F)

물론 (15)에서 제시한 화자 정보 표시 체계 중 한국인의 경우에는 (15a)와 (15c) 정보만 표시하면 되고, 외국인의 경우에는 모든 정보를 표시해야 한다. 위의 표시 체계에 따라 일본인 2급 여성 학습자의 화자 정보는 다음과 같이 표시할 수 있다.

(16) 화자 정보 표시 예(일본인 2급 여성 학습자의 경우)
<head>
<national>JAP</national>
<level>L2</level>
<gender>F</gender>
</head>

다음으로 억양 정보 표시에 대해 살펴보자. 대체로 억양은 문장의 종류에 따라 달리 실현되는 문미 억양, 억양구 경계에서 실현되는 경계 성조, 강세구 억양 등으로 구분된다. 먼저 문장의 종류에 따른 문미 억양은 한국의 발음 교육의 중요한 부분이므로 문장의 종류, 의문문의 종류 등과 같은 통사적인 정보가 필요하다.

(17) 문장 정보 태그
a. 문장의 종류: 서술문(DEC) / 의문문(INT) / 명령문(IMP) / 청유문(PRO)
b. 의문문의 종류: 확인의문문(CHQ) / 의문사의문문(WHQ) / 수사의문문(RHQ)
긍부정의문문(YNQ) / 의문사가 사용된 긍부정의문문(WYN)

또한 발화된 문장의 운율 구조에 따라 억양 유형이 달라지므로, 운율 구조에 대한 정보를 표시해야 한다. 운율 구조 정보에는 운율 단위의 종류뿐만 아니라 하

나의 강세구를 구성하는 음절수와 그 음절이 강세구 내에서 차지하는 위치에 대한 정보도 필요하다.

(18) 운율 구조 정보 태그
a. 운율 단위: 발화 시작(SIL) / 억양구(IP) / 강세구(AP) / 음절(SY)
b. 강세구의 음절수: 1음절 강세구(1AP) / 2음절 강세구(2AP) ……
c. 강세구 내 음절의 위치: 첫 음절(1SY) / 두 번째 음절(2SY) ……

이와 같이 운율 구조에 대한 정보를 표시한 다음에는 억양 실현에 관한 정보를 표시해야 한다. 억양의 유형은 억양구의 경계 성조는 K-ToBI의 기호 체계를 따르되, 자리수를 일정하게 해야 하기 때문에 H%는 HH% L%는 LL%로 표시하기로 한다. 강세구의 억양은 강세구 경계의 성조만을 표시하지 않고 각 음절의 억양을 표시하기로 한다. 한국인의 억양 유형을 기술할 때는 경계 부분의 억양만 표시해도 중간 음절의 억양을 예측할 수 있지만, 외국인들의 경우에는 한국어의 억양 유형과 다른 억양을 실현시키기 때문에 이들의 발화에 나타나는 억양 오류를 정확하게 진단하고 교정하기 위해서는 각 음절의 억양을 표시해야 하기 때문이다.

(19) 억양 정보 태그
a. 억양구의 억양: HH% / L% / HL%
b. 강세구의 억양: 각 음절마다 L이나 H로 표시(1SY_L / 2SY_H ……)

이 외에도 외국인들이 한국어 발화에서 억양의 오류로 지적할 수 있는 것은 강세구와 억양구의 끝음절이 길게 발음하지 않고 짧게 소리내거나, 폐음절을 개음절보다 훨씬 길게 발음한다는 것이다. 따라서 각 음절의 길이도 표시를 해 주어야 한다. 음절의 길이 역시 000ms로 표시할 수 있을 것이다.

다음으로 분절음 정보 표시 체계에 대해 살펴보자. 분절음의 경우에는 어떤 음소를 발음한 것인지를 표시해 주어야 하며, 그 분절음이 실제로 어떤 음으로 발음되었는지를 표시해 주어야 한다. 모음의 경우에는 그 음가를 제1포먼트와 제2포먼트값으로 나타낼 수 있으며, 자음의 경우에는 실제로 실현된 음성형을 IPA 기호로 나타내면 될 것이다. 그러나, 폐쇄음과 파찰음의 경우에는 폐쇄 구간의 길이, VOT의 길이를 표시해야 정확한 발음 오류 진단과 발음 교정이 가능하다. 한편 어떤 분절음을 삽입하거나 어떤 분절음을 탈락시켜 발음했는지도 표시해 주어야 한다.

(21) 분절음 정보 태그
a. 음소 표시
- 모음: 아(a) / 에,애(e) / 이(i) / 오(o) / 우(u) / 으(U) / 어(v)

- 활음: j / w
- 자음: ㅂ (b) / ㅍ (p) / ㅃ (B) / ㄷ (d) / ㅌ (t) / ㄸ (D) / ㄱ (g) / ㅋ (k) / ㄲ (G) / ㅅ (s) / ㅆ (S) / ㅎ (h) / ㅈ (z) / ㅊ (c) / ㅉ (Z) / ㅁ (m) / ㄴ (n) / ㅇ (N) / ㄹ (l)

b. 음성 표시

- 삽입 및 탈락: 삽입(INS) / 탈락(DEL)
- 음성형: 모음 포먼트 값 표시 (a_1F0500_2F1500)
  자음 IPA 기호 사용 (l_l / l_ɾ / l_ɬ ……)
- 폐쇄 구간 및 VOT 길이: 040ms / 053ms / ……

이 장에서는 한국어 교육용 음성 DB를 구축하는 방법을 살펴보았다. 구체적으로는 음성 자료를 확보하는 방법, 자료에 학습자 정보, 음성 정보 등을 표시할 때 사용되는 태그에 대해 살펴보았다.

5. 결 론

본고에서는 한국어 학습자의 발음을 경제적이면서도 효과적으로 교정하기 위해 한국어 발음 교육 프로그램 개발과 이를 위한 한국어 교육용 음성 DB 구축의 필요성을 제기하고, 그 구체적인 방안에 대해 살펴보았다. 이를 위해서 먼저 한국어 학습자들의 발음 오류 유형을 분석하고, 이를 토대로 음성 DB 구축에 필요한 실험 문장들을 고안하였다. 다음으로 이 자료를 이용한 음성 DB 구축 방안으로 자료 수집 방법, 정보 표시 체계를 제시하였다.

필자는 본 연구의 후속 작업으로 본고에서 제안한 방법에 따라 실제로 한국어 발음 교육용 음성 DB를 구축하고자 한다. 그리하여 실제로 한국어 발음 교육용 음성 DB가 구축된다면 한국어 발음 교육 프로그램 개발 과정과 실제 한국어 교육 현장에서 유용하게 사용될 수 있을 것이다.

참 고 문 헌

[1] 고석주, "한국어 학습자 말뭉치 구축의 현황", *한국어 교육과 학습자 말뭉치-제3회 한국어교육 국제워크숍 발표논문집*, pp.1-33. 2003.

[2] 김유미, "외국어로서의 한국어 학습자 말뭉치를 이용한 오류 분석", 연세대 교육대학원 석사학위 논문, 2000.

[3] 김정숙, 김유정, "한국어 학습자 말뭉치 구축을 위한 기초 연구", *이중언어학회지*, 21호, pp.97-119, 2002.

[4] 서상규, 유현경, 남윤진, "한국어 학습자 말뭉치와 한국어 교육", *한국어 교육*, 13권, 1호, pp.127-156, 2002.

[5] 유석훈, "외국어로서의 한국어 학습자 말뭉치 구축의 필요성과 자료 분석", *한국어 교육*, 12권, 1호, pp.165-180, 2001.

[6] 이정희, "학습자 말뭉치를 이용한 외국어로서의 한국어 교육 방법", *한국어 교육과 학습자 말뭉치-제3회 한국어교육 국제워크숍 발표논문집*, pp.91-96, 2003.

[7] 신지영, "음성 자료의 수집 · 조사 방법론", *국립국어연구원 심포지엄 자료*, 2001.

[8] 김정숙, "일본인의 한국어 학습 시 나타나는 음운론적 오류 분석", *韓國語學新硏究*, pp.63-75, 1988.

[9] 공일수, "한국어와 아랍어 음운론적 대조와 ERROR 분석", *이중언어학회지*, 9호, pp.162-192, 1992.

[10] 권현주, "일본인 학습자를 대상으로 한 한국어 발음 교수법 연구 -대조언어학적 방법을 통하여-", 이화여대 교육대학원 석사학위 논문, 1995.

[11] 김영아, "재중교포들의 한국어 학습 시 나타나는 오류 분석", *이중언어학회지*, 7호, pp.256-263, 1990.

[12] 맹주억, "韓中 中間言語의 硏究: 音韻論을 中心으로", 한국외국어대 박사논문, 1992.

[13] 박병채, "이중언어 교육상의 음운론적 대응 분석 - 한국어 교육을 중심으로-", *이중언어학회지*, 8호, pp.553-563, 1991.

[14] 우인혜, "일본인 학습자가 곤란을 겪는 한국어 음", *국제한국어교육학회 제11차 국제학술회의 발표논문집*, pp.287-295, 2001.

[15] 이경희, 정명숙, "일본인을 위한 한국어 파열음 발음 및 인지 교육", *한국어교육*, 10권, 2호, pp.233-255, 2000.

[16] 정명숙, 이경희, "학습자 모국어의 변이음 정보를 이용한 한국어 발음 교육의 효과 - 일본인 학습자를 대상으로 - ", *한국어교육*, 11권, 2호, pp.151-167, 2001.

[17] 이향, "중국어권 학습자들의 한국어 발음실태", *제14차 이중언어학회 국내 학술대회 발표논문집*, pp.125-138, 2002.

[18] 이호영, "한국어 억양의 교육", *제2차 한국어세계화 국제학술대회 발표논문집*, pp.290-300, 2001.

[19] 정명숙, "한국어 억양의 기본 유형과 교육 방안", *한국어교육*, 13권, 1호, pp.225-241, 2002.

[20] 정명숙, "일본인과 중국인의 한국어 억양", *한국어교육*, 14권, 1호, pp.233-247, 2003.

[21] 한재영, "한국어 발음 교육 현황과 문제점", *제2차 한국어세계화 국제학술대회 발표논문집*, pp.316-332, 2001.

[22] 신지영, 차재은, *우리말 소리의 체계*, 한국문화사, 2003.

[23] 정명숙, "현대 국어 말소리의 통시적 변화", 고려대 박사학위 논문, 2002.
[24] 차재은, 정명숙, 신지영, "공명음 사이의 /ㅎ/의 실현에 대한 음성, 음운론적 고찰", *대한음성학회 봄 학술대회 발표 논문집*, pp.48-51, 2003.

접수일자: 2003년 8월 30일
게재결정: 2003년 9월 17일

▶ 정명숙(Myungsook Jung)
주소: 136-701 서울시 성북구 안암동 5가 1-2번지
소속: 고려대학교 국제어학원 한국어문화교육센터
전화: 02) 3290-1549
FAX: 02) 921-0514
E-mail: phono@chollian.net