Computerized Sound Dictionary of Korean and English*

Jong-mi Kim**

ABSTRACT

A bilingual sound dictionary in Korean and English has been created for a broad range of sound reference to cross-linguistic, dialectal, native language (L1)-transferred biological and allophonic variations. The paper demonstrates that the pronunciation dictionary of the lexicon is inadequate for sound reference due to the preponderance of unmarked sounds. The audio registry consists of the three-way comparison of 1) English speech from native English speakers, 2) Korean speech from Korean speakers, and 3) English speech from Korean speakers. Several sub-dictionaries have been created as the foundation research for independent development. They are 1) a pronunciation dictionary of the Korean lexicon in a keyboard-compatible phonetic transcription, 2) a sound dictionary of L1-interfered language, and 3) an audible dictionary of Korean sounds. The dictionary was designed to facilitate the exchange of the speech signal and its corresponding text data on various media particularly on CD-ROM. The methodology and findings of the construction are discussed.

Keywords: bilingual sound dictionary, Korean and English, computerized pronunciation dictionary, Korean speech

1. Introduction

A compact bilingual sound dictionary has been compiled with a range of sounds across the span of Korean, English, and their L1-transferred speech.[1] This dictionary is a new attempt in that 1) it is a bilingual sound dictionary, 2) it includes an audible dictionary of Korean sounds, 3) it includes a sound dictionary of Konglish,[2] and 4) it includes a pro-

* This work was supported by Korea Research Foundation Grant (KRF 0898071-1-2).

** Dept. of English Language and Literature, Kangwon National University

[1] This bilingual sound dictionary is part of the SORIDA (Sound Rich Database) series that was first introduced in LREC-1998 (Kim, Dyer, and Day, 1998). Since then, SORIDA has developed thirteen different versions of its sound related database. An earlier version of this paper appeared in the Proceedings of Second International Conference on Language Resources and Evaluation held in Athens, Greece from May 31 to June 2, 2000. This paper has greatly been benefited from the academic and engineering expertise of its anonymous reviewers, Christopher Hogan and Robert E. Frederking of Carnegie Mellon University, Minsuk Song of Kwandong University, Myong-Wan Koo and Du-Seong Chang of Korea Telecom, Jin-Young Kim of Chonnam National University, and Ik-joo Chung of Kangwon National University.

[2] The word "Konglish" refers to a byproduct of English that has been modified, is socially

nunciation dictionary of the Korean lexicon in KORBET (Kim, 2000b), a keyboard compatible phonetic transcription system. The work is based on previous findings on monolingual pronunciation dictionaries and multi-lingual speech databases.

1.1 Background

When multimedia computers were not available to the public, typical sound dictionaries were written in orthographic, inaudible lists of sound transcriptions, commonly in the International Phonetic Alphabet (International Phonetic Association, 1993). Some examples are *A Pronouncing Dictionary of American English* by Kenyon and Knott and *A Korean Pronunciation Dictionary* by KBS.

As multimedia computers became more available to the public, such dictionaries like Merriam-Webster, Cambridge, Longman, etc. began to include audio forms of the words. The primary purpose of these dictionaries is not the sound itself, but the lexical description and the audio recording of lexical entries by one or two speakers. Each lexical entry contains a semantic interpretation and a sample voice recording. Korean dictionaries have not yet included audio forms.[3] In contrast, SORIDA's current dictionary includes several representative variations of sounds in different allophonic environments.

The development of the computer also invoked the need for pronunciation dictionaries in which sound forms are represented, stored, and searched by any ordinary computer keyboard. The translation of written sound symbols into audible sound waves is no longer the exclusive concern of phonologists, but the extended concern of engineers, secretaries, businesspersons, travelers, and others. One such example is the electronic dictionary by Carnegie Mellon University, *CMUdict.0.6d.1.* In this dictionary, all allophonic symbols are represented by the lowercase Roman letters on the ordinary keyboard. The phonological units are represented by spaces or abbreviations, and the degree of stress is specified by numbers. In Korean, however, such a keyboard-compatible pronunciation dictionary has not appeared until this first attempt.[4] The present bilingual dictionary includes the two written pronunciation dictionaries; *CMUdict.0.6d.1* for the English lexicon, which is free to use, and the newly-built pronunciation dictionary of the Korean lexicon. A Konglish dictionary has not appeared at all until this project, whether the entry or annotation is in audio or written form.

recognized, and is typically used by the non-fluent Korean speakers of English. The result is a new mixture of sound that is a combination of Korean phonology and English vocabulary. The term covers both accent and new-breed of vocabulary.

[3] Audio recording of a lexical dictionary is underway by Younghee Chae (Personal correspondence).

[4] A Korean dictionary has been compiled by Sang-Oak Lee (1995), in which the phonetic transcription system includes such diacritic markers as the breve and apostrophe. This dictionary follows the McCune-Reischauer system of Romanization, which was officially adopted by the government from 1984 to 2000.

Speech variations on a multilingual basis have been collected into speech databases, mainly for the engineering application of speech processing; namely, ACCOR, BDSONS, EUROM1, M2VTS, MULTEXT, ONOMASTICA-COPERNICUS, TED in Europe and OGI Multilanguage Corpus, CALLHOME, and CALLFRIEND in the United States (On-line speech corpora, 2001; Speech and related resources, 2001). Korean and English telephone speech data are included in the OGI Multilanguage Corpus and CALLHOME. For clear speech, Kim, Dyer, and Day (1998, 1999) have collected a bilingual speech database. This speech database has further been developed into the present dictionary with the addition of 1) the pronunciation dictionary of the Korean lexicon as mentioned above, 2) the phone dictionary of phonetic contexts and its corresponding audio registry, and 3) the annotation files of the manual time alignment for representative phones.

1.2 Objectives

The proposed bilingual dictionary aims to compile a compact and yet balanced coverage of sound variations for phonological, biological, dialectal, and cross-linguistic reasons in order to provide a phonetic resource for cross-linguistic research or application in Korean and English. We would like to know, for instance, "How many representative variations exist for the voiceless velar stop [k] in English and Korean?" To provide an adequate phonetic resource, the variations of the given sound are expected to cover 1) allophones and coarticulations, 2) male and female speech, 3) adult and child speech, 4) major dialects, 5) L1-transferred utterance of Konglish, and 6) bilingual speech of Korean and English. In order to achieve these specifications, the features and methodology had to be newly defined and developed.

The bilingual dictionary consists of four consecutive file sets. They are properly indexed and integrated into a single dictionary: 1) The audio file sets are read by 12 speakers in Korean, English, and Konglish words (see Section 4). 2) The label file sets of the phonetic phases are dependent upon the audio file sets. 3) The phone dictionary of Korean and English is indexed for the corresponding audio and label files. 4) The pronunciation dictionary of the Korean and English lexicon is transcribed in the keyboard-compatible alphabets. The phone information is linked to the phone dictionary and the corresponding audio and label files.

The data is categorized into a directory hierarchy for easier accessibility and use with a wide variety of computer systems. The data fits into a single CD with representative sound information for a broad range of individuals and supply researchers in speech-related fields.

The intended usage of this bilingual dictionary is for 1) cross-linguistic and acoustic research, 2) application to speech recognition, synthesis, and translation, and 3) foreign language education including pronunciation exercises. The languages of Korean and English are typologically so distinct as to provide rich phonetic variations. In addition, these languages meet the practical needs of people in Korea, where English is the most widely spoken foreign language.

We will discuss 1) the major features of the dictionary, 2) the phonological and speaker variations, 3) the organization of the data into digital audio files, 4) the phone information with respect to the pronunciation dictionary, and 5) the annotation of the audio files in terms of the time alignment of the acoustic events.

2. Sound Reference

The dictionary features bilingual sound reference for computer use. This contrasts with other conventional dictionaries that feature the book-prints of the monolingual lexical reference. The contrast is between "sound-based" and "lexeme-based" for expositional convenience. "Sound-based" refers to a dictionary whose entries are sounds and contain the description and variation of the sounds. This contrasts with a "lexeme-based" dictionary whose entries are lexemes and contain the description and variation of the lexical meaning. The audio forms of this sound-based dictionary are digitally saved, transcribed with alphabets on the ordinary keyboard, and may be operated by most ordinary sound cards in personal computers.

Table 1 shows samples from the phone dictionary of Korean, where the entries are sounds.

Table 1. Sample data entries of the phone dictionary of Korean. Entry phones are specified for the preceding and the following phones, and the context. Each sample ID number represents a speech sample. The same ID appears when the context, i.e., the speech sample, is identical. The transcription follows the conventions of KORBET. [1]The symbol "-" denotes the preceding sound. [2]The symbol "+" denotes the following sound. [3]*k1pam* denotes the Korean speaker #1 of the primary dialect, adult and male. [4]*k2paf* denotes the Korean speaker #2 of the primary dialect, adult and female. [5]*k3pcf* denotes the Korean speaker #3 of the primary dialect, child, and female. [6]*k4sam* denotes the Korean speaker #4 of the secondary dialect, adult and male. [7]*k5saf* in the chart denotes the Korean speaker #1 of the primary dialect, adult and male.

Entry phone	[1]*Pre phone (-)*	[2]*Post phone (+)*	*Recording Context*	*Sample ID of Audio Files*				
				[3]*K1pam*	[4]*K2paf*	[5]*K3pcf*	[6]*K4sam*	[7]*K5saf*
d	-m	+i	dh i m d i	11311	21311	31311	41311	51311
d	-m	+e	dh e m d e	11312	21312	31312	41312	51312
dh	-pau	+eu	dh eu m d eu	11314	21314	31314	41314	51314
dh	-pau	+eu	dh eu m d eu	11314	21314	31314	41314	51314
e	-dh	+m	dh e m d e	11312	21312	31312	41312	51312
e	-d	+pau	dh e m d e	11312	21312	31312	41312	51312
i	-dh	+m	dh i m d i	11311	21311	31311	41311	51311
i	-d	+pau	dh i m d i	11311	21311	31311	41311	51311
m	-e	+d	dh e m d e	11312	21312	31312	41312	51312
m	-eu	+d	dh eu m d eu	11314	21314	31314	41314	51314

The proposed dictionary is "sound-based," in that all sound entries are listed for the different allophonic environments, and all recording materials are spoken by several representative speakers. Table 2 presents the difference between lexeme-based and sound-based dictionaries.[5]

[5] Among the various phonetic environments that the sound dictionary may deal with, the proposed dictionary focuses on the segmental environment and diminishes the prosodic variations. The reasons are two fold. 1) Dictionaries, by common practice, are word-based, rather than sentence-based. 2) The prosodic rules are more varied and thus inadequate to provide agreeable reference materials in this compact dictionary.

Table 2. "Sound-based" versus "lexeme-based" dictionaries. The entries are "phonemes" with various speech samples in a sound-based dictionary.

Type *Criteria*	***Sound-based Pronunciation Dictionary***	***Lexeme-based Pronunciation Dictionary***
Recording and Searching Entries	Phonemes in Allophonic Contexts	Words in Lexicon
Speech Variety	Various Speech Samples per Phoneme Entry	One Speech Sample per Lexical Entry
Examples	SORIDA Dictionary of This Project	Merriam-Webster, Cambridge, Longman

Only the sound-based dictionary is capable of answering a question like, "How many representative allophonic variations exist for the phoneme [k] in English and Korean?" It can present variations in terms of 1) two major dialects of each language, 2) Konglish, 3) variations between the two genders, 4) adult and child speakers and 5) diverse allophonic distributions.

The lexeme-based pronunciation dictionary does not present adequate pronunciation samples for the following reasons:

1) Insufficient numbers of allophonic samples are represented, since the dictionary usually omits the audio recording of inflected forms of a word. Inflected forms, however, provide important coarticulation effects, as in the example of four consonants in a row in the syllable-final position in the English word *texts*. The present dictionary includes these forms.

2) Excessive overlapping of allophones occurs because lexical items prefer unmarked segments. The lexicon is inherently built in terms of meaning variation, and not sound variation.

3) Dialectal variation is excluded by limiting the recording to a selected, and usually single, speaker. The disk space is left for only one speech sample per lexical item, because the priority is given to lexical description.

The sound-based pronunciation dictionary, however, contains 1) diverse allophonic samples by positioning phonemes in all phonotactically possible locations, 2) reduced sound overlapping by taxonomically arranging adjacent segments and deleting identical allophones arising from phonological changes, and 3) speech variations by selecting the speakers from representative dialectal and biological groups in each language.

The sound contents of the dictionary are enriched in both 1) phonological phenomena and 2) speaker distribution. The phonological variations are collected in terms of phonemes, allophones, syllables, and stress. The sound variation among speakers covers cross-linguistic, dialectal, and L1-transferred variations along with speech from different gender and age groups.

The total number of recorded allophones is 17,825 for Korean (3,565 triphones × 5 speakers), 12,650 for English (2,530 triphones × 5 speakers), and 5,060 for Konglish (2,530 triphones × 2 speakers). More phones are recorded for the Korean database that uses artificial words, whereas the English one uses only real words.[6] The phone tokens in this phone dictionary are searchable from the audio registry in Section 5, and from the pronunciation dictionaries of the lexicon in Section 6. The creation of any reasonable sized speech corpus is labor-intensive. Therefore, the dictionary was designed to balance utility and manageability, containing small amounts of speech from a relatively diverse speaker population and a range of phonetic environments. The two sections following discuss these considerations.

3. Phonological Considerations

The bilingual sound dictionary reflects the phonological differences of the two languages. The phonological aspects are considered in preparing the recording materials. The recording material should include tokens of not only the compared sound, but also the preceding and the following sounds. In order to do this, 1) CV templates allocate different sequences of sounds, 2) all possible combinations for different phonemes are arranged within the template, and 3) overlapping of the allophones caused by phonological assimilation and deletion is eliminated.[7] The following is an example of the taxonomic arrangement of permissible sound sequences and their corresponding recording material.

[6] The method of using artificial words is first employed in EUROM speech database (Speech and related resources, 2001), and adopted in this project for the Korean part only.

[7] Phonological assimilation phenomena enable the data size to decrease, as in Korean examples of tensing, liquid assimilation, and vowel reduction. For instance, the Korean sample [minmi] combines two input forms of /mitmi/ and /minmi/. Similarly, the pattern [ninni] is from /nitni/ and /ninni/. This is because an obstruent is nasalized in front of nasal consonants in Korean.

Table 3. Representation of phonological aspects in recording material. "C" stands for one consonant, and "V," one vowel. Different consonants and vowels are combined with the given segment in a specified position within the template. English uses real words and Korean, the artificial ones. The transcription follows the conventions of CMU symbol set (TIMIT, 1990) and KORBET.

Phonological Aspect	*Template Example*	*Recording Material*
Coarticulation of the English Vowel Phoneme /ae/	(C)VC	pap, tat, attack, cad, tact, babble, dab, dad, gag, chat, attach, jazz, badge, ma'am, Nan, gang, Larry, latter, Al, rash, rare, fad, half, thatch, bath, sad, ass, shatter, ash, had, vat, that, zad, jazz, yam, wack
Coarticulation of Intervocalic Consonants in Korean	VbVdVgVjVsV	ibidigijisi, ebedegejese, aebaedaegaejaesae, eubeudeugeujeuseu, eobeodeogeojeoseo, abadagajasa, ubudugujusu, obodogojoso
Allophones of the Korean Phoneme /b/	bhVlbV	bhilbi, bhelbe, bhaelbae, bheulbeu, bheolbeo, bhalba, bhulbu, bholbo
Syllable Onset Clusters	kw, kr, kl, sk, skw, skr, skl	quest, cry, clue, sky, squash, scratch, sclerosis
Unstressed Vowel /er/	VCer	usher, butter, gather, trader, braver, sneaker, glimmer, dresser, creature

In English, these templates are taxonomically arranged for 14 vowels and 24 consonants, totaling 336 samples. Vowels in the templates include both stressed and unstressed forms.[8]

The recording prompts for Korean can be tokenized from a longer template than those for English, because artificial words are used. The Korean alphabet "Hangeul" provides a reliable sound-symbol correspondence; most dictionaries directly use the alphabet for phonetic transcriptions. The speakers are provided with the recording prompts in the Korean alphabet. The use of artificial words did not induce any hesitation on the part of speaker when pronouncing them.

The cross-linguistic phonological variations are considered in terms of 1) phonemes, 2) allophones, 3) syllables, and 4) stress. All languages use different sets of phonemes. As an example of cross-linguistic variation, several English phonemes are absent in Korean, which prompts Korean learners of English to substitute them for the most similar Korean phoneme. For instance, the English phoneme [th] is absent in Korean, therefore, the English word think would likely be pronounced as sink. The bilingual dictionary supplies various phonemes of both languages, as in such minimal sound pairs. Table 3 above includes the

[8] The 14 vowels for taxonomic arrangement include stressed central vowels (ah, er) and unstressed central vowels (ax, axr), the rest of the tense vowels (iy, ey, ow, aa, ao, uw) and lax vowels (ih, eh, ae, uh). The transcription follows the CMU convention (TIMIT, 1990).

English vowel phoneme /ae/ as an example. The recording prompts for English are read by both English native speakers and Korean speakers in order to compare the L1-transferred speech forms to the target forms.

Allophones are determined by language-specific phonological rules. For instance, the allophonic change of [l] in Korean causes learners to pronounce the English word *light* indistinguishably from *right*. Table 3 includes examples of tokenizing different allophonic and coarticulation forms in the recording material. The basic technique is to concatenate all the permissible sequences of phonemes allowed by the phonotactics of both languages.

For instance, a recording sample, [bhilbi] has word-initial [bh], the vowel [i], syllable-final [l], and syllable-initial [b]. Thus, the template covers five different allophones derived from three different phonemes. In particular, we find two allophones of lenis bilabial stop /b/, one in word-initial position followed by [i], and the other in between [l] and [i].[9] The former is a slightly aspirated voiceless stop and the latter is a fully voiced stop. The use of allophonic rules may predict many bound allophones as such, yet it is often insufficient to predict even the commonly observed surface phonetic forms, as in Korean epenthetic-s (Kim, 1991, 1992). All the pronunciations in this bilingual dictionary were hand-checked.

Syllable forms differ from language to language. Learners may therefore pronounce the English monosyllabic word *strike* as [seu-teu-ra-i-keu] with five syllables. The bilingual dictionary provides all possible syllabic forms in both languages. English consonant clusters are listed in the onset and coda positions; up to three consonants in a syllable-initial position and up to four in a syllable-final position. Table 3 exemplifies the tokenization of some syllable initial clusters into the recording material. In practice, the recording material for English contains a total of 39 kinds of onset clusters and 169 kinds of coda clusters. Thus, roughly 90 percent of all possible cluster forms have been presented in English. The remaining 10 percent include the clusters that the Korean speakers do not pronounce in their dialects.

In Korean, on the other hand, only one consonant is allowed in both syllable-initial and syllable-final positions. Accordingly, the recorded materials include at most two consonants in a row; one for the coda of the preceding syllable and the other for the onset of the following syllable.

Stress variation is modeled only for English and not for Korean. This is due to the

[9] The symbol [bh] expresses a slightly aspirated allophonic variation of the lenis phoneme /b/ in Korean. According to KORBET, there are five symbols to represent the phonetic quality of Korean bilabial stops: [bh] for the slightly aspirated variant of the lenis stop /b/ in word-initial position, [b] for the voiced variant of the lenis stop /b/ in inter-sonorant position, [p] for the voiceless unaspirated unreleased variant of all bilabial stops in syllable final position, [ph] for the heavily aspirated variant of the fortis stop /p/ in syllable initial position, and [pp] for the unaspirated tense variant of the tense stop /p/ in syllable initial position. All these allophonic variations are predicted by phonological rules and their speech samples are represented in this dictionary.

presence and absence of stress in the underlying structure of the English and Korean phonology.[10] Korean rhythm depends on the number of syllables, whereas English rhythm on the number of stresses. Hence, the proposed bilingual dictionary provides the variation of both stressed and unstressed English vowels, but needs not record stress in Korean. Some examples of an unstressed vowel are provided in Table 3.

Another aspect of stress-timing is related to artificial words. An English native speaker, in reading an artificial word, would not know where to put the stress and whether to pronounce the vowels as tense or lax. The bilingual sound dictionary uses real words in English to minimize the speaker's hesitation, and artificial words in Korean to maximize the coarticulation variation. Repetitive and predictable patterns of tokens caused the naturalness of coarticulation. Table 3 includes examples of intervocalic consonants and the bilabial stop /b/, where the vowels within a token are identical, and change into another vowel of the given order for the next token.

In addition to the phonological variation discussed so far, the dictionary includes speech variation from cross-linguistic, dialectal, and physical differences.

4. Speaker Variation

The data covers the sound variation from speakers of different languages, dialects, ages, and genders. Part of the methodology here is described in brief, as it has been developed in Kim, Dyer, and Day (1998, 1999). The speaker information is given below.

[10] Some lexical accents in Kyeongsang dialects are not yet included in this compact version of the bilingual dictionary. This mirrors pronunciation dictionaries of the Korean lexicon published up to the present. Some stress patterns in Kyeongsang dialects are expected to have been included in the recorded material, though it was not designed for that purpose.

Table 4. Speaker information for the audio registry. Reproduced from Kim, Dyer, and Day (1998).

Language	*Code*	*Dialect*	*Age*	*Gender*
Korean	1	primary (mid-west region)	adult (43)	male
	2	primary (mid-west region)	adult (38)	female
	3	primary (mid-west region)	child (9)	female
	4	secondary (Kyongsang region)	adult (30)	male
	5	secondary (Kyongsang region)	adult (31)	female
English	1	primary (middle class)	adult (40)	male
	2	primary (middle class)	adult (41)	female
	3	primary (middle class)	child (12)	female
	4	secondary (working class)	adult (35)	male
	5	secondary (working class)	adult (35)	female
Konglish	1	primary (mid-west region)	adult (43)	male
	2	primary (mid-west region)	adult (38)	female

The language variation includes 1) cross-linguistic registry of English and Korean phonemes, and 2) the L1-interfered English spoken by the Korean learners. The Konglish database is made up of Korean speakers of both genders reading English words. English speakers reading Korean words have not been modeled for this compact version of the bilingual dictionary. This is because Korean speakers use more English words in their verbal activity than vice versa.

The dialects are classified into social classes in American English, and geographical regions in Korean. These parameters are chosen because of the considerable difference between allophones and the large number of speakers of the two groups. The dialectal variance of social classes in American English is extensively studied by Labov (1972) and others. A difficulty was to recruit speakers who equally represent all dialect regions, given the time constraints and the compact size of the data. All dialects are represented by adult speech of both genders. Only the primary dialects of both languages contain child speech of one gender.

The audio registry is standardized with technical support to make the samples vary by both phonological and speaker-related factors.

5. Audio Registry

Construction of the audio registry necessitates 1) reducing the size of audio data to realistic numbers, 2) controlling pitch and tempo by the arrangement of the recording prompts, and 3) standardizing the audio file format. The amount of recording material is reduced to occupy approximately 300 megabytes so that all the content of the bilingual dictionary can easily fit onto a CD of 650 megabytes for wider distribution. Down-sampling waveforms to a 16 kHz sampling rate conserves the disk space.

We tried to reduce the amount of audio data for recording vowel combinations. Because the primary focus in designing the dictionary was the cross-linguistic comparison of the phones, emphasis was placed on providing consonantal combinations that can be uneasy for a computer to recognize in both languages. The combinable variations of monophthong sequences are kept only to the representative samples in Korean because each monophthong waveform occupies one syllable in the syllable-timed language, which therefore is acoustically salient, and its spectral pattern is easy for the computer to recognize. English is more complex, since vowel combination always triggers the stress reduction of one constituent vowel. We thus recorded for English data the samples of both stressed and unstressed vowels. The recorded monophthong samples are designed to retain vowel quality without being much affected by the neighboring consonants. Therefore, the onset and coda consonants are filled with identical consonants, if possible, or alveolar obstruents [d, t, s], which least affect the formant transition.[11] Each diphthong occupies one syllable, and thus is represented as one vowel unit.

The regulation of pitch and tempo abides by the rhythm of the two languages, although both English and Korean have typologically different prosodic rhythms. To standardize the recorded material, pitch and tempo in the recording were controlled in terms of the rhythmic arrangement of recording prompts. Rhythmic adjustments depend on language typology. English is a stress-timed language and the recorded words are arranged in terms of the similarity of the rhymes and morphemes. This minimizes the speakers' hesitation and regulates pitch and tempo. On the other hand, Korean is a syllable-timed language and the recording prompts are arranged in terms of the syllable numbers. Each "breath-group" takes only a designated number of syllables. Some examples of English recordings, then, are *hut, cut, but, gut, shut; rust, rusts, rush, rushed,* where *-ut* and *-us* rhymes are arranged together with the pause position marked with a semi-colon every 4-6 syllables. Examples of Korean words are *chinchi, chenche, chaenchae, cheuncheu, cheuncheu; cheoncheo, chancha, chunchu, choncho, choncho.* Here, there is a pause approximately every ten syllables. The dummy (or repeated) duplicates are inserted before and after a pause, where irregular pitch and tempo might occur.[12] These duplicates are deleted when acquiring audio data.

[11] The sound [h], least affects the transition, but is not suitable because it is often deleted when adjacent to a vowel.

[12] In a phonological phrase, pitch begins high, then settles to medium, and ends either high for a non-final phrase, or low for a declarative ending (Koo, 1986).

Other techniques are employed in recording and sampling in order to standardize samples and make them compatible with different database formats. The practice-reading rehearsal is conducted to elicit full coarticulation and to minimize hesitation. Other efforts to lessen the speakers' anxiety and the "dead-room effect" include the use of an ordinary laboratory room and a microphone instead of a soundproof room and direct computer input. Recording was done with an AKG Model C-100 microphone, which was directly digitized at a sampling rate of 48 kHz using a Fostex D-5 DAT recorder.

The audio data were then down-sampled to 16 kHz and 22.05kHz for audio files in the computer. The combination of 48 kHz sampling and 16 kHz acquisition reduces information loss when formatting the audio files. The 16 kHz audio samples are in wide circulation, while 22.05 kHz ones are requested by some PC users. The resolution is 16 bits.

The audio samples are encoded in three file formats: *.wav* files for Windows PCM header, *.nsp* files for CSL header, and *.pcm* for headerless raw PCM data whose byte order is (msb, lsb). The most accessible file format to date seems to be *.wav* with a sampling rate of 16 kHz and resolution of 16 bits. For the Windows PCM waveform, all *.wav* formatted files follow the RIFF (Resource Information File Format) specification. Most of the special information bits are saved with the wave file in this manner. The standard Windows PCM waveform contains PCM coded data, which is a truly uncompressed pulse code of modulation-formatted data. This type of audio file form is also easily read by other types of computers.[13] A sample audio file in *.nsp* format is presented in Section 7.

The audio registry consists of the speech samples of the phone reference in Section 2. The audio registry consists of 3,395 artificial words for Korean (679 words x 5 speakers), 3,415 real words for English (683 words x 5 speakers) and 1,366 artificial words for Konglish (683 words x 2 speakers). The phone information of any real word in the lexicon is to be represented in the following pronunciation dictionaries of Korean and English.

6. Pronunciation Dictionaries of Lexicon

A new pronunciation dictionary of Korean vocabulary is constructed where the transcription system uses the keyboard-compatible phonetic alphabet. For its English counterpart, *CMUdict 0.6.1* is adopted without any alteration. The aim was to create a dictionary of words whose allophonic information is searchable in the phone dictionary, and whose phonetic transcription is compatible with any software that recognizes ASCII characters. The word selection was verified and modified by five college students based on frequent usage. If none or few of them were familiar with a given word, the word was

[13] The audio registry of the *.nsp* and *.pcm* file formats is not supplied with extensive modification and annotation to create the dictionary.

eliminated from the dictionary entry. Consider the following examples in Table 5, which are transcribed by KORBET.

Table 5. Sample data entries of the pronunciation dictionary of Korean lexicon. Each lexeme is transcribed both in the Korean alphabet, "Hangeul" and in KORBET. Each coarticulation unit is tokenized for the preceding and following segments. These triphones[14] are searchable in the phone dictionary in Section 2. The symbol "-" denotes the preceding sound. "+" denotes the following sound. "pau" denotes pause.

Lexicon	***Transcription***		***Allophones***						
	Hangeul	***KORBET***	***1st***	***2nd***	***3rd***	***4th***	***5th***	***6th***	***7th***
견학	견학	gh yeo n h a k	gh -pau +yeo	yeo -gh +n	n -yeo +h	h -n +a	a -h +k	k -a +pau	
견학(2)	겨낙	gh yeo n a k	gh -pau +yeo	yeo -gh +n	n -yeo +a	a -n +k	k -a +pau		
반가	반가	bh a n g a	bh -pau +a	a -bh +n	n -a +g	g -n +a	a -g +pau		
방아	방아	bh a ng a	bh -pau +a	a -bh +ng	ng -a +a	a -ng +pau			
순리	술리	s u l l i	s -pau +u	u -s +l	l -u +l	l -l +i	i -l +pau		
울산	울싼	u l ss a n	u -pau +l	l -u +ss	ss -l +a	a -ss +n	n -a +pau		
맛있다	마시따	m a sh i tt a	m -pau +a	a -m +sh	sh -a +I	i -sh +tt	tt -i +a	a -tt +pau	
맛있다(2)	마싣따	m a sh i t tt a	m -pau +a	a -m +sh	sh -a +I	i -sh +t	t -i +tt	tt -t +a	a -tt +pau
맛있다(3)	마디따	m a d i tt a	m -pau +a	a -m +d	d -a +I	i -d +tt	tt -i +a	a -tt +pau	
맛있다(4)	마딛따	m a d i t tt a	m -pau +a	a -m +d	d -a +I	i -d +t	t -i +tt	tt -t +a	a -tt +pau

The phonetic transcription of Hangeul is also generally used for phonetic transcriptions

[14] A triphone refers to a phone specified for the preceding and the following phones.

of other Korean dictionaries. Additional phonetic transcription of KORBET supplies allophonic information as in the examples of [gh yeo n a k] "a field trip for study" and segmental information as in the example of [bh a n g a] "a noble family" and [bh a ng a] "a mill." The transcription is automatically derived from a software program, and hand-corrected to increase the precision level.[15] In a number of cases, we referred to the authority *A Korean Pronunciation Dictionary* by KBS (1993). The English equivalents for similar phones are given in a separate document file for sound comparison.

Among several transcription systems of Korean including those by Chung (1994, 1999) and Chin (2000), KORBET has been chosen because of the following useful features: 1) the use of the ordinary lower case Roman alphabet on the keyboard, 2) the use of the governments newly adopted Romanization system as long as it does not conflict with allophonic representation, 3) the detailed expression of phonetic quality compared to the official Romanization system, and 4) the representation of phonological units, *i.e.*, a space for phone boundary and the abbreviation *"pau"* for a word boundary. The notation of phone boundaries enables individual phones to be adequately indexed for the corresponding phone entry in the phone dictionary (Section 2).

The amount of lexical entries in the pronunciation dictionary is 9,183 for Korean, and 129,804 for English. A Konglish equivalent does not exist, because it is not an independent language with its own lexicon.

The pronunciation dictionaries of the Korean and English lexicon refer to matching phones in the phone dictionary (Section 2), and then their corresponding audio files are found in the audio registry (Section 7).

7. Annotation of Audio Registry

The speech waveform is time-aligned with the acoustic-phonetic sequence on the phone level. A *time-aligned phonetic transcription* relates the lexical representation of words as in Section 6 to their acoustic realizations. Annotation of audio files is done for the time alignment of acoustic events on the phone level. The annotation is manually checked for 1) representative samples for all speakers, and 2) all samples of one designated speaker (K1PAM) for Korean. The representative samples cover at least 10 occurrences of each phoneme. Remaining samples in Korean are time-aligned automatically with the acoustic-phonetic sequence, using an alignment program "Festival Synthesizer," developed at Carnegie Mellon University in USA.[16]

[15] A conversion program was kindly provided by Du-Seong Chang of Korea Telecom.

[16] The automatic alignment was kindly performed and supported by Christopher Hogan and Robert E. Frederking at CMU. Hand correction of the aligned transcription was not done

Hand annotation of audio files is performed at first by using the CSL program. We followed the procedure of segmenting files into word units, time aligning phonetic phases, and labeling phonetic transcription. The following figure is an example of annotation in *.nsp* audio files.

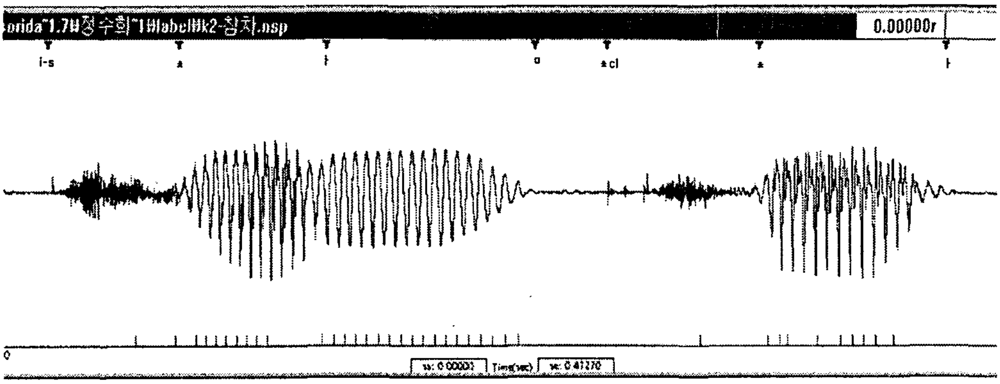

Figure 1. The annotated audio file of the word "chamcha" in .nsp format. Acoustic phonetic labels are time-aligned in the top portion of the figure. The waveform represents the speech from a female adult Korean speaker.

The following alignment protocols are followed when it is difficult to find explicit rules and practical considerations are required. 1) The beginning of soundwaves is marked as "*i-s*" (Initial Suspension of air-flow) that distinguishes from the preceding silence. 2) The phonetic phases are labeled at the end of the time sequence. 3) Word-medial voiceless stops include an independently labeled closure period as in *ㅂcl, ㄷcl, ㄱcl,* etc. that specifies the phonetic quality of the following release. 4) Clusters are separated by 1/2 length, if no transition mark is noticeable. A more extensive study is required to make any theoretical claim on this heuristic rule. Hangeul transcription is used for Korean to minimize inconsistency among the human transcribers.[17] The conversion table into KORBET is provided in a separate file, allowing the literal translation to the pronunciation dictionary of the Korean lexicon.

.nsp audio files are read by a limited number of computer programs including CSL or Multi-Speech by Kay Elemetrics. The labeled *.nsp* files are converted into *.wav* audio files and *.txt* label files for the convenience of a wide range of users.[18]

Annotation of audio files consists of time alignment of different phonetic phases. The

due to time constraints.

[17] To insure the consistency and accuracy of labeling and alignment, the transcribers conducted 1) cross-checking of the others' alignment, 2) regular meeting on alignment, and 3) constant refinement of the alignment protocol.

[18] To decode *.nsp* tag information, a program was kindly provided by Jin-young Kim of Chonnam National University.

label files have the same name with a different extension. The following is the example of an annotation file of the audio registry "*chamcha.wav*" of K2 speaker.

```
519   i-s
1438  ㅊ
2479  ㅏ
3947  ㅁ
4453  ㅊcl
5525  ㅊ
6824  ㅏ
```

The time is aligned in X-waves format, where the number indicates the time in milliseconds at the end of a given segment, starting with zero. The transcription follows Hangeul, and specifies more detailed phonetic phases, such as *i-s* for the beginning and ㅊ *cl* for the ㅊ segment in syllable final position.

The X-waves format of time alignment is chosen for its practicality for the UNIX system, instead of the NETSHOW format where the time is specified for the beginning of a given segment. Accordingly, the X-waves format can be transformed directly into Microsoft Excel format so that the time alignment information can be easily accessed by most PC based programs.

There are 904 manual time-alignment annotation files for Korean (57 words x 4 speakers + 679 words x 1 speaker), 335 for English (67 words x 5 speakers) and 134 for Konglish (67 words x 2 speakers). The label files are used for the identification of the phonetic quality of individual phones in a specified context. It is rather straightforward to translate the label files into the orthographic transcription of the lexicon pronunciation dictionary (Section 6). The alignment of a phonetic transcription with the corresponding speech waveform is essential for using the dictionary in speech research, since it provides direct access to specific phonetic events in the waveform. The time-aligned sound samples are also useful in speech synthesis and recognition applications.

8. Concluding Remarks

A "sound-based" pronunciation dictionary in Korean, English, and Konglish was developed for comprehensive sound reference. The Korean portion required more effort than the English one since the pronunciation component dictionary and transcription system had to be newly developed. The dictionary also includes typical erroneous pronunciations of non-native speakers, which is useful data for "error analysis" in the field of second language pedagogical research. Error awareness may benefit the learners in improving phonetic performance.

The sound dictionary is composed of four consecutive file sets in English, Korean, and Konglish, which are indexed with one another and integrated into a single dictionary. They are 1) a phone dictionary of various coarticulation environments, 2) a pronunciation dictionary of the lexicon, 3) audio files of the speech samples, and 4) annotation file sets of the audio files. The data size is less than 600 megabytes and fits on a standard CD.

Some new methodology has been explored in terms of the tokenization of representative speech samples and the standardization of data structure and formats.

The proposed sound dictionary achieves 1) parallel comparison of the target sound with native sound and L1-transferred sound, 2) incorporation of major phonological aspects of the target language, and 3) balance of speech variation among dialects, gender, and age groups. It is a useful sound reference tool for these languages. It contains numerous alternate sounds found in a number of phonetic contexts.

The Konglish database is useful for both speech engineering and language education. In speech engineering, the Korean speech context includes English words more naturally in Konglish form than the native English form. This applies to speech recognition, synthesis, and translation. In pronunciation education, the learners input speech can be pattern-recognized better with the Konglish database than the native English one. Accurate recognition enables the computer to give better suggestions and lessons to learners on improving their speech.

Scientific evaluation is left undone for the symmetry of data and the accuracy of annotation in this bilingual dictionary. The question of an evaluation method is itself left unresolved. Manual inspection is attempted by repeated crosschecking, which does not provide evaluation beyond the framework of this dictionary. A statistic evaluation of the acoustic and linguistic contents is still needed.

Useful data that were not included in this compact dictionary were the following. 1) The speakers in this project are primarily selected for research convenience of repeated recording for data replacement. Thus, common sound variation among speakers was observed, as needed in the phonetic research and speech recognition components. Additional recording of professional announcers are desired for the purpose of language education and speech synthesis. All speakers recorded in this project, however, consider their own voices easily comprehended. 2) The data for representative dialectal variation needs to be refined. This project used recorded material as phone reference in order to compare different sound qualities of dialects. Additional recordings of real words are needed to represent dialect-specific phonological features such as the vowel length variation of Seoul dialect and the accentual variation of Kyeongsang dialect. 3) The prosodic representation was not modeled due to its complexity. This shortcoming has now been remedied by a proposed parallel research project on prosodic representation, ToBI (Cho, 1998; Jun, 2000 among others). 4) Recording of L1-transferred speech of the Korean language by English native speakers is recommended for the purpose of Korean language education. 5) Hand annotation of all audio registries is recommended if time allows. It is not common to find a complete set of speech data hand-aligned for acoustic-phonetic events on the phone level. TIMIT, a rare speech database of this kind, has served as an important speech resource for the English language (On-line speech corpora, 2001). 6) The stress variation of English lax vowels (ih, eh, ae, uh) needs to be represented in the audio registry. These are not included in this project because the English speech database is more available and less necessary at this stage of research than a Korean one. Including all these aspects would increase the data size and develop the dictionary more extensively. A great deal of this information would be useful for many users with more complicate tasks than phone reference.

References

Cambridge Dictionary of American English on CD-ROM. 2000. UK: Cambridge University Press.

Chin, Yong-ok. 2000. "Orthophonic alphabet system and task of informational technology." Paper presented in the 12th International Conference on Korean Linguistics, July 13-15, 2000. Praha: International Circle of Korean Linguistics.

Cho, Yong-Hyung. 1998. "A prosodic labeling system of intonation patterns and prosodic structures in Korean." *Korean Journal of Speech Sciences.*

Chung, Hyunsong. 1999. "SAMPA description of Hanmal Romanization symbols." *Korean Language Diphone Database Hanmal (HN1).* [www.phon.ucl.ac.uk/home/hchung/home.htm (July 29, 1999)].

Chung, Kook. 1994. "Basic concepts and transcription for speech recognition and synthesis." *Proceedings of the Speech Communication and Signal Processing Workshop* 11, 37-41. The Acoustical Society of Korea.

CMUdict.0.6d.1. 2000. A pronunciation dictionary of English lexicon in keyboard compatible phonetic alphabet. Language Technologies Institute, Carnegie Mellon University.

International Phonetic Association. 1993. The International Phonetic Alphabet. [www2.arts.gla.ac.uk/IPA/ipachart.html (Aug. 14, 2000)].

Jun, Sun-Ah. 2000. "K-ToBI (Korean ToBI) Labeling Conventions." *Speech Sciences*, 7, 143-170.

Kenyon, J. S. & T. A. Knott. 1953. *A Pronouncing Dictionary of American English* Springfield, MA: Merriam-Webster Inc.

Kim, Jong-mi. 1991. "Epenthetic-s in Korean." *Linguistic Research: Proceedings of the Kyung Hee International Conference on Language and Linguistics*, 10, 119-144. Kyung Hee Language Institute, Kyong Hee University.

Kim, Jong-mi. 1992. "Lexical phonology in Korean epenthetic-s phenomenon." *Pan Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics*, 3, 1011-1025. Thailand: Chulalongkorn University.

Kim, Jong-mi. 2000a. "Design methodology for bilingual pronunciation dictionary." *Proceedings of the Second International Conference on Language Resources and Evaluation*, 291-296. Paris: European Language Resources Association.

Kim, Jong-mi. 2000b. "Computer readable phonetic alphabet of Korean: KORBET." Paper presented in the 12th International Conference on Korean Linguistics, July 13-15, 2000. Prague. *An abstract in the preliminary proceedings.* (Full proceedings in publication). International Circle of Korean Linguistics.

Kim, Jong-mi, Stephen A. Dyer, and Dwight D. Day. 1998. "Construction of a Speech Translation Database." *Proceedings of the First International Conference on Language Resources and Evaluation*, 1071-1078. Paris: European Language Resources Association.

Kim, Jong-mi, Stephen A. Dyer, and Dwight D. Day. 1999. "Speech coarticulation database of Korean and English." *Journal of the Acoustical Society of Korea*, 18.3, 17-26.

Koo, Hee San. 1986. *An Experimental Acoustic Study of the Phonetics of Intonation in Standard Korean.* Doctoral dissertation, University of Texas at Austin. (Seoul: Hanshin).

Korean Broadcasting System. 1993. *A Korean Pronunciation Dictionary.* Seoul: Eomungak.

Labov, William. 1972. *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press.

Lee, Sang-Oak. 1995. *Basic Korean Dictionary (for foreigners): Korean-English/English-Korean.* Seoul: Hallym Publishing Co.

On-line speech corpora, web site, Linguistic Data Consortium, USA, [www.ldc.upenn.edu/ldc/online/speech/index.html (Jan 31, 2001)].

Speech and related resources, web site, European Language Resources Association, [www.icp.grenet.fr/ELRA/cata/tabspeech.html (Jan 31, 2001)].

TIMIT: Acoustic-Phonetic Continuous Speech Corpus CD-ROM. 1990. DARPA database distributed by LDC, UPENN. [www.ldc.upenn.edu (Jan 31, 2001)].

Wells, J. C. 2000. *Longman Pronunciation Dictionary.* Second Edition. Harlow: Pearson Education Limited.

Received: Jan. 30, 2001.
Accepted: Mar. 1, 2001.

▲ Jong-mi Kim
Department of English Language and Literature
Kangwon National University
Chuncheon City, Kangwon Province 200-701
Fax: +82-33-250-8259
E-mail: kimjm@kangwon.ac.kr