Annotation of a Non-native English Speech Database by Korean Speakers

Jong-mi Kim*

ABSTRACT

An annotation model of a non-native speech database has been devised, wherein English is the target language and Korean is the native language. The proposed annotation model features overt transcription of predictable linguistic information in native speech by the dictionary entry and several predefined types of error specification found in native language transfer. The proposed model is, in that sense, different from other previously explored annotation models in the literature, most of which are based on native speech. The validity of the newly proposed model is revealed in its consistent annotation of 1) salient linguistic features of English, 2) contrastive linguistic features of English and Korean, 3) actual errors reported in the literature, and 4) the newly collected data in this study. The annotation method in this model adopts the widely accepted conventions, Speech Assessment Methods Phonetic Alphabet (SAMPA) and the TOnes and Break Indices (ToBI). In the proposed annotation model, SAMPA is exclusively employed for segmental transcription and ToBI for prosodic transcription. The annotation of non-native speech is used to assess speaking ability for English as Foreign Language (EFL) learners.

Keywords: Non-native Speech, Speech Annotation, Korean-interfered English, SAMPA, ToBI, EFL, CALL

1. Introduction**

With the development of multimedia computers, speech scientists have increasingly relied on digitized speech data as raw material for their research. A well-transcribed speech annotation can characterize important linguistic features of a speech database that can be used as reliable resources for the scientific study of language as well as for language-related technology.[1)] Speech annotations have been developed to accommodate

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

** This paper has been greatly benefitted by the comments from Sei-Poong Oh, Il-Jae Lee, and Kee-ho Kim, although the remaining errors are the author's responsibility.

1) "Annotation" is a cover term for transcription, description, labeling, coding, mark-up, and analysis of speech. It covers the description of orthographic, phonetic, prosodic, morphological, syntactic, semantic, and pragmatic representations as well. That is, the

the growing need. Numerous publications are listed on the linguistic annotation website [www.ldc.upenn.edu/annotation/]; and a series of international workshops has taken place on a large scale. The development has occurred mostly in the native speech databases of either monolingual speech, as in TIMIT (Garofolo et al., 1986), or multilingual speech, as in CALLHOME and CALLFRIEND [www.ldc.upenn.edu]. The annotation system described here, however, is focused on the less explored case of non-native speech, particularly between the homogeneous target and native languages: American English and Standard Korean.

An annotated non-native speech database can be used to assess speaking ability for foreign language learners. Computer Assisted Language Learning (CALL), in particular, requires access to such database to understand, score, and correct the native-language interfered speech of the learners. The database can also provide curriculum objectives for English teachers. Other business applications include speech-to-speech translation (JANUS and Verbmobil) and multilingual recognition.[2]

To serve these purposes, a non-native speech annotation is expected to represent i) the major features of the spoken target language and ii) the typical errors of the learners.[3] The typical errors can often be predicted from the linguistically marked forms in the target language, or from the linguistic distance between the native and the target languages. Thus, the design for annotating non-native speech must represent linguistically important features of the target language and categorize the learners' predictable errors based on comparative linguistics. This is especially apparent when both the target and native languages are homogeneous as in this study.

To be used for a wide range of corpus exchange involving more than one language, the annotation format must be compatible with widely accepted conventions. Linguistic annotation methods in wide use include the Speech Assessment Methods Phonetic Alphabet (SAMPA) and the Tones and Break Indices (ToBI). SAMPA is a computer-

---

term refers to any symbolic description of particular portions of a pre-existing linguistic subject (Bird and Liberman, 2001).

2) The JANUS speech translation system translates spoken language, much like a human interpreter. It operates on a number of limited domains such as appointment scheduling, hotel reservation, travel planning, etc., [http://www.is.cs.cmu.edu/mie/janus.html]. Verbmobil is a mobile translation system for the translation of spontaneous speech in face-to-face situations [http://verbmobil.dfki.de/].

3) Technically speaking, errors are distinct from mistakes. A mistake is defined as an unsystematic failure of performance that is either a random guess or a slip-of-the-tongue in which a learner fails to use a known system appropriately. An error refers to a systematic failure of competence that deviates from the norms of the target language. Mistakes are believed to be unimportant in the language learning process; therefore, the error analysis should be limited to the study of errors (Chin, 2001; among others).

readable phonetic alphabet. More than 20 languages and dialects have been transcribed in SAMPA, as listed on its website [www.phon.ucl.ac.uk/home/sampa/home.htm]. This website provides the basic segmental SAMPA and two extensions: SAMPROSA for prosodic notations and X-SAMPA for more-detailed segmental symbols. All together, SAMPA is a keyboard-compatible system of the International Phonetic Alphabet (IPA) (Wells, 1995). ToBI is a framework for developing community-wide conventions for transcribing the intonation and prosodic structures of spoken utterances. Nine languages and dialects have been transcribed in ToBI, as listed on its website [http://ling.ohio-state.edu/~tobi/]. Both SAMPA and ToBI have inventories for American English and Standard Korean.[4] By adopting the two annotation conventions in wide circulation, annotated resources can be easily shared among users and integrated with statistical packages for linguistic analyses and technical applications.

The proposed annotation system in this paper integrates both SAMPA and ToBI into a multi-tiered structure. Each tier in the structure is well formed on its own, and multiple tiers can be combined into a single annotation using a union operation. This reflects the nature of speech that is inherently a multi-tiered activity composed of both segmental and prosodic aspects. A typical linguistic annotation system represents the structural information linked to acoustic signals, as in Bird and Liberman (2001), Cassidy and Harrington (2001), and Taylor et al. (2001).

Non-native speech annotation is more difficult than its native counterpart, in that it must deal with a greater occurrence of different sounds. This is because non-native speech involves two different languages and the varying degree of fluency. Witt and Young (2000) annotate non-native fluency by comparing the speech of the learners with the target forms stored in the computer. The similarity of the input to the target form is scaled, rating from the most native-like form to the most target-like form. The scope of non-native speech annotation is not only limited to allophonic quality, as studied by Witt and Young (2000), but also includes prosodic contours, morpho-syntactic structures, and others. Nevertheless, these areas other than allophonic differences have remained hardly explored.

The purpose of this paper is to devise explicit formats for transcription and internal data structure, and to identify the representation units and symbols of linguistic features in the annotation structure. The features refer to the segmental, prosodic, and morpho-syntactic errors that arise from linguistic differences in phonotactics, syllable structure, rhythm, inflection, and word order.

The remainder of this paper is organized as follows. Section 2 presents an overview of the non-native speech annotation model. The subsequent sections demonstrate how each component tier in the proposed model can successfully represent the errors

4) See Kim (2001c) for the Korean SAMPA.

occurring from crosslinguistic differences: morpho-syntactic errors in the word *transcription tier* (Section 3); phone quality errors in the IPA-based transcription tier (Section 4); and prosodic errors in the ToBI based transcription tier (Section 5). All these tiers are combined into a single annotation phase in Section 6.

2. Annotation Model of Non-native Speech

The proposed annotation model for non-native speech contains two features that are not needed in native speech: 1) presence of an error tier and 2) overt transcription of predictable information in both the target and native languages.5) An error tier is needed to measure learners' proficiency level of the target language, while the other tiers categorize distinct spectral patterns and prosodic units in both native and target languages. The overt transcription of non-native speech annotation contrasts to that of native speech annotation which omits predictable information. For instance, ToBI for native speech does not transcribe lexical stress that is predictable from the dictionary entry. An overt transcription of predictable information is needed, because non-native speech includes ill-formed instances of either the target or the native language, or of both. Thus, non-native speech annotation is more complex and difficult to transcribe than native annotation, due to the two extra features.

Non-native speech annotation consists of a recording of the utterance, an associated electronic or paper record of the waveform, spectrogram, fundamental frequency contours, and symbolic labels for events arranged in six parallel tiers: 1) a word tier, 2) a phone tier, 3) a tone tier, 4) a break-index tier, 5) an error tier, and 6) a miscellaneous tier. Added in non-native speech annotation is only the error tier, while the rest of components are commonly employed in native speech annotation. Non-native speech errors in the transcription tiers from 1) to 4) are time-aligned and represented in the error tier. An error type is classified by the linguistic units of Word (W), Phone (P), Tone (T), or Break (B) and also by the event types of Insertion (I), Deletion (D), or Transfer (T). The convention is to write the error type using two capital letters and the referent (i.e., the forms to be inserted, deleted, or transferred) inside the following parentheses.

Table 1 illustrates the annotation symbols and the transcribed errors. The target language is English; in particular, the general, less marked American English in the

5) This assumes that the learners' native language to be homogeneous (e.g., Korean) as in this study. For heterogeneous learners (e.g., Korean and Japanese), the predictable information is overtly transcribed for only the target language, and not the native language.

United States.[6] The target expressions are some variation of the sentence, 'Don't ask me to carry an oily rag like that,' taken from TIMIT (Garofolo et al., 1986).

Table 1. The error annotation symbols representing the error types with respect to different transcription tiers.

Transcription Tier	*Error Annotation*	*Sample Transcription*
Word	WI(that)	Don't ask me ^that to carry an oily rag like that.
"	WD(an)	Don't ask me to carry ^oily rag like that.
"	WT(rag)	Don't ask me to carry an oily ^rags like that.
Phone	PI(U)	Don't ask me to carry an oily [rEg^U] like that.
"	PD(t)	[don^] ask me to carry an oily rag like that.
"	PT({)	Don't [^esk] me to carry an oily rag like that.
Tone	TI(H)	Don't ask me to carry an oily rag like that. *H ^H *H L-L%
"	TD(*H)	Don't ask me to carry an oily rag like that. ^ L-L%
"	TT(!H)	Don't ask me to carry an oily rag like that. *H ^*H *H L-L%
Break	BI(2)	I don't want to carry an oily rag like that. ^2
"	BD(3)	Don't ask me to carry an oily rag like that, you dummy. ^0
"	BT(1)	Don't ask me to carry an oily rag like that. ^3

NOTE: W denotes 'word,' P 'phone,' T 'tone,' I 'insertion,' D 'deletion,' and T 'transfer.' The circumflex '^' marks the error location. [{] is a SAMPA phonetic symbol of the IPA [æ].

In Table 1, the error tier is shown in the middle column. In the first example, the error annotation 'WI(that)' means that the word 'that' in the transcription tier is incorrectly inserted. Shown in the right column is the transcription tier. In the second example, the error annotation 'WD(an)' is read; the word 'an' in the transcription tier is incorrectly deleted. In the third example, the error annotation 'WT(rag)' is read; the word 'rags' in the transcription tier is an incorrect form of 'rag.' In the fourth example, the error annotation 'PI(U)' in the transcription tier is read; the phone [U] is incorrectly inserted, and so on. The errors of Break Insertion (BI) or Break Transfer (BT) type are always associated with disfluency. Disfluency in a native transcription would be marked

6) The general American English refers to the group of accents in the United States that does not bear the marked regional characteristics of either the East (more precisely, Eastern New England and New York City) or the South (mainly ranging from Virginia, the Carolinas and Georgia to Louisiana and Texas). It has the largest geographical spread and is the accent most commonly used in the television networks covering the whole of the United States (Giegerich, 1992: p.47).

with the letter 'p' as in '2p' or '3p' for prolongation or hesitation. In non-native speech, however, disfluency is not limited to prolongation and hesitation. All types of disfluency are annotated in the error tier, regardless of the presence of 'p.' Break Deletion (BD) is annotated when the break level is incorrectly assigned a '0' value to indicate the absence of a word-level break in ToBI. The second example from the bottom demonstrates an instance of Break Deletion. The speaker read the word sequence 'that' and 'you' with the break level 0; in this case, the two consonants [t] and [j] are realized as one palatalized consonant [tS].

An example of error annotation is demonstrated below. The speech in the recording is that of a 21-year-old Korean female student majoring in English.

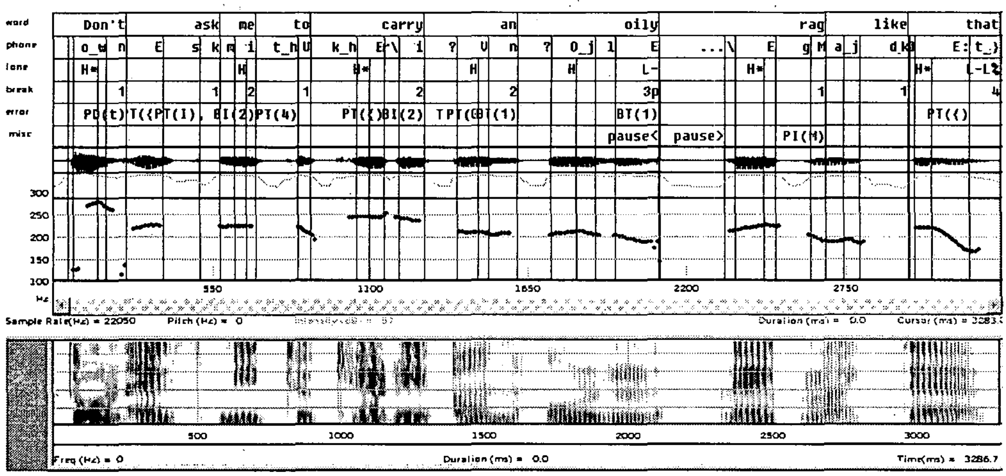


Figure 1. The non-native annotation model of an English sentence spoken by a Korean EFL learner: 'Don't ask me to carry an oily rag like that.'

In Figure 1, there are six annotation tiers for word, phone, tone, break, error, and miscellany. The transcription is based on the waveform, spectrogram, and F0 contour. Errors in the transcription tiers are explicitly annotated in the error tier. All transcriptions are time-aligned with the acoustic signals by vertical bars.

The annotation in the miscellaneous tier contains paralinguistic information such as speech turns, hesitations, vocal non-speech events, external noises, overlapping speech, or even speaker information. The transcriptions of such miscellany need not differ from those of native speech. Detailed information will be discussed, tier by tier, in the following sections.

3. Word Transcription and Morpho-Syntactic Error Annotation

The word tier in Table 2 shows a transcription of the utterance. This tier necessarily contains morpho-syntactic errors from word transcription. The data is taken from real-time visual communication in the CU-See-Me project between college students of Kangwon National University in Korea and Waseda University in Japan. The typescript is provided by the interlocutors themselves.

Table 2. Morpho-syntactic errors transcribed in the word tier and annotated in the error tier.

Word Tier	Where do you buy foods?	I buy convenient store.	Are there many
Error Tier		WD(them at a) WT(convenience)	
Misc. Tier	J1 speaker	K1 speaker	J1 speaker
Word Tier	'convenience store' in Korea?	Of course!!!	How many?
Error Tier	WT(stores)		
Misc. Tier		K1 speaker	J2 speaker
Word Tier	Lot.	We can't count it.	
Error Tier	WD(a)	WT(them)	
Misc. Tier	K1 speaker		
Word Tier	Near Waseda university, I found one Korean food restaurant a few years ago.		
Error Tier	WI(food)		
Misc. Tier	J1 speaker		

NOTE: WD denotes 'word deletion,' WI 'word insertion,' WT 'word transfer,' and Misc. 'miscellaneous.'

Table 2 demonstrates six morpho-syntactic errors, which include all three types: WD, WT, and WI. The first erroneous word 'convenient' shows two types of errors: deletion of 'them at a' and misspelling of 'convenient' for 'convenience.'

Common types of errors reported in the literature can also be annotated by the proposed convention in this paper. Chin (2001) reports, in her study of spoken errors, that transfer errors occur in the following order of frequency: article omission, plural misuse, subject-verb agreement, number agreement, misuse of copula, misuse of preposition, etc. Errors of subject or object omission or incorrect voice are less common. Among the reported errors, some are also listed in Table 2: article omission WD(a), plural misuse WT(stores), number agreement WT(them), misuse of preposition WD(at), and object omission WD(them). The other remaining types of errors are subject-verb agreement, misuse of copula, and incorrect voice. Table 3 illustrates the annotation of these error types.

Table 3. Morpho-syntactic errors in Chin (2001) annotated by the proposed system.

Word Tier	He look Asian.	She short.	This car <don't see> in Korea.
Error Tier	WT(looks)	WD(is)	WT(can't be seen)

The first example in Table 3 contains an error of subject-verb agreement and is annotated by Word Transfer (WT) with the target word 'looks.' The second example contains an error with the copula use, and is annotated by Word Deletion (WD) with the target word 'is.' The third example contains an error of voice that involves more than one word. The error is marked by a pair of angled brackets '< >.' This notation in ToBI indicates the beginning and ending of an event, as often used in the miscellaneous tier of native speech annotation. This is then annotated as an instance of Word Transfer (WT); where the erroneous form is 'don't see,' and the target form is 'can't be seen.'

The frequent errors discussed so far are not necessarily significant enough to hamper native speakers' understanding. Chin (2001) reports that native speakers of English judge such frequent errors either as insignificant (e.g., articles) or less significant (e.g., plural forms). In contrast, they considered incorrect voice and omission of a subject or a verb to be significant. These are errors related to overall sentence structure.

Salient morpho-syntactic differences of English and Korean are word order and morphological agreement. Errors related to word order are expected due to the parametric difference between English and Korean (Shin, 2001; among others). In English, words are ordered Subject-Verb-Object (SVO), and in Korean, Subject-Object-Verb (SOV). In English, a head is followed by a complement, whereas in Korean, vice-versa. The verb morphology in English must agree with the subject in person and number, while it need not in Korean. Instances of morphological agreement are illustrated by the WT(stores) in Table 2 and WT(looks) in Table 3. A word order transfer error is illustrated by the WT(can't be seen) in Table 3.

The word tier may represent semantic and pragmatic features of non-native speech as well. Non-native speech contains simplified vocabulary as well as pragmatic inappropriateness (Ko, 1998). These errors are not comprehensively studied in this paper, because they are difficult to define. A full systematic annotation is not done in this study.

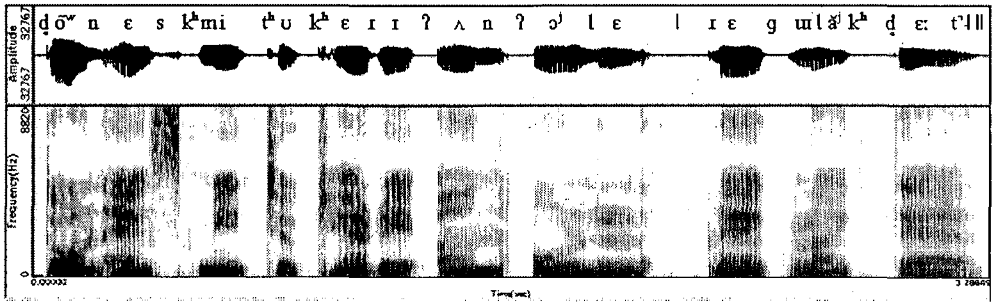
4. IPA-Based Transcription and Segmental Error Annotation

A phone tier is an IPA-based transcription of allophones, using a spectrographic analysis of input speech.[7] This tier necessarily contains pronunciation errors of

7) The acoustic details in Figure 2 have been analyzed in Kim (2001c). The analysis

individual phonemes. Figure 2 illustrates an IPA-based transcription of non-native English in the phone tier, and the time-aligned annotation for segmental errors in the error tier. The convention used is SAMPA, an IPA-based phonetic transcription for an ordinary keyboard input [www.phon.ucl.ac.uk/home/sampa/home.htm]. The speech and its annotation in Figure 2 are taken from Figure 1.

Word Tier	Don't ask me to carry an oily rag like that.
Phone Tier	d_0o_wnEskmit_hUk_hEr\I?Vn?O_jlE...r\EgMla_jkd_0E:t_}
Error Tier	PD(t) PT(I) PT({)PT(@) PI(M) PT({) PT({) PT(4)



NOTE: PD denotes 'phone deletion,' PI 'phone insertion,' and PT 'phone transfer.' SAMPA symbols are translated into the IPA: [d_0] for [d̥], [r\] for [ɹ], [V] for [ʌ], [M] for [ɯ], [a_j] for [aʲ], [t_}] for [t˺], [@] for [ə], [{] for [æ], and [4] for [ɾ].

Figure 2. Segmental errors transcribed in the phone tier and annotated in the error tier.

In Figure 2, the SAMPA transcription in the phone tier is translated into the IPA transcription [d̥oʷnɛskmitʰʊkʰɛɹɪʔʌnʔɔʲlɛɹɛgɯlaʲkd̥ɛt˺], which is a considerably narrow transcription of acoustic events. The degree of narrowness most useful for foreign language teaching has been argued to be the one that includes the allophones of both the native and target languages (Kim, 2001c). The SAMPA symbols in the error tier [@, M, {, 4, d_0] correspond to the IPA symbols [ə, ɯ, æ, ɾ, d̥]. Figure 2 demonstrates eight phone errors including all three types; PD, PT, and PI, as outlined in Section 2. The phone [t] is deleted in 'don't,' and thus annotated as PD(t) in the error tier. The phone [i] in 'me' is a transfer error of the correct allophone [I], and thus annotated as PT(I). The high back unrounded vowel phone [M] is inserted in 'rag,' and thus annotated as PI(M).

The error types reported in the literature can also be annotated by the proposed convention. The reported error types are mostly related to phonetic quality (Ahn, 2001;

---

includes phonetic correlation with the acoustic signals and sound forms, various types of transcription symbols, various degree of narrowness, and different usages of each transcription.

Hwang, 2001; Koo, 2001). Errors related to syllable structure are elaborated in Cho (1998) and Rhee and Choi (2001). Errors in phonemic distinction have been studied in Park and Ingram (1995). Cho et al. (2001) report glide-deletion; and Park (2001), positional variation of phone lengths. Among these reported errors, some have already been demonstrated in Figure 2. Phonetic quality difference is annotated by PT(4); a syllabic error by PI(M); and phonemic distinction by PT(I). The phonetic symbol [4] is a voiced alveolar tap as in 'better.' The types of errors yet to be discussed are glide-deletion and positional variation of phone lengths. Glide-deletion is marked as PD(w) or PD(j). The examples given in Cho et al. (2001) are 'queen, quiz, quest, swallow, mutual.' The positional variation of phone length in principle should be annotated as Phone Transfer. Thus, Park's (2001) examples of the sentence final words 'pack, thought, father' are transcribed in the phone tier as [p{k], [TOt], and [fAD@r], respectively. They are annotated in the error tier as PT({:), PT(O:), and PT(A:). Although possible, such detailed annotation of errors is not recommended for non-native speech. A transcription with such a narrow degree makes labeling difficult and inconsistent for transcribers.

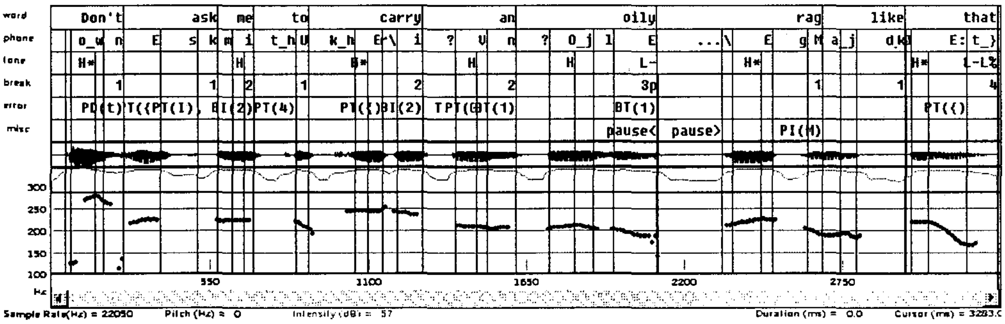
The proposed annotation system can successfully account for major segmental errors caused by language transfer. Some errors arise from the difference in syllable structure. English allows up to three consonants in syllable onset and four consonants in syllable coda, while Korean allows only one in each position. In Figure 2, the phone [t] is deleted in 'don't' to meet this syllabic requirement of Korean. Other types of phonotactic differences between English and Korean may be of phonological rules. The English words 'all night' are often pronounced as [OllaIt] by Koreans, due to liquid assimilation in Korean phonology. This is annotated as [OllaIt] in the phone tier and 'PT(n)' in the error tier. Other major errors arise due to phonemic and allophonic differences. The phonemic errors are PT(I), PT(@), and PT({); and the allophonic errors are PT(4) and PT(g_0). The vowels [E, i, V] in 'ask, me, an' produced by the Korean subject are phonemically different from the correct vowels [{, I, @] that a native speaker of the target language would utter. These errors are annotated as Phone Transfer in the error tier. Less importantly, the consonant [t_h] in 'to' is a heavily aspirated phoneme in Korean, but merely an English allophone of the phoneme /t/. The correct pronunciation should be with the alveolar tap [4] that may be the lenis stop phoneme /d_0/ in Korean, but not the heavily aspirated stop phoneme /t_h/. This phone error represents Phone Transfer, PT(4). The time alignment protocol of acoustic events may follow the outlines of Zue and Seneff (1988) and Kim (2001a: p.48).

The SAMPA transcription is useful for representing the speech phenomena using a standard keyboard, in as much detail as the IPA. The transcription converts to the IPA through one-to-one mapping of the symbols, as shown in Figure 2. The graphic symbols of the IPA are not convenient for annotating a speech database. The Hangeul transcription is also inappropriate for non-native speech annotation, which contains non-Korean speech forms. The official Korean Romanization is also inappropriate for non-native speech transcription. It is capable of transcribing monolingual Korean speech only (Kim, 2001a). There have been attempts to extend the Hangeul orthographic conventions that can encompass unfamiliar sounds of other languages (Chin, 2000; Seo and Cheong, 2000). However, the graphic forms in the Hangeul extensions are not compatible with an ordinary keyboard.

5. ToBI Transcription and Prosodic Error Annotation

The TOne and Break Index tiers represent prosodic analyses according to the ToBI system developed in Silverman et al. (1992), Beckman and Elam (1997), Jun (2000), and others. These tiers contain prosodic errors related to tone and break assignment. Figure 3 is a ToBI transcription of non-native speech with error annotation.

Word	Don't	ask	me	to	carry	an	oily	rag	like	that.
Tone	H*	H*	H		H*	H	H L-	H*		H* L-L%
Break	1	1	2	1	2	2	3p	1	1	4
Error			TI(H) BI(2)		BI(2)	TI(H) BT(1)	BT(1)			



NOTE: TD denotes 'tone deletion,' TI 'tone insertion,' TT 'tone transfer,' BD 'break deletion,' BI 'break insertion,' and BT 'break transfer.' ToBI symbols are: H for 'high,' L for 'low,' * for 'pitch accent,' for 'phrasal accent,' and % for 'sentence boundary.'

Figure 3. Prosodic errors transcribed in the tone and break tiers and annotated in the error tier.

The tone transcription in Figure 3 is narrower than the ToBI transcription of English as native speech (E-ToBI, henceforth), as outlined in Beckman and Elam (1997). Word stress 'H' is overtly transcribed in non-native speech, but not in E-ToBI for native speech. "Word stress" refers to the prominence of word internal rhythms, marking of stressed and unstressed syllables within each word, or more precisely within a foot.[8] The foot-internal prominence depends on pitch increase, but in other cases, on loudness, length, vowel quality, or any combinations of those. In Korean, accent is assigned to the raised pitch (Koo, 1986; Jun, 2000; among others). Assignment of a symbol 'H' or 'L' marks foot-internal prominence to maintain consistency with the ToBI system. These new symbols as well as other standard ToBI symbols cannot be transcribed to show all aspects of prosody or other aspects that are amenable to symbolic transcriptions (Beckman and Ayers, 1997: Section 1.2). Moreover, since word stress is largely predictable from the dictionary entry, it is not transcribed in E-ToBI.

E-ToBI only concerns pitch accent to describe word prominence with regard to other words in the same intonational phrase. "Pitch accent" refers to prosodic prominence that can be either a nuclear or a prenuclear stress. A "nuclear pitch accent" falls on the most prominent syllable within intonational phrase. This is positionally determined, in that it falls on the stressed syllable of an emphasized word, or of the last accented word within the phrase. A "prenuclear pitch accent" is assigned to a more stressed syllable whose prominence level ranges between nuclear pitch accent and word stress. The place that a prenuclear pitch accent is assigned cannot be predicted from the grammar (Beckman and Elam, 1997: Section 1.3). Neither does E-ToBI distinguish nuclear pitch accent from prenuclear pitch accent with the use of symbols, because the former can be derived positionally. For instance, the sentence 'Marianna made the marmalade' assigns five Hs for word stress to the following vowels in the capital letters, 'MAriAna MADE the MARmaLAde'; one nuclear pitch accent H* to 'mar' of 'marmalade'; and one prenuclear pitch accent H* to 'A' in 'mariAna.' H* is assigned to the top of the place where an H is currently licensed with word stress. E-ToBI omits word stress transcription, but transcribes the nuclear or prenuclear pitch accent with H* in 'marAna made the MARmalade.' In contrast, the proposed system for non-native annotation transcribes word stress, because a non-native speaker is likely to make errors that native speakers would not do in the same environment. The rest of the annotation conventions follow E-ToBI.

In Figure 3, the first word 'Don't' starts with the pitch accent H*, showing the highest F0 level around 300 Hz. The second word 'ask' is not accented, since it is in a lower F0 level than 'Don't.' The absence of declination of an unaccented sequence shows the assignment of a word accent H to 'me.'[9] The raised F0 in 'carry' contains a pitch

8) Function words such as prepositions or articles do not get any word stress within a foot, unless they are in an emphatic context. A foot consists of a word stress and unstressed syllable(s).

9) The following word 'to' also contains a word stress 'H' in the F0 trace according to

accent H*. The next word 'an' contains a word accent H, due to the surrounding glottal stops and the unrounded tense mid-vowel: [V] in SAMPA and [ʌ] in the IPA. Words pronounced in isolation receive a word stress in English. The next word 'oily' also contains a word stress H on the first syllable [O_j], as seen by the absence of declination of an unaccented sequence. The second syllable in 'oily' has a falling F0 contour that is followed by a pause; thus, the phrasal final accent L- is assigned.

The raised F0 contour shows that 'rag' receives a high pitch accent H*. The next word 'like' is not accented, as shown by a lower F0 level than 'rag.' What follows is the raised F0 trace, indicating that the word 'that' contains a high pitch accent H*. This pitch accent is not down-stepped, since its F0 range is the same as the preceding H* for 'rag.' The utterance ends with the typical declarative intonation L-L%.

When compared to tonal patterns of native speech, it is expected to de-accent 'ask.' In English, two adjacent words are not accented at the same time: This is due to the nature of English rhythm that a stressed syllable alternates with an unstressed syllable. Another instance of de-accenting in native speech concerns 'me' and 'an,' where word stress is absent from native speech, unless emphasized. These words are often incorrectly stressed by non-native speakers, as shown in Fig 1-3. Such annotation is an instance of Tone Insertion, TI(H).

The break transcription of non-native English follows the conventions used in E-ToBI without modification. The break level 1 is assigned to 'Don't ask me,' as expected in native speech. A break level 2 is assigned between 'me' and 'to', because 'to' starts with heavy aspiration. The break level 2 in E-ToBI is a disjoint word boundary. A native speaker does not usually allow a break in this level, and instead, pronounces as 'to' with an alveolar tap: [4] in SAMPA and [ɾ] in the IPA. The break level 1, as expected in native speech, splits the word sequence 'to carry'. The next sequence 'carry an oily' has the break level 2, because 'an' is isolated between two glottal stops. The word 'oily' is followed by the break level 3p, due to a hesitation pause with the phrasal accent L-. The symbol 'p' in E-ToBI refers to hesitation. The break level 1, as expected in native speech for word boundary, again separates the words in 'rag like that.' The break level 4 is assigned to the sentence final break, as it would be in native speech.

When compared to native speech, the break level 1 for word boundary is expected and not considered as an error. However, the break level 2 between 'me' and 'to' is an error, because a native speaker would produce the break level 0 with an alveolar tap, a consonant observed in word internal position. This is then annotated as Break Insertion, BI(2). For the same reason, the break level 2 between 'carry' and 'an' is annotated as

---

other software like CSL by Kay Elemetrics. The F0 height of it is the same as 'me.' Transcribing the function word 'to' with a word stress 'H' another case of tone insertion error, TI(H).

BI(2). The word 'an' in native speech is usually not accented and linked to the preceding word without an intervening break. The break level 2 between 'an' and 'oily' is Break Transfer. An expected break level in native speech would be the break level 1 for normal word boundary. For the same reason, the break level 3p between 'oily' and 'rag' is annotated as Break Transfer, BT(1).

Figure 4 compares those features of non-native speech with native speech signals. The fundamental frequency contour shows a clear declination. Notice that the words 'me,' 'to,' and 'an' all show a declination. The speech is taken from TIMIT (1990). The speaker's ID is Fntb0 with the dialect of the upper Midland in the US.

Word	Don't	ask	me	to	carry	an	oily	rag	like	that.
Tone	H*				!H*		!H*			L-L%
Break	1	1	0	1	0	1	1	1	1	4

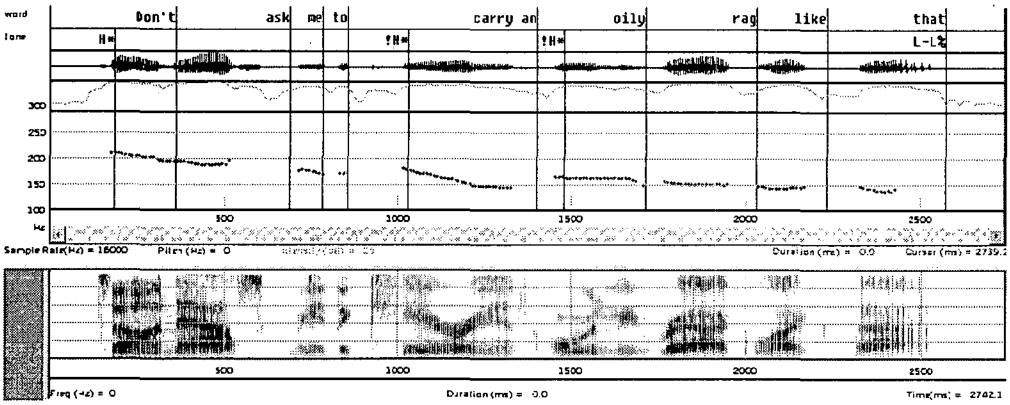


Figure 4. Figure 4. A native speech of American English transcribed in E-ToBI.

In native speech, a clear declination pervades the sentence except the three peaks, 'Don't,' 'carry,' and 'oily.' These words are annotated as H*. The second and third peaks are gradually down-stepped. Unlike the non-native speech in Figure 3, the words 'me,' 'to,' and 'an' are declined; in other words, unaccented. A break is not present inside 'me to' and 'carry an,' and is transcribed as the break level 0. The /t/ in 'me to' is a voiced alveolar tap, as if 'me to' were one word. An alveolar tap occurs word internally in American English. 'Carry an' also acts like one word: The nasality in the low front vowel in 'carry' persists to the final nasal consonant [n]. The low front vowel [{] (the IPA [æ]) is inherently nasalized.

The annotation conventions for non-native prosody are summarized below. They are different from the transcription of English native speech (Beckman and Elam, 1997), in that lexical stress is overtly transcribed, as stated in (1) below. The transcription conventions in (2) and (3) are a summary of E-ToBI. The error annotations as in (4) are exclusive for non-native speech, which are not needed in E-ToBI for native speech.

(1) A stressed syllable is marked by H or L with no diacritics attached.
(2) An asterisk '*' marks pitch accent, and the most common type is H*. Some variations of it are L*, L*+H, L+H*, !H*, L*+!H, L+!H*, and H+!H*. A nuclear pitch accent falls on the stressed syllable of an emphasized word or the last content word.
(3) Boundary conventions for tonal transcription and break indices follow E-ToBI. Break indices are on a scale from 0 (the strongest perceived conjoining) to 4 (the most disjoint). Level 0 is for clitics, level 1 for word boundary, level 2 for disjoint word boundary, level 3 for phrase boundary, and level 4 for sentence-final boundary. The level 3 boundary has three accents, L-, H-, and !H-. At level 4, two boundary tones are represented, either independently as in L% and H%, or combined with the level 3 tones as in L-L%, L-H%, H-L%, H-H%, !H-L%, and !H-H%. A standard declarative F0 contour is represented by L-L%, and a standard yes-no question contour ends in H-H%. A typical boundary tone for a sentence internal phrase is L-H%.
(4) The error tier has the representation of the error types using two capital letters and the referent inside the following parentheses. The error types under discussion are Tone Insertion (TI), Tone Deletion (TD), Tone Transfer (TT), Break Insertion (BI), Break Deletion (BD), and Break Transfer (BT). The referent refers to the form to be inserted, deleted, or transferred. An example of representation using the notations can be TI(H), a Tone Insertion error for an illegally inserted H tone.

The annotation system proposed in this paper can successfully annotate the error data reported in the literature. The first error data to be considered is the wrong assignment of boundary tone in an intermediate phrase (Oh et al., 2001).

Table 4. A boundary tone assignment error in Oh et al. (2001) annotated by the proposed system.

Word	Fortunately,	he	can	win	the	game.
Native Tone	H* L-H%					H%
Non-native Tone	H* H-L%					L-L%
Error	TT(L-H%)					

The erroneous boundary tone in non-native Korean speech is annotated as Tone Transfer, TT(L-H%). The tonal transcription is provided in Oh et al. (2001). The utterance is not considered for a more detailed transcription, to avoid misleading the point of discussion.[10)]

Jun et al. (2001) report that accents are realized differently for general negation than the negation denying presupposition in American English. In contrast, they are realized the same for English spoken by a Korean. Table 5 illustrates such difference.

Table 5. Pitch accent assignment errors in Jun et al. (2001) annotated by the proposed system.

Word	Was Patty in your car? No, Patty was never in my car.
Native Tone	LH* L-L%
Non-native Tone	H* !H* !H* L-L%
Error	TT(H) TT(!H)
Word	What was Patty doing in your car? Patty was never in my car.
Native Tone	H* L-L%
Non-native Tone	H* !H* !H* L-L%

Both examples in Table 5 show a number of pitch accents in English spoken by a Korean. The tonal transcription is provided in Jun et al. (2001), although the final boundary tone is supplemented in this study. The boundary tone transcription is based on the F0 trace in Jun et al. (2001), which shows a typical declination contour. The boundary tone is supplemented to prevent from misleading the point of discussion. In the proposed annotation system, the accent levels for 'Patty' and 'car' lower to word stress; that is, Tone Transfer from H* to H. Thus, they are annotated as TT(H).

On the other hand, the pitch accent for 'never' is not annotated as an error. Jun et al. (2001) report in their study that some native speakers often do not distinguish between the two cases of negation in some sentences. The error annotation is used only with obvious errors, leaving out marginal ones.

Ahn et al. (2001) report that Korean learners often fail to stress a head noun that is followed by a complement phrase. There is no peak, not even at a small scale.

Table 6. A pitch accent assignment error in Ahn et al. (2001) annotated by the proposed system.

Word	Tom	has	the	house	that	I	spoke	of.
Native Tone	H*			!H*			H	L-L%
Non-native Tone	H	H	H	H	H	H	H	L-L%
Error		TI(H)	TI(H)	TT(H*)	TI(H)	TI(H)	TI(H)	

10) If re-transcribed, a nuclear accent H* is assigned in the second intonational phrase. All intonational phrases, by definition, include at least one nuclear accent.

Table 6 shows a typical case where a nuclear pitch accent is missing: Tone Transfer from pitch accent to word stress. The ToBI transcription is not provided in Ahn et al. (2001), but supplemented in this study based on the fundamental frequency contour. In non-native speech, the F0 trace is flat throughout the utterance. That is, all syllables are assigned a word stress, as in a typical syllable-timed language like Korean. The insertion of word stress is annotated as the error TI(H). The central claim in Ahn et al. (2001) is that there is no nuclear accent in English spoken by Koreans, unlike in native speech. The absence of nuclear accent is annotated as Tone Deletion. The annotation TT(H*) does not specify the down-stepped information, because the down-step is a relational phenomenon inferred from the fact that the preceding H* is higher. The presence or height of the preceding H* is not presupposed without independent motivation.

There are other major error types that have not been discussed. For comprehensiveness, the proposed annotation system needs to account for the major prosodic errors of English. In particular, errors dealing with important prosodic features of the target language that are systematically different from the speakers' native language need to be tested.

Such errors are of three categories: 1) stress placement errors, 2) rhythm errors, and 3) intonation realization errors. Stress placement errors concern lexical stress as in the verb 'preSENT' meaning 'to demonstrate' and the noun 'PREsent' meaning 'gift.' Standard Korean does not have lexical stresses, and, for that reason, the learners are more likely to misplace them when speaking English. These errors are represented by a combination of Tone Deletion and Tone Insertion as in the following.

Table 7. Lexical stress assignment errors annotated by the proposed system.

Word	p r e s e n t (Verb)	p r e s e n t (Noun)	What a sensation!
Native Tone	H	H	H H*
Non-native Tone	H	H	H !H*
Error	TD(H) TI(H)	TI(H) TD(H)	TI(!H*) TD(H*)

In Table 7, the misplaced stress is annotated by Tone Deletion of the wrong tone, TD(H), and Tone Insertion of the correct tone, TI(H).

The second major prosodic feature concerns rhythm. In English rhythm, each stressed syllable is assigned equal time. Function words such as pronouns, auxiliary verbs, and prepositions are unstressed. Compare the stress placement of two sentences, 'DOGS EAT BONES' and 'The DOGS will have EATen the BONES.' Both sentences contain three stressed syllables each. A native speaker of English reads the two sentences in equal time (Prator and Robinett, 1985: p.40). In contrast, since Korean is a syllable-timed language, the learners take much longer time to read the second sentence that has more syllables (Han et al., 1999: pp.119-121). Table 8 illustrates the errors due to rhythmic difference.

Table 8. Stress-timed rhythm errors annotated by the proposed system.

Word	Dogs eat bones	The dogs will have eaten the bones
Native Tone	H !H H* L-L%	H !H H* L-L%
Non-native Tone	H H H L-L%	H H* H H H H !H* L-L%
Error	TT(H*)	TI(H) TI(H) TI(H) TI(H)

Recent literature does not discuss the ToBI transcription, which this study supplements based on the account in Prator and Robinett (1985) for native speech and in Han et al. (1999) for non-native speech. According to Table 8, in native speech, every stressed syllable permits a word stress H, among which the final one gets the nuclear accent H*. The function words 'the, will, have' are not assigned an accent. The second word accent H is down-stepped, due to being adjacent to another H. In contrast, every word in non-native speech has a word accent, regardless of being a function or content word. There is no time reduction for de-accenting a function word. In other words, the non-native speech rhythm is syllable-timed due to the extra assignment of word stress to every syllable. This extra assignment of word stress is annotated by Tone Insertion: TI(H). The phenomenon is the same, as in Table 6, where the F0 trace is flat throughout the utterance. Table 8 reveals another error of nuclear accent omission in the first example, 'Dogs eat bones.' This nuclear accent placement is discussed below.

The final major topic of English prosody concerns intonation. Intonation realization errors represent sentential stress patterns. A word with nuclear stress is defined positionally in the absence of emphasis. It is the last accented word in the phrase (Bauer et al., 1980: p.120). Consider the following sentence, in which the stressed syllables are written in capital letters: 'The intonation of a LANguage / can be described as its MELody / but this is only a first approxiMAtion.'

On the other hand, the highest F0 in a phrase is assigned to the last syllable in Korean, where accent is realized by pitch height (Koo, 1986). Therefore, the F0 contour will be realized differently from English when the phrase final syllable is unstressed. The given example would be pronounced with the wrong accent: 'The intonation of a lanGUAGE / can be described as its meloDY / but this is only a first approximation.' Such crosslinguistic differences are annotated in ToBI, as in Table 9.

Table 9. Errors of nuclear accent assignment annotated by the proposed system.

Word	The intonation of a language can be described as its melody but this is only a first approximation.								
Native Tone	H	H*	H- 3	H	!H* L-H% 4	H	H	H*	L-L% 4
Non-native Tone	H	H	H- 3	H	H H- 3	H	H	H	L-L% 4
Error		TT(H*)			TT(H*)			TT(H*)	

Again, the ToBI transcription is not provided in the literature; however, it is supplemented in this study. The native speech pattern is based on the explanation by Bauer et al. (1980; p.120). The Korean transcription is based on the explanation in Ahn et al. (2001) and Koo (1986). An informal phonetic experiment also supports the transcription.[11]

In Table 9, the native speech contains a nuclear accent in every phrase, while the non-native speech has none. The omission of nuclear accent is annotated by Tone Transfer from nuclear accent to word stress, TT(H*). A similar error is also with TT(H*) in Table 8. The last accented word 'bones' does not have a nuclear accent H*; but merely a word stress !H. This is annotated by Tone Transfer, TT(H*).

6. Time Alignment of Transcription Label Files

The linguistic relations among annotation symbols in different tiers can be read and categorized on a time-linked label file in plain text format. The relations include a number of labels and the overlaying annotations. In the proposed system, as demonstrated, each tier in the annotation model is represented in separate label files for the word tier, the break index tier, the tonal tier, the error tier, and the miscellaneous tier. All these independent label files contain the label symbols lined up with digitized utterance time. This time reference makes possible for merging all the multiple tiers. The illustration in Table 10 shows how the non-native speech annotation following the proposed system merges with a single annotation file in plain text format.

11) Depending on the level of fluency, the number of stressed syllables varies accordingly.

Table 10. Non-native speech annotation showing all six tiers time-linked and combined into a single annotation plane. The first column is the time alignment protocol. The following six tiers are of phone, word, tone, break, error, and miscellany.

500	i-s					
659	d_0					
1508			H*			
1816	o_w					
2499	n	Don't		1	PD(t)	
3777	E				PT({)	
5060	s					
5773	k	ask		1		
6282	m					
6687			H		TI(H)	
7023	i	me		2	PT(I), BI(2)	
8485	t_h				PT(4)	
8919	U	to		1		
10521	k_h					
10960			H*			
11486	E				PT({)	
12010	r\					
12881	I	carry		2	BI(2)	
14032	?					
14749			H		TI(H)	
15102	V				PT(@)	
16154	n	an		2	BT(1)	
17307	?					
18187			H			
18847	O_j					
19563	l					
21065	E	oily	L-	3p	BT(1)	pause<
23316	...					pause>
23689	r\					
24668			H*			
25067	E					
26275	g					
26811	M	rag		1	PI(M)	
27409	l					
27986	a_j					
29678	k	like		1		
29886	d_0					
30475			H*			
31774	E:				PT({)	
32845	t_}	that	L-L%	4		

NOTE: Time in the first column is in one-tenth of milliseconds. 'i-s' is an abbreviated form of Initial Suspension of airflow, which is used to be distinguished from the preceding silence.

All annotation types, discussed in Figures 1-4, are aligned by time in Table 10. The transcription of acoustic events follows the conventions of SAMPA in the second column, and ToBI in the fourth and the fifth columns. The non-native errors are annotated in the sixth column according to their types and the referents, 'Type(referent).' The phone tier in the second column shows the most frequent distribution in time, since phones can be sorted according to the shortest time duration when compared to other linguistic units in the same transcription. For example, the three phones [d_0, o_w, n] map into 'Don't' in the third column at 249.9 milliseconds. The tone tier in the fourth column places the label for pitch accents at the minimum or maximum of F0, if this F0 event is within the interval of a vowel segment at the phone tier. Pitch accents are placed within the interval that can be identified with the vowel. For instance, the initial tone H* is assigned to the peak of the fundamental frequency contour of the vowel [o_w] at 150.8 milliseconds; the next tone H to the mid-point of the vowel [i] at 668.7 milliseconds because there is no apparent F0 peak; and so on. The break levels in the fifth column are aligned at the end of each word. For example, the break level 1 divides the first word 'don't' and the second word 'ask'; and is aligned with the word 'don't' at 249.9 milliseconds. The error annotation in the sixth column is time-aligned with relevant transcriptions in the second to the fifth columns. For instance, the first phone error PD(t) occurs to the phone [n] at 249.9 milliseconds. The first tone error TI(H) is aligned with the associated H at 668.7 milliseconds.

Some concrete heuristic rules employed in this diagram renders consistent transcription, which are not specified in Zue and Seneff (1988), Kim (2001a: p.48), among others. For instance, vowel edges are segmented at the peak of a wave period that retains distinct characteristics of a given vowel. They can also be time-aligned on the spectrogram where the formants, F1 and F2, are distinctively discriminated. In Table10, the first consonant [d_0] is segmented at 65.9 milliseconds, because the final peak of the wave period occurs at the same time, retaining characteristic features of the following vowel [o_w]. Relevant waveforms and spectrograms are provided in Figures 2 and 3, although the accurate alignment requires further enlargement of the signals. Another instance is found at the consonant offset following a vowel as in 'that.' The phone [E:] divides the vowel and the consonant offset [t_}], the right edge of which is marked at the end of the amplitude track, i.e., 57dB at 3284.5 milliseconds. This type of alignment leaves more signals for consonants than vowels, because the coarticulation parts are included in the consonants. This is useful for speech recognition or CALL database, in that it allows more features to recognize consonants; otherwise consonants contain either no or low signal level. One may decide contrastingly, in that vowel edges are segmented at the end of voicing. When followed with this rule of segmentation, the consonantal duration can be more accurately measured. The way of setting up heuristic rules depends on the purpose of the database. For instance, adopting the segmental transcriptions in Table 10 for education purposes is less narrow than the case of acoustic phonetic databases in Zue and Seneff (1988) and Kim (2001a: p.48).[12)]

7. Conclusion

A convention for annotating errors was devised that can categorize systematic errors of non-native English spoken by Koreans. This annotation convention has been tested against the reported errors in the previous literature as well as salient crosslinguistic differences. The major crosslinguistic differences are successfully annotated by only seven symbols: Word, Phone, Tone, Break, Insertion, Deletion, and Transfer.

The proposed annotation system in this paper enhances accuracy, scalability, adaptability, and consistency. Accuracy concerns the accurate annotation of major salient linguistic features involved in non-native speech. This has been demonstrated throughout the paper by examining each major linguistic aspect involved in speech. Scalability deals with whether the annotation system is capable of transcribing very large corpora. A simple annotation system like the proposed one is considered as an advantage for scalability. Adaptability concerns using the systems of widespread popularity as in SAMPA and ToBI. This enables adopting the proposed system to different platforms for a wide range of corpus construction and annotation tasks. By sharing the widespread annotation system, developing procedures will cut down on the annotation time. Consistency concerns the annotation consistency within a transcriber and among different transcribers. A simple system like the proposed one can be more consistent for transcribing.

Annotation of non-native speech is used in Computer Assisted Language Learning (CALL), in particular for understanding, scoring, and correcting English with Korean accent. An annotated database is also useful for research in linguistics as well as technical applications. An annotated speech corpus provides an effective way of testing theories of speech and language, without each laboratory having to devote extensive resources for creating and annotating individual task-specific databases.

Some issues remain unresolved regarding the proposed annotation. ToBI prosodic labeling of speech typically takes for even an experienced labeler from 100 to 200 times real time (Syrdal et al., 2001). Annotation of non-native speech will take even more time, due to the addition of errors and phones. To alleviate this problem, further developments

12) Kim (2001), for instance, has the independent labeling of stop closure for word medial voiceless stops that specifies the phonetic quality of the following release. Such precision is not adopted in Table 10.

are needed in the future. In the case of E-ToBI, Syrdal et al. (2001) has proposed such a time-saving system.

References

Ahn, Gil-Soon, Pyung-man Jeon & Hyun-Gee Kim. 2001. "A prosodic study of focus in English relative clauses." [In Korean]. *Korean Journal of Speech Sciences,* 8(4), 207-214.

Ahn, Soo-Woong. 2001. "An acoustic study of relative articulatory positions of English vowels." *Korean Journal of Speech Sciences,* 8(4), 171-184.

Beckman, Mary E. and Gayle A. Elam. 1997. *Guidelines for ToBI Labelling.* Version 3. The Ohio State University Research Foundation.

Bauer, Laurie, John M. Dienhart, Hans H. Hartvigson, and Leif K. Jakobsen. 1980. *American English Pronunciation.* Wellington: Gyldendal.

Bird, Stephen and Mark Liberman. 2001. "A formal framework for linguistic annotation." *Speech Communication,* 33, 23-60.

CALLFRIEND. speech database. [http://www.ldc.upenn.edu/Catalog/by_type.html#speech].

CALLHOME. speech database. [http://www.ldc.upenn.edu/Catalog/by_type.html#speech].

Cassidy, Steve and Jonathan Harrington. 2001. "Multi-level annotation in the EMU speech database management system." *Speech Communication,* 33, 61-77.

Chin, Cheongsook. 2001. "Error analysis: an investigation of spoken errors of Korean EFL learners." *English Teaching,* 56(2), 97-124.

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

Cho, Hyeonkwan. 1998. "English loanwords and syllable structures." *Studies in Phonetics, Phonology and Morphology,* 4, 1-29.

Cho, Mi-Hui, Mae-Ran Park, and Shinsook Lee. 2001. "The acquisition of English onglides by EFL students in Korea." *English Teaching,* 56(2), 79-96.

Garofolo, John S., Lori F. Lamel, William M. Fisher, Jonathon G. Fiscus, David S. Pallett, and Nancy L. Dahlgren. 1986. *DARPA TIMIT: Acoustic-Phonetic Continuous Speech Corpus CDROM.* NIST. [http://www.ldc.upenn.edu/lol/docs/TIMIT.html].

Giegerich, Heinz J. 1992. *English Phonology: An Introduction.* [Cambridge Textbooks in Linguistics]. Cambridge: Cambridge University Press.

Han, Jaehyeon, Jong-mi Kim, In Lee, Dae-Bin Kim, Young-hern Lee, Jae-woong Choe, Byong-Man Jeon, Jeongkhn Ahn, and Jun-Eon Park. 1999. *Understanding English Linguistics,* 2nd ed. Seoul: Hankuk.

Hong, Soonhyun. 2001. "Output-oriented *t* and *s* in English loanwords in Korean." *Studies in Phonetics, Phonology and Morphology.* 7(1), 243-263.

Hwang, Young-soon. 2001. "A comparative analysis on English vowels of Korean student by formant frequencies." [In Korean]. *Korean Journal of Speech Sciences,* 8(4), 215-220.

The International Phonetic Alphabet. 1993. International Phonetic Association. [www2.

arts.gla.ac.uk/IPA/ipachart.html].

Janus. web site. [http://www.is.cs.cmu.edu/mie/janus.html].

Jun, Ji-Hyun, Soon-Boak Park, and Kee-Ho Kim. 2001. [In Korean]. "The accentual realization of negation of presupposition in English." *Korean Journal of Speech Sciences*, 8(4), 259-274

Jun, Sun-ah. 2000. "K-ToBI (Korean ToBI) labeling conventions." *Speech Sciences* 7(1), 143-170. Korean Association of Speech Sciences.

Kim, Jong-mi. 2001a. "Computerized sound dictionary of Korean and English." *Speech Sciences*, 8(1), 271-296.

Kim, Jong-mi. 2001b. "Phonetic extension of the official Korean Romanization." *Korean Journal of Linguistics*, 26(4), 669-694.

Kim, Jong-mi. 2001c. "Computer codes for Korean sounds: K-SAMPA." *Journal of Acoustical Society of Korea,* 20(4E), 1-14.

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.

Koo, Hee San. 2001. "An acoustic study of the pronunciation of English vowels uttered by Korean regional dialect speakers." [In Korean]. *Korean Journal of Speech Sciences,* 8(4), 193-206.

Official Romanization System. 2000. Ministry of Culture and Tourism, Republic of Korea. Available in [http://www.sejong.or.kr/b_4.htm] and in [http://www.korean.go.kr/bbs/pds.html].

Oh, Sei-Poong, Young-Soo Jang, and Yong-Jae Lee. 2001. [In Korean]. "A study on the intonation of Korean speakers in English sentence adverbials." *Korean Journal of Speech Sciences,* 8(2), 119-132.

Park, Hee-Suk. 2001. "A study on the English vowel lengths /æ/, /ɔ/ and /ɑ/." [In Korean]. *Korean Journal of Speech Sciences,* 8(4), 215-220.

Park, See-gyoon and John Ingram. 1995. "/l/ & /r/ perception by Korean and Japanese speakers learning English: the relative importance of L1 transfer effects and universal acoustic factors." *Korean Journal of Linguistics*, 20(4), 87-110.

Prator, Clifford H. and Betty W. Robinett. 1985. *Manual of American English Pronunciation.* Chicago: Holt.

Rhee, Seok-Chae and You-Kyung Choi. 2001. "A statistical observation of vowel epenthesis in English loanwords in Korean and its significance." [In Korean]. *Studies in Phonetics, Phonology and Morphology.* 7(1), 153-176. Phonology-Morphology Circle of Korea.

Seo, Chung-soo and Dal-young Cheong. 2000. "A study on the international phonetic Hangeul." [In Korean]. Paper presented in the 12th International Conference on Korean Linguistics, July 13-15, 2000. Prague. International Circle of Korean Linguistics.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. "ToBI: a standard for labeling English prosody." *Proceedings of the International Conference on Spoken Language Processing*, 2, 867-870.

Shin, Jung-Sun. 2001. "Interlanguage grammar of Korean EFL learners: Cross-sectional study on the phrase structure development." *English Teaching*, 56(2), 167-192.

Speech Assessment Methods Phonetic Alphabet (SAMPA): computer readable phonetic alphabet. website. [www.phon.ucl.ac.uk/home/sampa/home.htm].

Syrdal Am K., Julia Hirschberg, Jalia McGory, and Mary Beckman. 2001. "Automatic ToBI prediction and alignment to speed manual labeling of prosody." *Speech Communication*, 33, 135-151.

Taylor, Paul, Alan W. Black, and Richard Caley. 2001. "Heterogeneous relation graphs as a formalism for representing linguistic information." *Speech Communication*, 33, 153-174.

ToBI. website. [http://ling.ohio-state.edu/~tobi/].

Verbmobil. website. [http://verbmobil.dfki.de/].

Verbmobil: Foundations of Speech-to-Speech Translation. 2001. W. Wahlster (ed.). Berlin: Springer-Verlag.

Wells, John C. 1995. "Computer-coding the IPA: a proposed extension of SAMPA," ms. Downloadable from [www.phon.ucl.ac.uk/home/sampa/x-sampa].

Witt, Silke M. and Steve J. Young. 2000. "Phone-level pronunciation scoring and assessment for interactive language learning." *Speech Communication,* 33, 135-151.

Zue, Victor W. and Stephanie Seneff. 1988. "Transcription and alignment of the Timit database." Paper presented at the Second Symposium on Advanced Man-Machine Interface through Spoken Language, Oahu, Hawaii. Available in Garofolo, Lamel, Fisher, Fiscus, Pallett, and Dahgren (1986), *DARPHA TIMIT*, pp. 36-45.

Received: Jan. 28, 2002.
Accepted: Mar. 11, 2002.

▲ Jong-mi Kim
Department of English Language and Literature
Kangwon National University
Chuncheon City, Kangwon Province 200-701
Tel: +82-33-250-8150 Fax: +82-33-250-8259
E-mail: kimjm@kangwon.ac.kr