

# 한국인의 영어 음성코퍼스 설계 및 구축<sup>+</sup>

이석재\*, 이숙향\*\*, 강석근\*\*, 이용주\*\*\*

연세대학교 영어영문학과\*, 원광대학교 영어영문학과\*\*, 원광대학교 전기전자 및 정보공학부\*\*\*

## Design and Construction of Korean-Spoken English Corpus (K-SEC)

Seok-Chae Rhee\*, Sook-Hyang Lee\*\*, Seok-keun Kang\*\*, Yong-Ju Lee\*\*\*

Dept. of English Language and Literature, Yonsei University\*,

English Department, Wonkwang University\*\*,

Dept. of Computer Engineering, Wonkwang University\*\*\*

E-mail : scrhee@yonsei.ac.kr\*, shlee@wonkwang.ac.kr\*\*,

skkang@wonkwang.ac.kr\*\*, yjlee@wonkwang.ac.kr\*\*\*

### Abstract

K-SEC(Korean-Spoken English Corpus) is a kind of speech database that is being under construction by the authors of this paper. This article discusses the needs of the K-SEC from various academic disciplines and industrial circles, and it introduces the characteristics of the K-SEC design, its catalogues and contents of the recorded database, exemplifying what are being considered from both Korean and English languages' phonetics and phonologies. The K-SEC can be marked as a beginning of a parallel speech corpus, and it is suggested that a similar corpus should be enlarged for the future advancements of the experimental phonetics and the speech information technology.

### I. 목적

‘한국인의 영어발음 음성코퍼스(Korean-Spoken English Corpus, K-SEC)’는 “한국인에 의해 발화되는 영어음의 디지털 녹음 자료로서 컴퓨터를 통해 재구현될 수 있고, 주석과 그 밖의 자료를 동반하여 체계적으로 정리된 음성 자료”를 뜻하는 것으로서, 본 연구는 다양한 대상 기준에 따른 한국인의 영어발음 음성을 대용량으로 녹음한 후 데이터베이스화하여 실험음성학과 음운론, 그리고 영어교육과 음성기술산업의 미래

발전을 위한 기초적 자료와 연구 대상을 제공하기 위한 음성코퍼스의 설계 및 그 구축 과정을 보여주고자 한다.

### II. 배경 및 K-SEC의 필요성

실험음성학 분야에서 한국인의 영어 발화시 한국어와 영어의 음성·음운의 차이에 의한 간섭 현상의 연구는 비교 실험음성학의 이론적 발전에 기여할 수 있는 중요한 연구과제이다. 그러나 현 상황에서 상기한 목적을 위하여 실험음성학 학자들이 ‘공동’으로 사용할 수 있는 음성코퍼스의 존재가 미미하여 각자의 연구 목적에 따라 소용량의 음성을 녹음하여 연구하고 있으나, 이는 연구 결과의 객관성에 문제가 생길 수 있으며 학계 전체적인 차원에서 보았을 때도 비경제적인 일이다. 이러한 상황에서 다양한 기준에 따른 대용량의 한국인 영어발음 음성코퍼스의 구축과 그 결과물의 공유는 실험음성학의 공통적 자료 기반을 마련한다.

현재 한국 영어교육의 틀은 독해 중심의 영어에서 말하기 중심의 영어로 큰 틀이 변화하고 있지만, 영어 말하기에서 중요한 발음교육은 어느 발음에 문제가 있는지 구체적으로 정확히 진단을 하지 않은 상태에서 교육이 실행되는 형태를 보이고 있다. 즉 영어교육 현장에서 영어발음에 대한 진단 및 교정이 거의 피상적인 인상에 의거하여 이루어지고 있어, 현재 행하여지고 있는 발음교육의 효과가 의문스러운 상태이다. 이러한 상황에서 본 연구가 추진하고자 하는 한국인의

영어발음 음성코퍼스는 한국인에게 나타나는 영어발음의 문제점을 구체적이고 객관적으로 진단해 줄 수 있는 자료를 제공하여 효과적인 영어발음 교육의 출발점이 된다. 또한 본 연구는 영어발음의 진단 차원을 넘어 영어발음 음성코퍼스를 여러 기준에 따라 구축함으로써 지금까지 있어 왔던 영어발음 교육에 대하여 지역간, 성별, 연령, 영어교육의 정도 등에 따른 향후 영어교육정책의 자료적 기반을 제시하고, 앞으로 그 필요성이 절실하게 될 '영어발음 자동 평가안'의 기초 자료 및 기준점을 제시하는 방안이 된다.

이에 덧붙여 여러 기준에 따른 한국인의 대용량 영어발음 음성코퍼스는 사회언어학적 그리고 언어심리학적 연구의 심층적 연구에도 활용될 수 있는 자료를 제공하게 될 것이며, 장기적으로 연구 수행 과정에서 얻게 될 기술 및 그 결과는 컴퓨터와 인간 사이의 자연어에 의한 의사소통을 다루게 될 미래 언어정보학의 발전에 초석이 될 것이다. 특히, 본 연구를 수행하는 과정에서 익히게 될 기술 및 그 결과는 음성정보기술(Speech Information Technology, SIT) 산업에도 중요한 자료로 사용될 수 있다. 현재 전세계적으로 컴퓨터와 인간의 의사소통을 음성으로 구현하려는 추세이고, 그 응용 분야는 산업쪽에서 무한한 가능성을 보이는 바, 한국인의 영어발음 음성코퍼스는 한국인이 영어를 말해야 하는 상황이 요구되는 컴퓨터 개발에 있어 음성인식과 관련되는 기술 그리고 STILL(Speech Technology in Language Learning) 분야 영어학습 소프트웨어 개발에 필수적인 기초 자료이다.

### III. 설계

녹음전 설계에서는 발성목록과 발성화자의 형태 및 수를 결정한다. K-SEC 발화목록 설계는 다음의 기준에 의하여 설계되었다.

- 낭독체 발화(read speech)를 녹음한다.
- 어휘음성(독립어)과 연속음성(문장 및 이야기)을 모두 녹음한다.
- 정형음성을 녹음한다.
- 어휘음성은 PBW(Phoneme Balanced Words)가 되도록 한다.
- 어휘음성은 초등, 중등, 일반인으로 구분한다.
- 연속음성은 초등, 중등, 일반인의 구분을 두지 않는다.
- 어휘 및 연속음성 모두에서 분절적 요소와 초분절적 요소를 모두 고려한다.
- 해외 음성DB와의 연계성을 갖도록 설계한다.

K-SEC의 발성화자는 연령별로는 3구분(초등학생 5, 6학년/고등학생/일반성인), 지역별로는 7구분(서울/경기/강원/충청/전라/경상/제주), 성별(남/여)로는 2구분이며, 각 구분에 따른 화자 인원은 8명이다. 이에 따라 한국인 총 336명의 발성화자가 K-SEC 녹음에 참여하였다.

(3연령 x 7지역 x 2성별 = 42구분) x 구분당 8인 = 총 336명

이러한 설계에 따른 결과의 분석은 사회언어학적인 특성을 보여 줄 것이고, 영어교육의 정도에 따른 발음의 양상도 보여줄 것이며, 남성·여성별로도 그 차이가 명시적으로 밝혀질 것이다.

한편 추후 대조 비교를 목적으로 하여 영어 원어민도 2성별, 3연령으로 구분하고 각 구분당 2명씩 녹음하였다(원어민 성인 남녀에서는 지역 구분도 고려하여 NAE(North American English), BE(British English), AE(Australian English)로 지역 구분을 두었다).

발성목록은 영어의 분절적인 면과 초분절적인 요소의 특징을 복합적으로 나타낼 수 있는 어휘, 문장 그리고 이야기 읽기로 설계되었다. 발성목록의 문장이나 이야기 읽기는 아래 영어의 음성 음운론적 특징 및 기존의 연구에 의해 예견되는 한국어 간섭에 의한 영어 오류 발화를 보여줄 어휘나 어휘의 연결체로 구성되도록 고려되었다.

- 순음성 폐쇄음과 마찰음의 구분
- 저해음 유·무성의 구분
- 저해음 유·무성의 차이와 앞 모음 길이의 상관성
- 단설음화의 양상
- [s] 다음에 폐쇄음의 기음(aspiration) 정도
- 단어 내 위치에 따른 무성파열음의 기음 정도
- 파열음의 단어 내 위치에 따른 파열 유·무와 정도
- 치경과 경구개치경 마찰음과 파찰음의 구분 정도
- 저해음의 부분적 무성음화의 정도
- 연음의 양상
- 원순음화
- 자음-비음 연결체
- 모음 사이의 무성폐쇄음 발음
- 공명음 [n], [r], [l]의 구분
- 축약모음(schwa)의 양상
- 리듬
- 억양
- 휴지

K-SEC은 한국인의 영어발화 음성 DB이지만 각 동

일 화자의 모국어 L1 발화 때와 외국어 L2 발화 때의 음성 비교를 위하여 기본적인 한국어 발화의 녹음도 포함되어 있다.

상기의 원칙 및 고려 요인을 바탕으로 발화목록은 총 6개의 SET로 구성되어 있다.

- SET#1: 한국어 독립어(실제어 및 가상어, 일부는 이끔 문장(carrier phrase)으로 시작)
- SET#2: 한국어 이야기
- SET#3: 영어 자음 및 모음 관찰용 독립어(일부는 이끔 문장 포함)
- SET#4: 영어 독립어 어휘(초등, 중등, 일반인용으로 구분)
- SET#5: 영어 문장
- SET#6: 영어 이야기

SET#1은 한국어 자모의 관찰 및 영어와의 대조 비교를 위하여 설계되었다. 초성에 올 수 있는 한국어 자음 18개를 틀형 [Cada]에 대입하고, 단모음 7개를 틀형 [hVda]에 넣었으며, 경음, 평음 치경마찰음을 전고설 모음 앞에 위치시킨 후, 이렇게 모아진 모든 어형을 2회씩 반복하여 총 목록을 작성하고 이것을 컴퓨터 소프트웨어를 사용하여 무작위화(randomized)하였다. 또한 각 자음의 특징을 모음 사이에서 관찰하기 위하여 틀형 [Cal]에 각 자음을 대입하고, 모음 사이의 유음 관찰을 위하여 추가적으로 [가라], [자라], [타라]를 추가 한 후, 수집된 (실제 및 가상) 어형을 2회씩 반복하여 총 목록을 작성하고 이것을 다시 무작위화한 후 이끔 문장(carrier phrase)에 넣었다. 이 결과가 SET#1이다. 이는 동일화자 발화에 있어 한국어와 영어의 자모 비교를 위한 것이고, 녹음 자료는 한·영 발화 대조연구는 물론 한국어 자체의 실험음성학적 연구에도 사용될 것이다.

SET#1 한국어 독립어 목록

바다, 흐다, 까다, 헤다, 나다, 카다, 라다, 췌다, 가다, 파다, 자다, 식다, 하다, 마다, 후다, 다다, 허다, 빠다, 싸다, 차다, 짜다, 호다, 따다, 사다, 타다, 히다, 아다, 까다, 호다, 파다, 마다, 후다, 췌다, 바다, 흐다, 빠다, 허다, 카다, 라다, 나다, 싸다, 다다, 하다, 자다, 헤다, 차다, 식다, 사다, 히다, 아다, 타다, 따다, 짜다, 가다, 이것도 탈이다; 이것도 말이다; 이것도 잘이다; 이것도 밭이다; 이것도 팔이다; 이것도 랄이다; 이것도 찰이다; 이것도 빨이다; 이것도 갈이다; 이것도 잘이다; 이것도 달이다; 이것도 깔이다; 이것도 가라다; 이것도 딸이다; 이것도 타라다; 이것도 짤이다; 이것도 살이다; 이것도 딸이다; 이것도 타라라; 이것도 찰이다; 이

것도 알이다; 이것도 쌀이다; 이것도 탈이다; 이것도 날이다; 이것도 팔이다; 이것도 할이다; 이것도 가라다; 이것도 달이다; 이것도 자라다; 이것도 칼이다; 이것도 알이다; 이것도 찰이다; 이것도 살이다; 이것도 깔이다; 이것도 갈이다; 이것도 날이다; 이것도 쌀이다; 이것도 랄이다; 이것도 할이다; 이것도 칼이다; 이것도 밭이다; 이것도 빨이다; 이것도 말이다; 이것도 자라다.

SET#2는 한국어 이야기 읽기로서 연속발화에 있어 한국어 자모의 관찰 및 영어와의 대조 비교를 위하여 설계되었는데, 특히 연속발화에 있어 동일 화자의 한국어와 영어의 운율(prosody) 차이 특성 연구를 위하여 도입되었다. 이야기 '바람과 햇님'이 선택된 이유는 이에 대한 기존의 음성DB가 있기 때문에 이를 더 많은 그리고 다양한 화자를 통하여 DB의 용량을 대용량화하기 위해서이다. 같은 내용의 영어 이야기 읽기는 SET#6에 제시되어 있다.

SET#2 한국어 이야기

바람과 햇님이 서로 힘이 더 세다고 다투고 있을 때, 한 나그네가 따뜻한 외투를 입고 걸어왔습니다. 그들은 누구든지 나그네의 외투를 먼저 벗기는 이가 힘이 더 세다고하기로 결정했습니다. 북풍은 힘껏 불었으나 불면 불수록 나그네는 외투를 단단히 여몄습니다. 이 때 햇님이 뜨거운 햇빛을 가만히 내려쬐니 나그네는 외투를 얼른 벗었습니다. 이리하여 북풍은 햇님이 힘이 더 세다고 인정하지 않을 수 없었습니다.

SET#3은 동일 환경 하에서의 영어 각 자음 및 모음의 특징을 관찰하기 위하여 설계되었다. 우선 영어 독립어 모음 관찰을 목적으로 [hVd] 틀형에 영어 모음 10개를 대입하여 리스트화한 후, 2회 반복하여 무작위화 하였고, 자음 관찰을 위해서는 Ladefoged(2001)가 제시한 최소대립쌍(minimal pairs)을 차용하여 2회 반복한 후 무작위화 시키고, 이를 이끔 문장 내에 두었다. SET#3의 문장은 영어의 대조적 강세에 따른 각 자모음의 특징 차이를 알아보기 위하여 선택되었다. 여기서 영어 모음 관찰을 위한 어휘는 Peterson & Barney(1952)로부터 차용되었고, 영어 대조적 강조 문장은 Turk & White(1999)로부터 차용되었다.

SET#3 동일 환경 영어자모, 문장

heard, had, heed, hod, heard, hud, hid, hood, had, who'd, hud, head, hod, hid, hawed, head, hood, who'd, heed, hawed, Say pie, too; Say vie, too; Say my, too; Say thy, too; Say lie, too; Say kye, too;

Say by, too; Say chime, too; Say guy, too; Say fie, too; Say why, too; Say sigh, too; Say tie, too; Say nigh, too; Say by, too; Say lie, too; Say shy, too; Say dye, too; Say shy, too; Say rye, too; Say dye, too; Say jive, too; Say pie, too; Say nigh, too; Say vie, too; Say fie, too; Say kye, too; Say thy, too; Say my, too; Say thigh, too; Say guy, too; Say high, too; Say jive, too; Say sigh, too; Say rye, too; Say thigh, too; Say tie, too; Say why, too; Say high, too; Say chime, too,

I said "BUY cakes", not MY cakes".

I said "buy CAKES", not "buy COKES".

I said "BIKE aches", not "MIKE aches".

I said "bike ACHES", not "bike APES".

SET#4는 초등, 중등, 일반인용으로 단어 수준에 따라 3분화 되어 있다. 3분화의 가장 큰 이유는 동일 목록을 초등, 중등, 일반인에게 읽힐 경우, 초등학생들에게 상당히 어려운 어휘가 포함될 수 밖에 없어, 정상적인 발화를 할 수 없을 것으로 예측되었고, 실제 시험 녹음에서도 우려했던 바가 그대로 나타나 연구 취지에 맞는 녹음을 얻을 수 없었기 때문이다. 따라서 교육부 고시 제 1997-15호 [별책 1]인 초·중등학교 교육 과정(국민 공통 기본 교육 과정)과 American Education Publishing사의 *Comprehensive Curriculum of Basic Skills*을 바탕으로 각 수준별 어휘 분류 작업을 기초로 연구진의 판단에 따라 어휘의 수준을 초등용, 중등용, 일반용으로 구분하였다. 한편 각 수준별 목록에서 어휘들은 PBW가 될 수 있도록 노력하였다. 구성은 24개의 영어 자음과 21개의 모음(이중, 삼중모음 및 rhotacized 모음 포함)을 조합하여, i) 각 모음 앞에서 각 자음의 특성, ii) 각 모음 뒤에서 각 자음의 특성, iii) 강세가 없는 모음과 강세가 있는 각 모음 사이에서 각 자음의 특성, iv) 강세가 있는 각 모음과 강세가 없는 모음 사이에서 각 자음의 특성(즉, CV, VC, VCV(+stress), V(+stress)CV)이 한국인의 영어 독립어 발화에서 어떠한 양상을 보이며 원어민의 발화와 어느 정도 차이가 나는지를 실험음성학적으로 분석하기 위하여 설정된 것이다. 해당 특성을 보이는 어휘는 우선 초등 어휘에서 선택하고, 초등 수준에서 발견되지 않는 것은 중등 어휘에서, 다시 중등 수준에서 발견되지 않는 것은 일반인 수준에서 선택되었다. 한편 중등과 일반인 발화목록에서는 SET#4에 대한 보강 어휘가 첨가되었는데 해외 음성 DB와의 연계성을 고려하여 M. Shimuzu, Tomiyama, Dantsuji & K. Shimuzu에서 밝히 Extended Hughes and Trudgill's Word List(EWL, 2002) 중에서 원 SET#4에 포함되지 못한 어휘를 추가

로 도입하여, 추후 미국의 음성DB 및 일본인의 영어발음 음성 DB와의 비교 분석을 위하여 설계되었다.<sup>1)</sup> 여기에 덧붙여 어두 자음군의 발음 양상을 보여주는 어휘가 추가되었고, 비음과 마찰음 사이에 폐쇄음 첨가 현상(예를 들면, print/prince), 강세와 음절 수 및 모음 길이의 상관관계를 보여주는 어휘들(예를 들면, speed/speedy/speedily), 성문폐쇄음의 발화 양상을 보여주는 어휘들이 포함되어 있다(예를 들면, button). 초등용 어휘는 총 435개, 중등용 어휘는 총 956개(초등 435개 포함), 일반인용 어휘는 총 1125개(중등용 어휘 956개 포함)를 수집하여 각각을 목록으로 만들고 이를 다시 무작위화하여 각 수준과 환경에 따른 각 소리의 특징을 최대로 보여주는 발성 목록을 완성하였다.

#### SET#4 성인(대학생, 대학원생, 일반)

hobby, keep, heat, jam, dig, soap, mouse, tobacco, map, dining, pork, cook, dark, quiz, so, rethink, follow, loiter, account, weekend, shall, have, gall, jibe, washer, lair, silence, eager, lady, sang, hot, demand, fit, woke, remark, cure, cow, boy, racing, mail, heard, cuter, repay, June, worthy, thank, ball, gorilla, hook, part, fear, refine, car, pot, viewless, Paul, Yale, alike, navy, rehearsal, sick, hurt, ken, hut, firepower, brook, rouge, spider, prohibit, rubber, vogue, loathe, arrange, duke, author, such, clean, beggar, fearsome, again, reward, word, parse, voicing, jail, remind, avoid, term, meeting, thorny, enable, reluctant, lodge, moment, doorway, mailing, teacher, lamb, valley, fire, footer, visit, hit, remember, rewire, exchange, make, shock, check, bear, recheck, abort, remove, truth, safe, zeal, deaf, spoon, puff, abed, babe, juice, mirror, lake, morpheme, vacation, avow, yet, fairly, rob, dire, animal, Sammy, idea, summer, lip, nothing, dupe, within, rare, shirt, typist, them, hued, individual,

1) 일본에서 제작된 Hughes and Trudgill의 확장 어휘 리스트(EWL)에 포함된 어휘는 다음과 같다: pit, pet, pat, put, putt, pot, bee, bay, buy, boy, boot, boat, bout, beer, bird, bard, board, city, seedy, hat, dance, daft, half, father, farther, pull, pool, pole, Paul, doll, cot, caught, fir, fern, fur, fair, nose, knows, plate, weight, poor, pour, pore, paw, tide, tied, pause, paws, meet, meat, mate, choke, joke, day, gay, pay, tea, key, see, she, he, fee, vehicle, theory, thee, moon, noon, zone, measure, lay, ray, way (이상 72어휘).

apart, with, retouch, sheep, naught, medical, fun, desk, wiretap, bare, soccer, reshuffle, carbon, talk, peace, shake, bribery, chalk, south, zip, pause, law, bomb, putt, liar, woman, research, review, bowling, jeer, void, if, retort, pull, ear, boycott, hers, Joyce, name, tool, deer, kidding, assume, frog, webbing, should, reload, pose, ashamed, god, beach, appoint, birth, gun, drowse, event, alone, vine, court, siren, orbit, bore, party, avert, supper, more, fool, cartoon, arrive, which, stove, defocus, redial, much, mud, mall, lesson, soil, balloon, fussy, goose, bug, this, ocean, tie, join, beat, pound, allow, bathe, design, date, mommy, door, safer, March, zed, assort, room, arm, renew, tide, down, return, raw, upon, beef, sit, batch, ago, shore, vegetable, tune, cookie, chop, hopper, city, league, icy, earwax, mount, life, theft, dance, reheat, kicking, curfew, believe, please, youth, defend, engineer, fourteen, Asian, thorn, burp, harm, order, kind, cheese, form, full, emotion, nearly, lucky, merge, veal, cupid, bark, vent, bag, hat, strawberry, possible, sad, future, view, jean, cab, disease, war, bed, pearl, tight, hate, porch, fork, weird, pepper, bus, vase, there, proofing, afar, jacket, home, hoist, toilet, Peru, either, alert, native, rejoice, feel, yacht, occur, out, push, give, chore, feeb, jaw, read, announce, achieve, collect, shark, because, avoiding, liberty, earring, recent, keen, cloud, large, scarf, roar, machine, revolve, what, dull, rising, Vick, pole, orb, retard, evolution, tongue, voice, assist, rather, affair, greyhound, peak, resist, high, watcher, you, they, live, toyman, took, need, correct, chess, adore, bazaar, vehicle, few, murmur, guy, tall, harsh, work, dead, attach, pick, buzz, heart, rabbit, eardrum, chip, harsher, airtight, searching, talking, deface, caught, mat, sufficient, orchard, long, foggy, geese, van, hour, even, matching, passenger, good, cut, nark, low, yarn, dirt, five, theirs, her, wood, reboot, teeth, student, noise, smoother, assault, face, attune, thought, human, ashore, tired, joke, comb, year, fall, item, feeble, letter, canoe, fine, beam, thin, time, hanger, obey, rose, puzzle, raccoon, sugar, saucer, toy, soybean, begun, ivy, lock, noon, accord, nobody, leave, card, whore, yearn, top, burner, primary, take, town, purse, coil, won, doll, both, adapt, awake, soon, vivid, retold, remain, ridiculous, serve, then, bone, cross, during, smooth, poise,

defeat, dove, feet, red, version, York, mate, cup, alloy, do, zoom, off, watch, deep, poor, tonight, fowl, model, noun, repair, lap, illusion, save, refrigerator, barn, cap, wore, spire, broad, dog, stupid, casual, today, jazz, amuse, peer, cop, size, nature, maze, rate, burger, tooth, loose, dare, house, my, hero, fishing, immediate, turkey, huge, method, rerun, rock, moan, saw, nose, feeling, jump, bad, cease, curb, connect, odd, wax, frugal, reject, gut, late, hotel, shawl, burn, search, fat, accuse, without, lemon, note, silly, murk, thigh, afford, fruit, moist, repeat, financial, notice, faith, horn, mouth, service, cork, far, appear, base, that, seedy, forbid, pattern, finish, descend, meet, pool, rich, along, hiking, chow, sketching, choice, open, hose, vision, tiresome, ill, mook, noisy, potato, female, mad, web, lazy, bathing, organ, where, army, use, swim, baby, gearbox, although, farewell, moon, color, murder, reveal, adult, toad, force, purple, big, drip, tiers, Carl, sharp, tier, tires, wet, carve, usher, mute, cake, though, haired, earthy, recant, behind, Yahoo, rifle, never, fate, Mars, George, cloudy, duck, associate, put, pipe, leisure, major, glove, nook, ruby, forty, sir, beard, lawyer, call, let, firehouse, proof, devote, spooky, resume, cooling, mare, loud, surf, happy, honor, ton, rub, amount, sharpen, voucher, pie, cute, tart, die, mother, depend, couch, shout, rear, coaching, set, refund, avail, cord, turn, move, recharge, news, giant, fence, young, recoil, abode, lark, desire, become, zap, flashing, singer, having, moth, special, loading, beware, kid, roof, lie, foot, boot, foil, made, refer, tire, zone, republic, abuse, gym, falling, region, sooner, hen, pass, laughing, closing, cock, meat, back, market, net, forget, share, barley, oil, relax, north, leg, rain, hurry, thief, king, famous, paper, smile, zoo, edge, reunion, soup, difficult, not, two, gear, rebuking, fume, shy, wait, adopt, toothache, dirty, show, bit, math, devise, core, limit, loving, report, assassin, merry, bib, drowning, fan, reef, forbear, rehear, kier, badge, hall, mule, daughter, bird, page, herb, lore, season, earphone, week, vow, laugh, nice, smoke, tap, afire, cold, myth, piss, repute, gaining, suppose, some, pat, among, care, particular, chair, Korean, farther, doze, tore, night, sheer, nor, begin, around, evaluate, dorsal, win, fresh, pudding, mercy, nephew, second, boat, mark, coin, third, nap, began,

denote, airforce, lead, cat, cube, tin, step, doors, harpoon, chat, rush, chart, breathe, ship, cutter, retire, academic, touching, Nick, ten, carry, hood, vote, guard, about, career, job, pig, tiger, piggy, same, refuse, deck, hire, length, shut, pop, rice, child, delay, regard, kiss, sue, learn, melon, fuller, met, whose, loser, movie, perhaps, decode, annoy, sketch, pet, Japan, cool, cowboy, thumb, tireless, sock, token, head, bridge, early, gown, why, abide, fir, bun, possess, marching, hair, shoot, beside, muffin, mature, pushing, round, run, baggage, sore, wire, kitchen, death, shape, judge, allure, using, gap, dish, coach, thus, globe, team, look, hugging, wow, adjust, audience, food, fin, Zeus, neck, iron, target, wash, choke, wrap, meaning, gate, offer, judging, Gore, get, dripping, kissing, herbal, over, crown, small, Arthur, awoke, reduce, bother, curse, prose, harder, shooting, rewind, seen, ahead, beyond, parser, measure, here, dew, sun, miss, nurse, cash, fox, bonus, magic, fair, novel, foe, study, bottom, buy, loaf, Cathy, merger, jet, fireman, hop, jar, fireball, result, verb, doubting, but, paid, dormitory, hard, love, book, cheer, these, marvel, Irwin, daddy, foul, pure, akin, beer, keeper, delicious, receive, stair, point, above, like, church, Roy, change, muse, mock, sorry, buyer, road, bell, firework, box, lose, near, shy, beige, usual, sigh, yawn, came, thousand, choose, Barney, fight, superb, together, behave, loosen, goat, sofa, money, nut, lawn, larger, rumor, desert, many, girl, maker, pair, shelf, airport, key, snake, bay, he, see, paw, prince, cry, lay, fee, paws, father, ray, speedy, fur, button, plate, raise, weight, race, sleep, quiet, twin, vie, daft, bee, pit, strike, nigh, kye, pour, way, tied, tense, fie, prints, she, scream, thee, cot, speed, spray, try, gay, board, pore, speedily, half, day, knows, bard, tents, rye, bout, fern, theory, pay, tea

SET#5는 영어 문장 연속 발화 목록으로, 목적은 i) 한국인의 영어 문장 발화에서 억양과 리듬을 보기 위하여 다양한 억양과 리듬 패턴이 들어가도록 설계되었고(예를 들면, 평서문, Wh-Q 의문문, YES-NO Q 의 문문, 부가의문형, 선택의문형, 강조구문, 감탄문 등), ii) 단어와 단어 사이에서 연음의 정도와, iii) 한국인에게 잘 발견되는 분절음 발화 상의 오류(예를 들면, 유성음 사이의 오류적 유성음화, 오류적 비음화 등), iv) SET#3과 SET#4에서 이미 관찰했던 현상들이 문장에

서는 어떠한 양상으로 나타날지를 고려하며 설계되었다. SET#5는 초등, 중등; 일반 공용으로 초등생에게 어렵지 않도록 어휘의 선택에도 유념하며 설계되었다.

#### SET#5 영어 문장 읽기

Miss Henry drank a cup of coffee.  
 What are you looking for?  
 Put your toys away right now.  
 The dancing queen likes only the apple pies.  
 I closed the door and waited for the bus.  
 Hit the ball with this bat.  
 Thirteen years later, Mary met him at the same place.  
 Jenny walked home from school in the rain.  
 Raise your right hand, if the teacher calls your name.  
 Nothing can stop me now.  
 How long does it take to the next bus stop?  
 My brother is coming on Friday.  
 I believe the class begins at ten.  
 The cats should have eaten the hotdog.  
 People couldn't sleep well last night because of the noise.  
 Elizabeth wanted to be a judge.  
 Children like candy very much.  
 It's my sister who talked to the kid.  
 The police took the cab to Seoul.  
 What a surprise!  
 Is the turtle alive?  
 I have friends who are just like me.  
 I see the moon shining brightly all over the lake.  
 If a tree could talk, what would it say?  
 Did he fail the test again?  
 The students are going to the zoo.  
 Call me tonight around seven.  
 What's the weather like today, sunny or cloudy?  
 You like orange juice, don't you?  
 The players sent the cap to Susan.  
 Heat the water to thirty degrees.  
 I can't remember the scene vividly.  
 When did he go skiing, Wednesday or Thursday?  
 We went to London, Paris, Cairo, and Boston.  
 Where do you think you left it?  
 They suspect that the suspect killed Ted.

SET#6는 영어 이야기 읽기를 위한 것으로, (문장 단위를 넘어) 발화되는 이야기 읽기에서 다양한 운율적

특징을 관찰하기 위하여 도입되었다.

#### SET#6 영어 이야기 읽기

*The Wind and the Sun were discussing who was the stronger of the two. Since they could not agree, they decided to test their strength. They saw a traveler walking on the road and the Sun said, "Whoever can make him take off his coat will be judged the strongest. You go first." So the Wind blew as hard as he could, but no matter what he did, he couldn't make the traveler take off his coat. Instead, the traveler wrapped his coat more tightly around himself. The Wind finally gave up. Now it was the Sun's turn. The Sun came out and began to shine on the traveler. At first the traveler unbuttoned his coat. Then, as the air grew warmer and warmer, the traveler took off his coat altogether. The Sun turned to the Wind and said, "You see, persuasion is better than force."*

### IV. 녹음

#### 4.1. 녹음 시설 및 장비

녹음은 방음시설을 갖춘 스튜디오에서 실시함을 원칙으로 하여, 서울 경기 지역은 연세대학교 정보통신처 방음실에서, 지방 화자 발화는 사전 섭의를 통하여 각 지역 대학교 방음실을 이용하였다(강원: 연세대학교 원주캠퍼스 방음실 방음실, 충청 및 전라: 원광대학교 음성정보기술산업지원센터 방음실, 경상은 부산해양대학교 방음실, 제주는 제주대학교 방음실). 방음 스튜디오 사용 및 녹음의 기술적 지원은 음성정보기술산업지원센터의 협조 하에 진행되었으며 사용된 녹음 기기는 아래와 같다.

- 디지털 녹음기: TASCAM DA-20MKII
- 오디오 믹서: Behringer MXB1002
- 헤드셋 마이크: Sennheiser HMD25-1
- DAT 테이프: SONY 10DT-120RA

각 지역 녹음을 위하여 출장 때에는 휴대용 DAT를 사용하였으며(TASCAM DA-P1), 이때 마이크는 SHURE SM10A를 사용하였다.

아래는 발화자의 섭의 및 녹음 준비 사항을 기술한 것이다.

#### 4.2. 섭의 과정

- 발화예정자에게 미리 준비된 공문을 배포하여 녹음 의지를 확인하고, 미리 준비를 갖추게 한다.

- 발화 확정자에게 발성목록을 미리 나누어 주어 (녹음 3-4일 전) 연습하게 하고, 그 목록에 발화자가 필요한 정보(발음기호, 끊어 읽기 등)를 표시하고 싶으면 그렇게 하여 연습하도록 하고, 다만 녹음 시에는 표기한 것이 없는 것으로 읽을 것이라고 예고한다.

- 초등용 녹음을 위해서는 초등학생 5-6학년, 중등용 녹음을 위해서는 고등학생 1-2학년, 일반용은 대학생이나 대학원생을 섭의하였다.

- 발화자 섭의 시 영어를 제1언어로 사용하는 국가에서 거주기간이 1년 이상인 자는 섭의 과정에서 제외시켰다.

- 녹음 당일 녹음 시작 전에 녹음일시, 녹음 엔지니어, 사용된 장비, 화자특성, 녹음 환경 등은 음성코퍼스에 반드시 포함시킬 수 있도록 기록하였으며, 이를 자세한 화자 데이터베이스로 따로 작성 운영하고, 이것을 후에 본 음성코퍼스에 포함할 수 있도록 하였다.

#### 4.3. 녹음 전 준비

- 녹음 모니터 요원은 3인 1조로 한다(발음체크 1명, 볼륨 및 장비체크 1명, 전체지도 1명).

- 모니터 요원은 발화자로 하여금 '발성화자정보지'를 녹음 전에 미리 작성케 하고, 그렇게 작성된 화자정보(발화자가 작성한 부분만)를 본 녹음 전에 녹음한 후 본 녹음에 들어간다.

- 발화자 정보(기타란)에 영어 학습에 관한 자세한 사항을 기록하도록 한다.

- 발화자의 긴장을 풀어주어 자연스러운 발화가 되도록 최대의 노력을 기울인다.

#### 4.4. 본 녹음

- SET#3의 발화 능력을 녹음하기 전에 미리 확인하며, 그 확인과정으로 녹음 볼륨 레벨을 조절한다.

- 읽을 때 각 단어 끝을 열거식 억양으로 읽지 않게 발화자에게 주지시킨다. 또한 단어와 단어, 문장과 문장 사이를 빨리 붙여 읽지 않도록 상시 체크한다.

- 발화오류의 경우, 단어는 해당 단어만을 문장은 해당 문장만을 다시 읽게 한다. 국문 및 영문 이야기의 경우는 오류 문장만을 다시 읽게 한다. (화자가 녹음 중 자신의 오류를 감지하고 다시 읽은 부분은 재녹음을 할 필요가 없으나, 다시 읽은 단어나 문장, 이야기 내의 문장을 별도의 '오류기록지'에 기록한다.)

- 화자가 빼놓고 읽지 않았거나 오류 발화, 갑작스런 화자 음성의 변화가 있는 단어나 문장등은 반드시 재녹음을 한다. 재녹음 리스트도 별도의 '재녹음기록지'에 기록한다.

- 단어 녹음에서는 굴절접사(어미의 복수형 's'나 과거형 어미 등을 생략했을 시) 오류를 재녹음한다. 그러나 문장 녹음에서는 이를 재녹음하지 않는다.
- 휴식시간은 매 SET마다 반드시 갖도록 한다. SET#4에서는 전체 SET의 1/3마다 쉬도록 한다. 그러나 발화자가 원하면 언제든지 휴식을 취하게 한다.
- 녹음을 끝낸 발화자로 하여금 '확인 및 영수증(첨부자료 #6)' 문서를 읽고 소속과 이름을 기록 후 날인하게 하며, 별도의 '발화자 목록표'를 기록하도록 한다.

#### 4.5. 녹음 후: 가공 예정

- 녹음된 자료를 디지털화하고 자료의 효율적인 검색을 위하여 체계적인 화일명을 부여한다.
- 음성신호의 분할 및 레이블링(labeling)을 선별적으로 실시한다. 분할 및 레이블링과 관련된 분석도구 및 그 활용은 음성정보기술산업지원센터의 협조를 얻는다.
- 선별적으로 시행할 전사(transcription)의 수준은 철자 수준(orthographic level)과 음성적 수준(phonetic level)을 모두 고려한다. 이는 연구 결과의 활용이 음성학자와 음운론자 그리고 영어발음 교육자들에게 모두 필요하기 때문이다.
- 운율 전사(prosodic transcription)도 반드시 포함하고 이는 ToBI 레이블링 체계를 따르도록 한다. 이는 한국인의 영어발음시 나타나는 억양의 오류를 객관적 체계에 따라 구현하기 위함이다.
- 총괄적으로 코퍼스에 다음 사항이 포함되도록 한다.
  - 신호 (음성신호, 일부는 후두신호 포함)
  - 특성 (화자특성, 녹음특성)
  - 각종 표시 (pitch구간 표시, 모음시작 표시)
  - 주석 (전사 표시)

## V. 결론

본 연구는 한국인의 영어발음 음성코퍼스(K-SEC)의 전반적 구축 배경 및 필요성, 그리고 설계상의 특징과 실제적 구축 과정을 기술하였다. 현재 K-SEC은 설계를 끝마치고 녹음의 막바지 작업에 있으며 (녹음 공정을 91%), 녹음이 곧 마무리되면 원광대학교 음성정보기술산업지원센터의 기술 지원 하에 각 발화 단위(어휘, 문장, 문단) 별로 화일화 한 후, 선별적 대표 발화에 대한 전사 작업에 임하게 될 것이고 제작 완료 후 본 코퍼스는 음성정보기술산업지원센터의 감독 하에 CD형태로 배포될 예정이다. 한편 K-SEC은 크게는 병렬 음성코퍼스의 일환으로서 이러한 병렬 음성코퍼스

는 다양한 언어 간 또는 동일 언어에 대한 L1 대 L2 병렬 관계 음성코퍼스로 제작을 확대하여 지속적인 구축이 필요한 분야로 인식되어야 하며, 이러한 분야의 개척은 추후 학제간 그리고 산학간 협동연구의 근간을 보여줄 수 있는 바탕 연구가 됨을 이해하여야 한다.

## 참고문헌

- [1] 고려대학교 민족문화연구원 국어연구소, 음성언어정보연구실. *음성언어의 연구 및 응용*. 학술대회 자료집, 2001.
- [2] 교육부. *초·중등 학교 교육 과정: 국민 공통 기본 교육 과정*, 교육부 고시 제1997-15호 [별책1], 1997.
- [3] 김기호, 이용주, 신지영, 이봉원. "한국어 방언음성 데이터베이스의 설계 구축 및 응용 방안 연구," *한국어학*, 11, 1997.
- [4] 김봉완, 이용주. "공동이용을 위한 음성DB의 구축 및 음성 자료 수집을 위한 workbench의 구현," *말소리*, 35-36, 1998.
- [5] 김선미, 문수미. "외래어 발음에서 나타난 영어와 한국어의 운율적 차이," *말소리*, 35-36, 1998.
- [6] 박전규. "영어 발성 능력 평가 자동화를 위한 지식 기반 발성해석기와 영어 음성인식기의 개발," 정보통신산업 기술개발사업 과제신청서, 2002.
- [7] 음성정보기술산업지원센터. *제1회 SITEC 심포지움 자료집*. 전북: 원광대학교, 2001.
- [8] Celce-Murcia, M., Brinton, D.M. and J.M. Goodwin. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. New York, NY: Cambridge University Press, 1996.
- [9] Edwards, J. and M. Lampert (eds.) *Talking data*. Hillsdale, NJ: Lawrence Erlbaum, 1993.
- [10] Ladefoged, P. *A Course in Phonetics*. (4th ed.) New York, NY: Harcourt College Publishers, 2001.
- [11] Leech, G., Meyers, G. and J. Thomas. (eds.) *Spoken English on computer*. Harlow, Essex: Longman, 1995.
- [12] Meyer, C.F. *English corpus linguistics: an introduction*. New York, NY: Cambridge University Press, 2002.
- [13] Peterson, G.E. and H.L. Barney. "Control methods used in a study of the vowels," *JASA* 24, 1952.
- [14] Shimuzu, M., Tomiyama, Y., Dantsuji, M. and K. Shimuzu. "Creating an English speech database



spoken by Japanese students for use with CALL system," *Proceedings of SNLP-Oriental COCOSDA 2002*, 2002.

- [15] Turk, A.E. and L. White "Structural influences on accentual lengthening in English," *Journal of Phonetics*, 27:2, 1999.