

# 한·영 동시조음 데이터베이스의 구축

## Speech Coarticulation Database of Korean and English

김 종 미\*, Stephen A. Dyer\*\*, Dwight D. Day\*\*  
(Jong Mi Kim\*, Stephen A. Dyer\*\*, Dwight D. Day\*\*)

\* 본 논문은 1997년도 연암 해외연구교수 지원비에 의해 Kansas State University에서 수행되었음.

### 요 약

본 연구는 국어, 영어, 콩글리쉬<sup>1)</sup>의 동시조음 표본을 최대 포괄한 "소리다" 음성 데이터베이스<sup>2)</sup> 개발하였다 [1]. "소리다"는 최소의 녹음목록에 최대의 삼중음소를 포함하도록 설계되어 자음의 경우는 해당언어의 모든 삼중음소를, 모음의 경우는 음운환경에 의한 모든 외존변이음을 포함시켰다. 또한, 강세의 유무에 따른 언어특성의 차이를 반영하여, 국어는 682개의 인공어를 조합하였고, 영어는 717개의 실제어를 색출하였다. 콩글리쉬는 717개의 영어녹음 목록을 국어화자로 하여금 발음하게 하였다. 화자는 성별, 연령별, 지역별로 안배된 국어 5인 화자, 영어 5인 화자, 콩글리쉬 2인 화자로 선별하였다. 녹음환경은 조용한 방에서 AKG 모델 C-1000 마이크와 Fostex 모델 D-5 DAT 녹음기를 사용하였다. 전체 녹음 분량은 4시간 분이다. 본 연구 결과물로서는 22.05kHz, 16bit로 샘플링된 씨디(Compact Disk), 48kHz로 샘플링된 DAT 녹음테이프, 언어별 음소 균형 단어 목록이 있다.

### ABSTRACT

We present the first speech coarticulation database of Korean, English, and Konglish<sup>3)</sup> named "SORIDA"<sup>4)</sup>, which is designed to cover the maximum number of representations of coarticulation in these languages [1]. SORIDA features a compact database which is designed to contain a maximum number of triphones in a minimum number of prompts.

SORIDA contains all consonantal triphones and vowel allophones in 682 Korean prompts of word length and in 717 English prompt words, spoken five times by speakers of balanced genders, dialects and ages. Korean prompts are synthesized lexicons which maximize their coarticulation variation disregarding any stress phenomena, while English prompts are natural words that fully reflect their stress effects with respect to the coarticulation variation. The prompts are designed differently because English phonology has stress while Korean does not. An intermediate language, Konglish has also been modeled by two Korean speakers reading 717 English prompt words.

Recording was done in a controlled laboratory environment with an AKG Model C-100 microphone and a Fostex D-5 digital-audio-tape (DAT) recorder. The total recording time lasted four hours. SORIDA CD-ROM is available in one disk of 22.05 kHz sampling rate with a 16 bit sample size. SORIDA digital audio-tapes are available in four 124-minute-tapes of 48 kHz sampling rate. SORIDA's list of phonetically-rich-words is also available in English and Korean.

\* 강원대학교 영어영문학과  
\*\* BECE, Kansas State University  
접수일자: 1998년 9월 15일

- 1) "콩글리쉬"는 속어로 한국어 엑센트의 영어발음을 일컫는다. 콩글리쉬 데이터베이스는 외래어의 인식 및 합성에 필요하다.
- 2) "소리다"는 국어 명칭이며, 영어로는 SOund RIch DATABASE의 약칭이다. "소리다"가 최초로 발표된 것은 1998년 5월 스페인에서 개최된 First International Conference on Language Resources and Evaluation 학술대회이며, 본 논문의 일부는 당시 Proceedings 에 실렸던 영문 논문에서 발췌, 번역하였다. 본 데이터베이스의 구상과 제작시 귀한 조언을 하여 주신 한국 통신의 구명완 박사님께 감사 드린다. 불문 오류는 집필진의 책임이다.
- 3) The word "Konglish" is Korean slang referring to the English language spoken with a Korean accent. Konglish database is needed for the recognition and synthesis of English loan words in Korean speech.
- 4) SORIDA is a Korean term meaning "(these) are sounds." For our database, we also use it as an acronym of the English phrase, "SOund RIch DATABASE." SORIDA has first been introduced in the first LREC conference, and part of the conference presentation has been translated and incorporated to the present paper.

## I. 서 론

### 1.1. 개발 동기

“소리다”는 한국어와 영어의 동시조음현상이 최대한으로 포함된 음성 데이터를 최소의 음성목록에 저장하여, 음성 합성, 음성인식 및 한·영 자동통역을 위한 기초자료로 사용하고자 고안되었다. 음성현상은 인접음의 영향에 따라 굴절이 심하므로 인접음을 무시하고 개별음만을 표본화하는 것은 인식능력과 합성성능을 현저히 떨어뜨리기 때문이다. 그러나 모든 인접음 조합의 음성표본을 포괄하는 것은 계산량이 막대하므로 기존 음성 데이터베이스는 데이터 결핍현상을 빚고 있었다 [2]. 더욱이, 통역기는 두 개의 언어의 인식과 합성이라는 네 가지 기능을 기본적으로 요구하고 있으므로, 기존 단일 기능 음성 데이터베이스로서는 자동통역의 최소요건을 충족시킬 수 없을 뿐 아니라, 서로 다른 4개의 단일기능 데이터베이스를 동시 사용한다 해도 계산량의 초과 현상을 빚게 된다.

본 “소리다” 음성 데이터베이스는 각 해당언어의 음운적 특성에 맞게 효율적으로 구성하여 전체 동시조음 현상을 포괄하고, 대표적인 성별, 연령별, 방언별 화자를 선별하여 국어, 영어, 콩글리쉬의 세 영역에 걸쳐 녹취함으로써, 두 언어의 인식·합성·통역 기능을 복합적으로 수행할 음소 균형 데이터를 단 4시간의 녹음시간에 총망라하였다. 이와 같은 접근방식은 기존 대다수의 단일 목적형 일회성 데이터베이스가 아닌 다목적형이고 축적성 데이터베이스를 만들고자 고안된 것이다.

본 연구 결과물인 음성 데이터베이스를 저장한 CD-ROM과 Digital Audio Tapes(DAT), 그리고 음소균형 단어목록은, 1)음성·음운론의 변이 및 교체현상 연구와 2)음성언어처리 연구, 즉 한·영 자동통역기, 국어 음성합성기, 국어 음성인식기, 영어 음성합성기, 영어 음성인식기 개발에 종사하는 학자들에게 제공하는 것을 목적으로 하고 있다.

### 1.2. 연구 배경

다국어 동시조음 데이터베이스는 국내에서는 공개된 바가 없고, 외국에서는 다국어 데이터베이스의 형태로 유럽의 ACCOR, BDSONS, EUROM1, M2VTS, MULTTEXT, ONOMASTICA-COPERNICUS, TED와, 미국의 OGI Multilanguage Corpus, CALLHOME, CALLFRIEND 등이 있다 [3] [4].

이 중 가장 많은 언어를 포함한 데이터베이스는 EUROM1 으로서 11개국의 언어를 같은 형식의 녹음내용과 화자 분포를 통하여 같은 오디오 파일 형식으로 저장하였다. 특히 언어마다 CVC 자모음들 속에 모든 자음과 [a, i, u] 3개 모음을 번갈아 결합하였다. 여기서 C는 자음을, V는 모음을 지칭하는데, 각각 Consonant와 Vowel의 약자이다.

자모음들의 사용은 다국어 데이터베이스에서 종종 발견되는 바, 위에 언급한 EUROM1 데이터베이스의 CVC

를 외에도, ACCOR 데이터베이스의 VCV들과 BDSONS 데이터베이스의 CVCV들이 있다.

또한, 비모국어 데이터베이스로서는 다양한 외국인들의 영어발음을 수록한 Translanguage English Database (TED)가 있다 [3].

한국어를 다룬 외국의 데이터베이스는 OGI multi-language Corpus와 CALLHOME이며, 각각 전화 음성을 녹음하였다.

본 연구 “소리다” 동시조음 데이터베이스는 위에 열거한 대다수의 다국어 데이터베이스처럼, 해당 언어인 영어와 국어를 통일된 형식의 녹음 내용과 화자 분포를 통하여 녹화한 후, 통일된 오디오 파일 형식으로 저장하였다. 특히, “소리다”는 EUROM1, ACCOR, BDSONS처럼 일정한 자모음들을 사용하였다. 또한, TED에서처럼 비모국어 발음을 고려하기 위하여, 국어화자의 영어 발음인 콩글리쉬 데이터베이스도 제작하여 포함시켰다.

한편, 이들 기존 데이터베이스 보다 더 나아가, 본 연구는 언어별 특성 차를 반영하여 제작하였다. 즉, 1) 언어별 음소 규칙에 따라 어휘표를 구성하였고, 2) 언어별 음운 특성에 맞게 녹음 목록을 작성하였으며, 3) 언어별 언어 사회군에 따라 화자분포를 선정하였다.

데이터베이스의 구축 방법은 1) 어휘표의 작성, 2) 녹음 목록의 작성, 3) 화자의 선정, 4) 녹음 및 5) 컴퓨터 처리의 다섯 단계에 걸쳐 다음과 같이 기술된다.

## II. 언어별 음소 규칙에 따른 어휘표의 작성

어휘표의 작성은 국어와 영어의 언어별 음소 배열 규칙에 따랐다. 또한, 언어 특성상 국어는 문자와 소리의 일치도가 높고 영어는 일치도가 낮기 때문에, 각 경우에 맞게 최선과 최대의 데이터를 취하기 위하여, 국어는 인공어를, 영어는 실제어를 선정하였다.

언어별 자모음들은, 국어의 경우 CVCCV를 기준으로 하고 VCV, VC, VV 유형을 보충하였으며, 영어는 CVC를 기준으로 하고 CV, VC, V 유형 및 성절자음과 자음군에 해당하는 모든 조합을 포괄하였다. 예컨대 국어의 경우는 “듭드, 답더, 답다, 들두, 돕도”와 같은 CVCCV의 인공어를, 영어의 경우는 “pip, tit, kick, bib, did, gig”와 같은 CVC의 실제어를 선택하였다. 또한, 언어별 음소 배열 규칙에 따라 국어는 음절내 자음군이 없는 한편, 영어는 “strike, texts”와 같은 음절내 자음군을 망라하였다.

### 2.1. 국어 어휘표

국어의 어휘표는 1)자모음 교체, 2)모음간 자음, 3)중성 장애음, 4)기준 모음의 네 가지 유형으로 작성하였다. 이들 네 유형의 자모음들을 사용하여 최다의 삼중음소의 개수를 최적의 어휘의 개수에 포함시키는 방법을 순서대로 기술하겠다.

첫 번째 유형인 자모음 교체는 다음 표1에서 예시된 바와 같이 CVCCV 자모음들을 사용하였으며, 전체 682개로 된 국어 어휘표 중 대다수인 624개를 차지한다. 국

어의 발음기호는 정국 외 (1994)를 따랐으며, 불특정 자모음을 나타내기 위한 C와 V는 밑줄로 표기하여 발음기호와 구분하였다 [5].

표 1. 국어의 자모음 교체 어휘표  
Table 1. The Korean lexicon of consonantal alternation.  
(No. 41-46 of 78 samples)

No.	Pattern	Lexicon
41	C <u>V</u> <sub>n</sub> C <u>V</u>	친치,철헌,철헌,촌츠,친차,찬차,춘추,촌츠
42	K <u>V</u> <sub>n</sub> K <u>V</u>	킨키,켄켄,켄켄,크크,킨카,칸카,쿤우,콘우
43	h <u>V</u> <sub>n</sub> h <u>V</u>	헨히,헨헨,헨헨,혼호,헨하,한하,훈후,혼호
44	m <u>V</u> <sub>n</sub> m <u>V</u>	민미,멘메,멘메,므므,민머,만마,문무,문모
45	n <u>V</u> <sub>n</sub> n <u>V</u>	넌니,넌네,넌네,느느,넌너,난나,눈누,논노
46	b <u>V</u> <sub>n</sub> b <u>V</u>	벌비,벨베,벨베,블블,벌버,발바,블부,블보

위 표1에서는 다양한 자음을 단어초, 음절말, 음절초의 위치에 배열시킨 후, 8개의 단순모음과 번갈아 결합시킨다. 예컨대 (46)번 표본의 경우, 단어초 자음 [b], 음절말 자음 [r], 음절초 자음 [n]과 다양한 모음을 결합시켰다. 따라서 CVCCV 자모음들은 #CV, CVC, VCC, CCV, CV# 형태의 삼중음소를 포괄한다. 이 중 #표시는 단어 경계를 의미하며, 단어초나 단어말의 묵음 표시와 동일하다.

어휘수를 최적의 개수로 제한하기 위하여, 음운동화 작용으로 인한 중복음은 삭제하였다. 예컨대, 위 (44)번의 [민미]는 다수의 철자형 /민미/, /민미/, /민미/, /밋미/, /밋미/, /밋미/로부터 발화되는 음이므로, 이 중 /민미/ 하나만을 녹음한다<sup>5)</sup>. 이와 같은 현상은 국어 음운론에서 발음법칙과 비음 동화로 인하여 발생된다. 다른 국어 음운 현상, 예컨대 경음화, 유음화등으로 인한 데이터의 중복도 위와 같은 방법으로 축소한다.

두 번째 유형의 국어 어휘표는 모음간 자음을 조합하기 위하여 VCV 자모음들을 사용하였다.

표 2. 국어의 모음간 자음 어휘표  
Table 2. The Korean lexicon of intervocalic vowels.  
(No. 01-08 of 32 samples)

No.	Pattern	Lexicon
01	aCa	아바다가자사,
02		아빠따까짜짜,
03		아파타카차하,
04		아아나라양
05	eCe	에베대개재재,
06		에베대개재재,
07		에베대개재재,
08		에베대개재

5) 음성학의 관계에 따라, 기저 음운형은 /민미/와 같이 사선으로 표민 음성형은 {민미}와 같이 적색으로 표기한다.

위 표2에서는 다양한 자음을 모음간의 위치에 배열시킨 후, 8개의 단순모음과 번갈아 결합시킨다. 예컨대, (01)-(04)번 표본의 경우, 19개의 자음과 [아]모음이 결합하였다. 따라서 이 방식으로 수록한 VCV 자모음들의 어휘표는 VCV 형태의 삼중 음소를 포괄한다.

어휘수를 최적의 개수로 제한하기 위하여, 단순모음을 자음과 결합시켰을 뿐, 이중모음은 생략했다. 그 이유는, 1)이중모음은 따로 아래 표4에서 다룰 것이며, 2)이중모음과 자음간의 다양한 조합은 단순모음의 합에서 유추될 수 있고, 3)음성언어처리 연구시 모음의 경우는 자음보다 상대적으로 용이하기 때문이다.

세 번째 유형의 국어 어휘표는 중성 장애음을 조합하기 위하여 VC를 사용하였다. 국어 중성 자음은 총 7개로 [ɾ, ɸ, ɸ, ɸ, ɸ, ɸ, ɸ]이며, 이미 첫 번째 유형 CVCCV틀에서 다양한 결합형태를 섭렵한 바 있다. 그러나, 이들 중성자음 중에서도 장애음 [ɾ, ɸ, ɸ]은 특히 청취 인지도가 낮으므로, 그 대표 음가를 아래 표3과 같이 표본화한다.

표 3. 국어의 중성 장애음 어휘표  
Table 3. The Korean lexicon of coda obstruents.  
(No. 01-03 of 3 samples)

No.	Pattern	Lexicon
01	<u>Y</u> p	입, 읊, 업, 압, 읊, 읊
02	<u>Y</u> r	잇, 옷, 잇, 잇, 옷, 옷
03	<u>Y</u> k	익, 옥, 억, 악, 옥, 옥

위 표3에서는 국어의 모든 음절말 장애음을 6개의 단순모음과 결합시켰다. 예컨대, 위 (02)번 표본의 경우 장애음 [ɸ]과 다양한 모음이 결합하였다. 여기서, 표본 잇 [일]은 “잇, 잇, 잇, 일, 일” 등의 다양한 받침자로 표기되는 발음 [일]을 대표한다. 따라서, 표3과 같은 유형은 #VC 형태의 장애음 삼중 음소를 포괄한다.

어휘수를 최적의 개수로 제한하기 위하여 다음 세 가지 방법을 사용하였다. 1)경음, 격음, 마찰음 계열의 장애음을 모두 삭제하였다. 그 이유는 이들 단어는 중성에서 발음법칙의 영향을 받아 [ɾ, ɸ, ɸ]으로 변하기 때문이다. 2)국어 7개의 중성자음 [ɸ, ɸ, ɸ, ɸ, ɸ, ɸ, ɸ]중 청취 인지도가 높은 공명음 [ɸ, ɸ, ɸ]은 별도의 #VC틀로 설정하지 않고 앞서 첫 번째 자모음틀인 CVCCV틀에서 수록한 것으로 사용한다. 3)모음을 6개의 단순모음에 국한시키고, 나머지 2개의 단순모음 [에, 에]를 제외한다. 이는, [에, 에]의 음가 구분을 하지 않는 개인이 많기 때문이다. 따라서, 데이터의 안정성을 위해서, [에, 에]의 경우 VC틀의 대표 음가를 설정하지 않고 첫 번째 유형의 CVCCV 자모음틀에 나타난 개별 표본을 사용한다.

마지막으로, 네 번째 유형의 국어 어휘표는 기준 모음의 안정구간을 추출하기 위하여 VV 모음들을 사용하였

다. 이들 모음이 자음과 결합한 형태는 이미 앞서 기술한 세 유형에서 도출된다.

표 4. 국어의 기준모음 어휘표  
Table 4. The Korean lexicon of reference vowels.  
(No. 01-06 of 6 samples)

No.	Pattern	Lexicon
01	VV	아 야 애 예
02	VV	어 여 에 예
03	VV	오 요 외 예 와
04	VV	우 유 위 웨 워
05	VV	으 외
06	VV	이

위 표4에서는 전체 21개의 단순모음과 이중모음을 나열하였다. 이들 모음을 연결하면 #V#, #VV, VV#, VVV 유형의 삼중 음소를 생성할 수 있다.

그러나, 어휘수를 최적의 개수로 제한하기 위하여, 모음끼리의 모든 가능한 조합은 시도하지 않았다. 모음은 형성음이 뚜렷하여 그 음향적 특징이 쉽게 판별되므로 음성 인식과 합성이 용이하기 때문이다.

이상 이 장에서 다른 국어 어휘표의 내용을 요약하자면 표1-표4에서 본 바와 같이 "소리다"의 국어 어휘표는 다양한 단어내 음운 환경을 포괄한다. 초성이 없는 경우(표3, 표4), 초성이 있는 경우(표1, 표2), 종성이 있는 경우(표1, 표3), 모음간 자음이 있는 경우(표2), 단어초 자음이 있는 경우(표1)이다. 이에 따라 다양한 삼중음소 유형이 도출되는데 표1에서는 #CV, CVC, VCC, CCV, CV# 형태를, 표2에서는 VCV 형태를, 표3에서는 #VC 형태를, 표4에서는 #V#, #VV, VV#, VVV 형태를 전체 682개의 녹음 목록에 모두 수록하였다.

또한 어휘수를 최적의 개수로 제한하기 위하여, 1)음향 특징이 잘 나타나는 모음과 공명음의 결합 형태를 축소하였고, 2) /민미, 밀미, 밋미, 밋미, 밋미/ 처럼 동음이형 때는 발음 나는 대로 [민미] 라는 동일형태로 축약하였다.

이처럼 어휘 표본의 자모음을 설정과 어휘 유형별 중복 제거는 언어학적인 근거에 의한 것으로서, 1)본질음의 음소 배열 규칙과, 2)공명음의 음향적 변별력에 기인한 것이다.

다음 장에서 기술하는 영어 어휘표도 위와 같은 언어학적 원칙에 의하여 제작한다. 다만 국어와 영어의 언어 유형이 다르기 때문에 그 결과물은 판이하게 다르게 나타난다.

2. 2. 영어 어휘표

영어 어휘표는 1)자모음 교체, 2)기준 모음, 3)초성 자음군, 4)종성 자음군의 네 가지 유형으로 작성하였다.

6) 단어말 파열음은 길이를 일정하게 표준화할 수 없으므로, 일정한 녹음틀을 갖춘 표11의 형태로 녹취한다.

첫 번째 유형인 자모음 교체는 다음 표5에서 예시된 바와 같이, 영어의 각 모음별로 초성자음과 종성자음을 교체 시켰으며 실제 단어를 수록하였다. 이와 같은 유형은 전체 717개 단어의 영어 어휘표 중 대다수인 463개 단어를 차지한다. 영어의 발음기호는 TIMIT을 따랐다 [6] [7].

표 5. 영어의 자모음 교체 어휘표  
Table 5. The English lexicon of consonantal alternation.  
(No. 505-511 of 336 samples)<sup>7)</sup>

No.	Pattern	CV Lexicon	VC Lexicon
505	[daed]	dad	(dad)
506	[gaeg]	gag	(gag)
507	[chaech]	chai	attach
508	[jhaejh]	jazz	badge
509	[maem]	maam	(maam)
510	[naen]	Nan	(Nan)
511	[aeng]	XXXX	gang

위 표5에서는 다양한 자음이 음절초와 음절말의 위치에서 [ae] 모음과 결합되었다. XXXX 표시는 영어의 실제어에 없는 경우를 표기한다. 위 (511)번의 XXXX는 영어의 음소 배열 제약 조건상 단어초 [ng] 발음이 없음을 나타낸다. 또한, 발호안의 단어는 다른 유형에서 이미 채택된 어휘를 더불어 사용하는 것이다.

표5의 단어는 다음 선정기준에 의하여 선택되었다. 1) 초성과 종성이 같은 CVC; 자모음들을 우선적으로 추출하여 모음의 안정구간을 최대화한다.<sup>8)</sup> 조음점이 서로 다른 자음은 형성음을 굴곡시키기 때문이다. 2)만일 초성과 종성이 같은 CVC; 유형의 실제어가 없을 때는 두 자음 중 중요도가 낮은 자음을 경구개 장애음인 [d,t,s] 등으로 덜 중요한 부분에 대처시켜 단어 목록을 만든다. 이들 음소 [d,t,s]는 동시조음시 전이구간을 최소로 변화시키기 때문이다.

표5의 단어 유형으로 인한 삼중음소는 #CV, VC#, CVC 형태이다. 예컨대 위 표의 샘플번호 (505)번의 경우, "Dad"가 도출하는 삼중음소는 #da, ad#, dad 의 3가지이다. 이와 같은 자모음들에 맞춰 14개의 모음과 24개의 자음을 번갈아 결합시켰다. 모음의 경우는 장세모음 모두와 비강세 모음 몇 개를 포함한다.<sup>9)</sup>

7) 표본의 번호는 유형별 단위배정에 의하여 때의 자리수가 결정되므로 총 샘플수보다 낮은 숫자일 수 있다.

8) 여기서 C;는 identical consonant를 표시하며, CVC;는 두 자음이 동일함을 나타낸다.

9) 비강세 모음으로서는, "bag, adopt"에 나타나는 [ax] 발음, "butter"에 나타나는 [axr] 발음을, 모든 자음과 번갈아 결합시켰고, "grec"에 나타나는 [ix] 발음을 선정된 몇몇 자음과 결합시켰다.

어휘수를 최적의 개수로 제한하기 위하여, C<sub>1</sub>VC<sub>2</sub> 형태의 순차적 조합을 실시하지 않았다. 그 이유는 C<sub>1</sub>VC<sub>2</sub> 형태는 수록된 표본중 #C<sub>1</sub>V와 VC<sub>2</sub>#를 추출, 결합하여 사용해도 모음부분의 결합이라는 점에서 오차의 소지가 적기 때문이다<sup>10)</sup>. 음향 표본의 결합, 도출 방식은 Young et. al (1994)을 참조하기 바란다 [2].

두 번째 유형의 영어 어휘표는 기준 모음을 설정한 것이다. 기준 모음에는 강세모음, 비강세모음, 이중모음은 물론 성절자음도 포함시켰다. 성절자음이란 아래 표6의 (20)-(22)표본에 예시된 바와 같이 공명자음이 모음처럼 음절핵을 이룬 것을 가리킨다.

표 6. 영어의 기준 모음 어휘표  
Table 6. The English lexicon of reference vowels.  
(No. 10-22 of 22 samples)

No.	Pattern	V(C)Lexicon	CV(C) Lexicon
10	[uw]	woo	boot
11	[ah]	us	but
12	[er]	err	bird
13	[ax]	about	soda
14	[axr]	ergodic	butter
15	[ix]	erect	orbit, cryptic, bandit, cabin, goodness
16	[aw]	our, bow	house
17	[ay]	I [ay], are	bite
18	[oy]	Oy! soy, annoy	boil, toil, loin, Freud
19	[el]	XXXX	apple, bottle, middle, wrinkle, ogle, (feeble)
20	[em]	XXXX	bottom, chasm
21	[en]	XXXX	button, redder
22	[eng]	XXXX	XXXX

위 표6에 수록된 예들은, 1) [uw] 처럼 열린모음, 2) [ah] 처럼 닫힌모음, 3) [uw, ah, er] 처럼 강세모음, 4) [ax, axr, ix] 처럼 비강세모음, 5) [aw, ay, oy] 처럼 이중모음, 그리고 6) [el, em, en, eng] 처럼 성절자음을 두루 포함하였다<sup>11)</sup>.

표6으로부터 얻을 수 있는 삼중 음소는 #V#, #VC, VC#, CV#, CVC, #CV 유형이다. 영어 음소배열 제약조건에 의하면, 이 중 #V#와 CV#유형의 삼중음소는 소위 열린 모음에만 가능하고, 나머지 #VC, VC#, CVC, #CV 유형은 열린 모음과 닫힌모음에 모두 가능하다.

10) 여기서 C<sub>1</sub>, C<sub>2</sub>는 서로 다른 두 자음을 나타낸다.  
11) 열린모음이란 강세 음절의 중성 위치에 올 수 있는 모음을 일컫고, 닫힌 모음은 동위치에 올 수 없는 모음을 일컫는다. 열린모음은 /iy, ey, aa, ao, ow, ay, aw, oy/가 있고, 닫힌 모음은 /ih, ix, eh, ae, er, ah, uh, ax, axr/가 있다.

어휘수를 최적의 개수로 제한하기 위하여, V<sub>1</sub>V<sub>2</sub>형태로 다양한 모음을 결합하는 것은 실시하지 않았다. 모음은 인식 합성을 하기가 쉽기 때문이다.<sup>12)</sup>

지금까지 표5와 6에서는 보다 일반적인 유형의 모음과 자음의 결합을 포괄하였다. 다음에 기술할 표7과 표8에서는 영어의 모든 자음군을 망라하고자 한다. 자음군은 특히 음성 인식과 합성을 하기가 어려우므로 전체를 포괄하는 것이 중요하다.

아래 표7에서는 세 번째 유형의 영어 어휘표를 발해 예시한 것으로서, 영어의 초성 자음군을 포괄하기 위한 것이다. 영어는 국어와 달리, 음절초에 자음군을 허용한다. 영어의 음절초 자음군으로서의 최대 세 개의 자음이 결합되어 나타날 수 있다.

표 7. 영어의 초성 자음군 어휘표  
Table 7. The English lexicon of onset clusters.  
(No. 10-23 of 52 samples)

No.	Clusters	Lexicon
10-13	[ ky, kw, kr, ki ]	cue, quest, cry, clue
14-17	[ sky, skw, skr, ski ]	skew, squash, scratch, sclerosis
18-20	[ by, br, bi ]	beauty, bribe, blind
21-23	[ dy, dw, dr ]	dew, Dwight, drop

이와 같은 방식으로 “소리다” 데이터베이스에는 영어의 초성 자음군이 모두 나열되었다. 표7의 초성 자음군 데이터로부터 얻을 수 있는 삼중 음소는 #CC, CCV, CCC 유형이다.

표본 작성시에는 지금까지와는 달리, 초성 자음군 어휘수를 최적의 개수로 제한하기 위한 시도를 전혀 하지 않고, 전체를 망라하였다. 그 이유는 1)자음군의 인식 및 합성이 가장 어렵고, 2)전체 목록을 다하여도 그 수효가 52개 단어 정도로 적기 때문이다.

마지막으로 네번째 유형의 영어 어휘표는 종성 자음군으로서 아래 표8에 예시하였다. 영어는 국어와 달리, 음절말에도 자음군을 허용한다. 영어의 음절말 자음군으로서 네 개의 자음까지 결합하여 나타날 수 있다.

표 8. 영어의 종성 자음군 어휘표  
Table 8. The English lexicon of coda clusters.  
(No. 10-19 of 179 samples)

No.	Clusters	Lexicon
10-11	[ kt, kts ]	act, acts
12-14	[ ks, kst, ksts ]	box, text, texts
15-16	[ ksth, ksths ]	sixth, sixths
17-18	[ bz, bd ]	bobs, bobbed
19	[ dz ]	Ted's

12) 여기서 V<sub>1</sub>V<sub>2</sub>는 두 개의 서로 다른 모음이 인접함을 나타낸다.

표8에 나타난 종성 자음군은 영어 음소배열 제약 조건이 허용하는 모든 경우의 수를 총망라하였다. 이와 같이 작성된 종성 자음군 표본은 삼중음소 CC#, VCC, CCC 유형을 도출한다.

어휘수를 최적의 개수로 제한하려는 시도는 하지 않고 전체를 망라하였는데, 그 이유는 1)초성 자음군의 경우에서와 마찬가지로 자음군의 인식 합성이 어렵고 2)전체 종성 자음군의 수는 180여 단어로서 표본 추출용으로는 약소한 숫자이기 때문이다.

요약하자면, "소리다"의 영어 어휘표는 삼중 음소를 총망라하는데, 1) 자음 뒤에 모음이 이어지는 CV#, #CV, CVC 유형들, 2) 모음으로 시작하는 VC#, #V#, #VC 유형들, 3) 자음군을 포함한 #CC, CCV, CCC, CC#, VCC 유형들을 포괄한다. VCV 유형의 삼중음소는 포함하지 못하였는데, 이 경우는 #VC와 CV#의 결합으로부터 도출한다. 다만 VCV 삼중 음소에만 나타나는 변이음인 [r]의 발음은 butter, button, bottom 등의 단어를 통해 포함되었다.

어휘수를 최적의 개수로 제한하는 합축 기술은, 언어별 음소 특성 중 1)음소 배열제약 조건, 2)강세의 유무, 3)음향 특징의 변별성에 기반한다. 첫째, 음소 배열제약 조건에 의하여 영어에서는 음절초와 음절말의 자음군을 총망라하였고, 국어에서는 음절간 자음 접점의 교체를 실시하였다. 그 이유는 음절 내 자음군이 영어에는 있고 국어에는 없기 때문이다. 둘째, 강세의 유무로 인하여, 어휘표 작성시 영어는 실제어로, 국어는 인공어로 수록하였다. 그 이유는 영어는 강세가 있어 문자와 소리의 일치도가 낮아 화자가 인공어를 일정하게 읽을 수 없고, 국어는 강세가 없어 문자와 소리의 일치도가 높아 화자가 인공어를 일정한 방식으로 읽기 때문이다. 셋째, 음향 특징의 변별성에 기반하여 전체 자음을 다양한 음운환경에서 총망라하여 교체 결합시킨 반면, 모음 부분은 상대적으로 교체의 폭을 줄였다. 그 이유는 자음의 음향 특징이 모음의 경우보다 미약하게 판독되기 때문이다. 이러한 기준으로 작성된 "소리다" 어휘표는 국어와 영어의 언어별 운율 특성에 따라 재배열된 후 실제 녹음 목록으로 사용된다.

III. 언어별 운율특징에 따른 녹음 목록의 작성

녹음 목록의 작성은, 국어와 영어의 언어별 운율 특성에 맞게, 국어는 음절의 개수에 따라 영어는 각운(Rhyme)의 유사도에 따라 단어를 재배열하였다. 국어는 음절언어이므로 음절수가 중요하고, 영어는 강세언어이므로 각운의 속성이 중요하다.

3.1. 국어 녹음 목록

국어의 녹음 목록은 음의 길이와 높낮이를 고르게 하기 위하여 다음과 같이 작성하였다. 첫째, 음의 길이를 고르게 하기 위하여 네 일정 음절수마다 끊어 읽기 표시 “;”를 하였다. 언어유형상 국어의 소리길이는 음절의 개수로 결정되기 때문이다. 둘째, 음의 높낮이를 고르게 하기 위하여 문장의 처음과 어절의 끝에는 반복어를 삽입하였다.

국어의 악센트 유형상 이들 위치는 음정의 변화가 심한 부분이기 때문에, 이들 반복어가 대신 배치되면, 실제 데이터만큼은 일정 음정을 유지하게 되기 때문이다. 예컨대, “답다; 답다, 뎡뎡, 뎡뎡, 뎡뎡, 뎡뎡; 뎡뎡, 뎡뎡, 뎡뎡, 뎡뎡; 뎡뎡, 뎡뎡, 뎡뎡, 뎡뎡”처럼, 매 10음절마다 끊어 읽게 했고, 문장의 처음과 어절의 끝에 허사를 넣었다.

그러면 이와 같은 원칙에 의해 앞 단원의 국어 어휘표가 실제녹음목록으로 재구성된 바를 1)자모음 교체, 2)모음간 자음, 3)종성파열음, 4)기준 모음의 순으로 기술하고자 한다. 그 첫 번째 유형으로서 아래 표9는 앞서 표1에서 제시한 자모음 교체에 관한 어휘목록을 실제녹음목록으로 재구성한 것이다.

표 9. 국어의 자모음 교체 녹음목록  
Table 9. The Korean prompts of consonantal alternation.  
(No. 41-46 of 78 samples)

No.	Phonation	Prompts
41	[ 친치 ]	친치, 친체, 친체, 친츠, 친츠; 친치, 친차, 친추, 친초, 친초
42	[ 킨키 ]	킨키, 킨개, 킨개, 킨크, 킨크; 킨키, 킨카, 킨쿠, 킨코, 킨코
43	[ 힌히 ]	힌히, 힌해, 힌해, 힌호, 힌호; 힌히, 힌하, 힌후, 힌오, 힌오
44	[ 민미 ]	민미, 민매, 민매, 민모, 민모; 민미, 민마, 민무, 민모, 민모
45	[ 넌니 ]	넌니, 넌내, 넌내, 넌느, 넌느; 넌니, 넌나, 넌누, 넌노, 넌노
46	[ 빌비 ]	빌비, 빌배, 빌배, 빌보, 빌보; 빌비, 빌바, 빌부, 빌보, 빌보

표9는 일정한 음절수마다 끊어읽기 표시 “;”를 하여 매음절 일정 길이를 유지하고, 발화구조와 억양구말에 반복어를 삽입하여 본 데이터의 음정 변화를 최소화하였다. 예컨대, (41)번 표본의 경우, “친치, 친츠, 친초”를 반복·삽입하였다. 이들 반복어는 추후 실제 데이터로서는 사용하지 않고 삭제한다.

두 번째 유형의 국어 녹음 목록은 앞서 표2에서 제시한 모음간 자음에 관한 어휘목록을 실제녹음목록으로 재구성한 것이다.

표 10. 국어의 모음간 자음 녹음목록  
Table 10. The Korean prompts of intervocalic vowels.  
(No. 01-08 of 28 samples)

No.	Phonation	Prompts
01	[ 아바 ]	아바다가자사사,
02		아빠따까짜짜짜,
03		아파타카차하하,
04		아마나라앙앙아
05	[ 에베 ]	에베테게제제제,
06		에베떼게제제제,
07		에베테케제제제,
08		에베테레앵앵에

위 표10에서도 앞서 표9의 경우와 같이, 모든 단어의 음절수가 같도록 끊어 읽기를 하고, 운율단위 경계인 발화구조와 억양구말에 반복어를 삽입하여 본 데이터의 음정변화를 최소화하였다. 예컨대 위 (01)-(04)번 표본의 경우, 반복음 [아, 사, 싸, 하, 앙] 등이 삽입되었으며, 이들은 실제 데이터 편집시 삭제하여 활용한다.

세 번째 유형의 국어 녹음 목록은 앞서 표3에서 제시한 중성 장애음에 관한 어휘목록을 실제 녹음 목록으로 재구성한 것이다.

표 11. 국어의 중성 장애음 녹음목록  
Table 11. The Korean prompts of coda obstruents.  
(No. 01-03 of 3 samples)

No.	Phonation	Prompts
01	[ 입따]	입다, 읊다, 엮다, 압다, 읊다, 읊다, 읊다
02	[ 일따]	잇다, 옷다, 엮다, 앓다, 옷다, 옷다
03	[ 익따]	익다, 옥다, 엮다, 악다, 옥다, 옥다

위 표11에서는 매 중성 장애음 [브, 드, ㄱ] 뒤에 “다”라는 음절을 넣어 본 데이터의 음의 길이를 일정하게 했고, 발화구조에 반복어를 넣어 음정 변화를 흡수하였다. 또한 발화자의 망설임을 최소화하기 위해, 용언활용의 으뜸형인 “다”를 허사로 사용하였고, 어근의 표기를 흔히 쓰는 받침표기인 “ㅏ”으로 하여 시각적 인식을 용이하게 하였다.

네 번째 유형의 국어 녹음 목록은 앞서 표4에서 제시한 기준 모음에 관한 어휘목록을 국어운율에 맞도록 재구성한 것이다.

표 12. 국어의 기준모음 녹음목록  
Table 12. The Korean prompts of reference vowels.  
(No. 01-06 of 6 samples)

No.	Phonation	Prompts
01	[ 아]	아야 애와는
02	[ 어]	어여 예외가
03	[ 오]	오요 요위로
04	[ 우]	우유 웬일로
05	[ 으]	으이 의외로
06	[ 왜]	왜왜 재워야

위 표12에서는 두 모음씩을 짝하여 일정 음절수의 단어를 이루게 하고, 실제어 격조사인 “는, 가, 로, 야” 등으로 억양구를 마감하도록 하였다. 일정 음절수의 단어는 음의 길이를 일정하게 하고, 격조사의 삽입은 음의 높낮이 변화를 흡수하게 한다. 또한 실제어 조사의 삽입은 녹음시 화자로 하여금 쉽게 읽을 수 있게 한다. 삽입된 격조사는 추후 데이터 편집시 삭제한다.

이상 표9-12에 걸쳐 설명한 국어의 녹음 목록은, 음절의 개수를 정확히 하고, 운율단위 경계부분에 반복어를 삽

입하였다. 또한 가급적 화자의 인지도가 높은 어미나 조사를 사용하여 화자의 망설임을 최소화하였다. 이 모든 것은 음의 길이와 높낮이를 일정하게 하기 위한 것이다.

이처럼 음의 길이와 높낮이를 일정하게 유지하는 것은 다음 장에서 설명할 영어 녹음 목록 작성시에도 일관되게 추구된다. 다만 구체적인 운율조정방식은 언어별 특성에 따라 서로 다르게 적용되어, 지금까지 제시한 음절 개수와 반복어 삽입에 관한 방법은 음절 언어인 국어의 녹음 목록 작성에만 국한한다. 즉, 강세 언어인 영어의 경우는 다음 3.2절에서 보는 바와 같이 전혀 다른 방법을 사용할 것이다.

3.2. 영어 녹음 목록

영어의 녹음 목록은 자연스러운 음가를 얻기 위하여 다음과 같이 작성하였다. 첫째, 음의 길이와 높낮이를 고르게 하기 위하여 각운이 같은 단어들을 인접시켰다. 둘째, 화자의 망설임을 최소화하기 위하여 어근이 같은 단어들을 인접시켰다. 셋째, 음의 길이를 일정하게 하기 위하여 매 일정수의 강세마다 끊어 읽기 표시를 하였다.

예컨대, “wack, smack, snack, attack; attach, scratch, thatch, thatched”처럼 같은 [æk] 각운과 [æch] 각운을 인접시켰으며, 강세의 수가 4개일 때마다 끊어 읽도록 하였다. 그러면, 이러한 원칙에 의해 앞 단원의 영어 어휘표가 실제 녹음 목록으로 재구성된 바를 1)기준자음, 2)기준모음, 3)강세각운, 4)비강세각운의 순으로 제시해 보겠다.

그 첫 번째 유형으로서 아래 표13은 앞 단원에서 제시한 영어 어휘 목록중 CVCVC에 해당하는 단어만을 골라 녹음 목록으로서 재구성한 것이다. 이 유형의 녹음 목록은 해당 모음을 둘러싼 전후자음이 동일하므로, 기준자음으로서의 음성표본을 제공한다.

아래 표13의 녹음 목록은 앞서 표5의 자모음교체에 관한 어휘목록에서 대부분 추출하였으나, (01)번의 “popped”처럼 표8의 중성 자음군에 관한 어휘목록에서도 추출하였다. “popped”는 “pop”과 어근이 같으므로 인접 배치하여, 화자의 망설임을 최소화하였다. 또한 4-5음절마다 끊어읽기 표시 “;”를 하여 운율을 맞추었다. 다만, 위 CVCVC 단어는 각운이 같은 단어끼리 배열할 수가 없는데, 그 이유는 이미 단어내 초성과 중성이 같은 두 단어간에 각운인 VC마저 같다면, 이들 두 단어는 서로 완전히 동일한 형태가 될 것이기 때문이다.

표 13. 영어의 기준자음 녹음목록  
Table 13. The English prompts of reference consonants.  
(No. 01-03 of 6 samples<sup>13)</sup>)

No.	VC Prompts
01	peep, pip, pep, pap; pup, pope, pop, popped
02	teat, tit, tat, tot; tote, toot, loll, hull
03	kick, cake, keck, cock, coke; cook, Kirk, roar, rear

13) “keck”: vomiting sound

두 번째 유형의 영어 녹음 목록은 앞 단원의 어휘 목록 중 기준 모음에 관한 표6을 각운의 유사도에 따라 재구성한 것이다. 아래 표14에 예시된 바를 본다.

표 14. 영어의 기준모음 녹음목록  
Table 14. The English prompts of reference vowels.  
(No. 03-06 of 6 samples<sup>14)</sup>)

No.	VC Prompts
03	ear, air, are, or, our, err, erect, ergodic
04	add, odd, goodness, cabin; ought, about, us, house
05	beat, bit, bait, bet; bat, bott, bought, boat
06	boot, but, bit, bite; butter, button, bottom, bottle

이들 표14의 단어들은 (05)번의 "bat, bott, bought, boat" 처럼 각운이 비슷한<sup>15)</sup> 것끼리 인접 배치하여, 발화 시 음의 고저와 장단을 일정하게 하였다. 또한, 강세의 수가 4~5개 일 때마다 끊어읽기 표시 ";"를 하여, 발화 시 음의 길이를 일정하게 유지하였다.

위 표13과 표14에서는 기준자음과 기준모음을 목록화한 것이다. 이 두 유형을 제외한 나머지 단어들은 강세 모음을 포함하는지 아닌지에 따라 분류되어, 다음 표15와 표16에서 수록된다.

세 번째 유형의 영어 녹음 목록은, 영어의 전체 어휘 목록 중 강세 모음이 같은 단어를 모아 각운의 유사도에 따라 재구성한 것이다. 아래 표15에 예시된 바를 본다.

표 15. 영어의 강세모음 녹음목록  
Table 15. The English prompts of stressed vowels.  
(No. 01-04 of 6 samples<sup>16)17)</sup>)

No.	VC Prompts
01	had, cad, fad, sad, Zad; vat, that, chat, latter, shatter
02	dance, danced, gang, gangs, ganged; gags, gagged, gasp, gasps, gasped
03	ass, ash, rash, slash, splash; bath, thank, thanks, thanked
04	wack, smack, snack, attack; attach, scratch, thatch, thatched

위 표15의 녹음 목록은 앞 단원의 어휘표에서 [ae] 모음을 가진 모든 단어 중 이미 표13, 표14의 유형에서

수록된 형태를 제외한 나머지 단어를 모두 수록하였다. 이와 같은 방법으로 "소리다" 데이터베이스에서는 총 12개의 강세 모음을 표본화한다. 이들 각 모음별로 단어를 수록하되, 음의 고저와 장단을 일정하게 하기 위하여 위 (01)번 표본처럼 비슷한 각운끼리 인접시켰고, 강세의 수가 4~5개 일 때마다 끊어읽기 표시 ";"를 하였다. 또한, 화자의 망설임을 최소화하기 위해 "gang, gangs, ganged" 처럼 어근이 같은 단어를 인접시켰다.

마지막으로, 네번째 유형의 영어 녹음 목록은 앞 단원의 어휘목록 중 비강세모음이 같은 단어를 모아 각 모음별로 재구성한 것이다. 아래 표16에 예시된 바를 본다.

표 16. 영어의 비강세모음 녹음목록  
Table 16. The English prompts of unstressed vowels.  
(No. 01-03 of 3 samples)

No.	VC Prompts
01	error, offer, author, razor, mayor; inner, Bauer, dollar, copper
02	usher, butter, gather, trader, braver; sneaker, glimmer, dresser, creature
03	usurp, superb, caliber, indulger, conquered; midterm, nocturne, desert, hauberk

위 표16의 녹음 목록은 앞 단원의 어휘표에서 [axr] 모음을 가진 모든 단어 중 이미 표13, 표14, 표15의 유형에서 수록된 형태를 제외한 나머지 단어를 모두 수록하였다. 이와 같은 방법으로 "소리다" 데이터베이스에서는 3개의 비강세 모음 [ax], [axr], [ix]를 표본화한다. 그런데 영어의 특성상 비강세 모음의 존재는 항상 강세 모음에 인접해 있어야 하므로, 이들 어휘표는 항상 2음절 이상이다. 이들 비강세 모음을 수록할 때도 앞서 설명한 강세모음의 경우처럼, 음의 고저와 장단을 일정하게 하기 위하여 비슷한 각운끼리 인접시켰고, 강세의 수가 4~5개 일 때마다 끊어읽기 표시 ";"를 하였다.

요약하자면, 본 절에서 논한 "소리다" 녹음 목록은 음정과 장단을 일정하게 유지하기 위해 국어와 영어의 언어별 특성에 맞게 어휘표를 재배열하였다. 국어는 음절 단위 언어이고, 영어는 강세 단위 언어이다.

그 배열방법을 구체적으로 기술하자면, 1)국어 녹음 목록은 일정 음절 수마다 끊어읽기 표시 ";"를 하는 한편, 2)영어 녹음 목록은 일정 강세 수마다 끊어읽기 표시 ";"를 하며, 3)국어 녹음 목록은 반복어를 음운 단위의 처음과 끝에 넣어 음정을 고르는 한편, 4)영어 녹음 목록은 각운이 유사한 단어를 인접하게 하여 음정을 고른다.

이렇게 작성된 녹음 목록은 각 언어와 방언별로 적정 화자를 선정하여 읽힌다.

14) 'bott': botanical

15) 어휘표 작성 자체가 각운 형태와는 상관없이 만들어졌기 때문에 각운이 같은 단어만을 골라 인접시키는 일은 별도의 작업이며, 목록의 형태를 바꾸게 된다.

16) 'Zad': Zadkiel calendar

17) 'wack': bad

IV. 언어별 언어사회군에 따른 화자 분포의 선정

화자의 선정은 영어 데이터베이스, 국어 데이터베이스, 콩글리쉬 데이터베이스별로 방언, 성별, 연령을 고루 배치하였다. 영어 데이터베이스는 미국인 5명이 영어 녹음 목록을 읽도록 하였고, 국어 데이터베이스는 한국인 5명이 국어 녹음 목록을 읽도록 하였으며, 콩글리쉬 데이터베이스는 한국인 2명이 영어 녹음 목록을 읽도록 하였다.

콩글리쉬 데이터베이스의 필요는 일상적으로 언어구사 시 외래 어휘와 발음을 삽입 사용하기 때문이다. 또한, 본 한·영 동시조음 데이터베이스의 콩글리쉬를 한국어 화자의 영어 발음으로 하는 이유는, 한국인은 영어를 섞어서 쓰는 경우가 많은 반면, 미국인은 국어를 섞어서 쓰는 경우가 적기 때문이다. 국어 화자는 대부분 영어 몇 단어쯤은 알지만, 영어 화자는 국어 단어를 잘 모른다.

“소리다” 데이터베이스의 언어 구성과 화자 분포는 다음 표17과 같다.

표 17. “소리다” 데이터베이스의 화자 분포도  
Table 17. The Speaker diversity of SORIDA.

언어	화자 번호	방언	연령	성별
국어	1	제1방언 (중서부)	성인(43)	남
	2	제1방언 (중서부)	성인(38)	여
	3	제1방언 (중서부)	아동(9)	여
	4	제2방언 (동남부)	성인(30)	남
	5	제2방언 (동남부)	성인(31)	여
영어	1	제1방언 (백인)	성인(40)	남
	2	제1방언 (백인)	성인(41)	여
	3	제1방언 (백인)	아동(12)	여
	4	제2방언 (흑인)	성인(35)	남
	5	제2방언 (흑인)	성인(35)	여
콩글리쉬	1	제1방언 (중서부)	성인(43)	남
	2	제1방언 (중서부)	성인(38)	여

상기 도표에서 방언의 분포는 국어의 경우는 지리적인 경계로, 영어의 경우는 인종 사회군 별로 표본화함으로써, 각 언어사회의 특성에 맞게 구성하였다. 지구상 수많은 영어 사용국 중에서도 교류빈도가 높은 미국의 화자를 대상으로 하였다. 국어와 영어 공히, 사용자의 수가 많은 제1방언의 경우는 남녀성인과 아동을 포함한 3인의 목소리를, 사용자의 수가 적은 제2방언의 경우는 남녀성인만 포함한 2인의 목소리를 녹화하였다.

연령의 분포는 삼사십대의 성인을 주요 사용 연령층으로 추정하여 방언마다 2인씩 녹화하였으며, 제 1 방언의 경우에만 아동 1인의 목소리도 첨가하여 연령차로 인한

소리변이를 포함하였다. 아동의 연령은 어휘구사력이 충분히 발달했을 단계의 최소 연령을 택했다. 따라서 문자와 소리의 일치도가 높은 국어는 9세로, 문자와 소리의 일치도가 낮은 영어는 12세로 정하였다.

성별의 분포는 성인의 경우 남녀 음성특징이 뚜렷이 다르므로 남녀를 고르게 분포시켰으며, 아동의 경우는 남녀 음성특징이 비교적 비슷하므로 여아만 녹화하였다. 남아보다 여아를 택한 이유는 합성시 여아목소리의 선호도가 높기 때문이다.

이처럼 언어별, 방언별, 연령별, 성별로 고르게 분포된 화자들은 해당 언어의 전체 녹음 목록을 먼저 숙지한 후, 다음 시절을 이용하여 녹음하였다.

V. 녹음 및 컴퓨터 처리

녹음은 AKG Model C-1000 마이크와 Fostex Model D-5 디지털 녹음기를 사용하였다. 미국 Kansas State University의 Digital Signal Processing Laboratory에서 1 회 1인 1시간 정도씩 녹음하였다. 매회 단일 녹화시간은 시행착오에 소요되는 시간을 포함하여, 최대 2시간 이내로 국한하였다. 이는 녹음자의 음질과 억양을 일관성 있게 유지하기 위함이다. 또한, 같은 이유로, 동일 데이터베이스의 녹화는 당회 녹화 시간내에 처리하도록 하였다. 실험실은 방음이 되지 않았으나, 최대한 조용한 시간을 사용하여 잡음을 최소화하였다.

음성표본 추출속도는 48kHz로 하였으며, 총 녹화된 음성 데이터는 4시간 분량이다. 3개의 디지털 오디오 테이프를 사용하여, 국어, 영어, 콩글리쉬를 각각 다른 테이프에 저장하였다.

컴퓨터 처리는 22.05kHz와 16kHz의 두 종류 속도와, 16bit의 크기로 음성표본을 추출하여, 세 가지 PCM방식의 오디오 파일로 저장하였다. 오디오 파일 형태는 윈도우즈용 .wav와 모토롤라 방식의 .pcm과 인텔방식의 .nsp 세종류로 되어 있다. 음성 데이터베이스를 저장한 CD-ROM은 음성표본 추출속도와 오디오 파일형태에 따라 다섯 가지 종류로 제작되었는데, 각 경우마다 Compact Disk 1장에 4시간에 걸친 녹음 분량을 모두 수록하였다. 이중 음성분할이 가장 상세히 되어 있는 CD-ROM버전은 “소리다 1-1.5”로서 22.05kHz, 16bit, .wav의 형식의 오디오 파일 형태를 어절별로 분할 편집하였다.

VI. 결과 및 고찰

본 연구 결과물인 “소리다” 데이터베이스는 한·영 동시조음 데이터베이스로서 최초의 개발품일 뿐만 아니라 국어 음성합성, 국어 음성인식, 영어 음성합성, 영어 음성인식, 한·영 통역의 다섯 가지 연구용 기초 데이터베이스로 활용될 수 있다.

기능상으로는, 국어와 영어의 언어 특질에 적합하게, 전 음소 조합을 망라한 어휘표를 작성하고, 운용에 맞는

녹음 목록을 만들었으며, 각 언어사회에 맞는 화자를 다양하게 분포시켰고, 필요 데이터의 개수를 최대, 최적으로 포괄하였다. 연구 결과물인 CD-ROM "소리다 1-1.5"는 학문 사회의 발전을 돕기 위하여 개발된 것으로, 모든 연구자의 자유로운 복사 사용이 가능하다.<sup>18)</sup>

### 참 고 문 헌

1. Kim, Jong-mi, S. A. Dyer and D. D. Day, "Construction of a speech translation database," First LREC conf. proc., pp. 1071-1078, Granada, Spain, May 1998.
2. Young, S. J., J. J. Odell and P. C. Woodland, "Tree-based state typing for high accuracy acoustic modelling," Human Language Technology Workshop proc., ARPA, pp 307-312, 1994.
3. *Speech Resources Catalogue*, web site, European Language Resources Association, <http://www.icp.grenet.fr/ELRA/cata/tabspeech.html>, 1998.
4. *Speech corpora*, web site, Linguistic Data Consortium, USA, [http://morph ldc.upenn.edu/Catalog/by\\_type.html#speech](http://morph ldc.upenn.edu/Catalog/by_type.html#speech), 1998.
5. 정국, 구희산, 이찬도, 김중미, 한선희, "음성 인식/합성을 위한 국어의 음성-음운론적 특성연구", 한국음향학회지, 13권, 6호, pp. 31-44, 1994.
6. Zue, Victor W. and Stephanie Seneff, "Transcription and Alignment of the Timit Database", Paper presented at the Second Symposium on Advanced Man-Machine Interface through Spoken Language, Oahu, Hawaii, Nov. 1988.
7. *TIMIT: Acoustic-Phonetic Continuous Speech Corpus*, CD-ROM, Distributed by LDC, UPENN, <http://morph ldc.upenn.edu/ldc/catalog/html/speech.html/arpa.html>, 1990.

### ▲Stephen A. Dyer

Professor

Department of Electrical and Computer Engineering,  
Kansas State University

### ▲Dwight D. Day

Associate Professor

Department of Electrical and Computer Engineering,  
Kansas State University

### ▲김 중 미(Jong-mi Kim) 1959년 9월 14일생



1981년: 전북대학교 영어교육과 학사  
1983년: 미국 USC대학교 언어학과 석사  
1986년: 미국 USC대학교 언어학과 박사  
1987년~현재: 강원대학교 영어영문학과 교수  
1997년: 미국 캔사스 주립대학교 전자공학과 방문 교수

\* 주관심분야: 음성 분석 및 응용

E-mail : kimjm@cc.kangwon.ac.kr

18) 또한, 16 kHz의 표본 추출 속도로 된 음성 데이터베이스가 필요한 연구자는 음성분할이 안된 "소리다 2-2.1, 2-2.2, 2-2.3"을 사용할 수 있다. 단, 사용사, 학문사회의 규칙에 따라 출처 기록은 요구된다.