

# Studying L2 Suprasegmental Features in Asian Englishes: A Position Paper

Helen Meng<sup>1</sup>, Chiu-yu Tseng<sup>2</sup>, Mariko Kondo<sup>3</sup>, Alissa Harrison<sup>1</sup> and Tanya Visceglia<sup>4</sup>

<sup>1</sup>The Chinese University of Hong Kong, <sup>2</sup>Academia Sinica, Taiwan

<sup>3</sup>Waseda University, Japan and <sup>4</sup>Ming Chuan University, Taiwan

hmmeng@se.cuhk.edu.hk, cytling@sinica.edu.tw, mkondo@waseda.jp, alissa@se.cuhk.edu.hk, tvisceglia@gmail.com

## Abstract

This position paper highlights the importance of suprasegmental training in secondary language (L2) acquisition. Suprasegmental features are manifested in terms of acoustic cues and convey important information about linguistic and information structures. Hence, L2 learners must harness appropriate suprasegmental productions for effective communication. However, this learning process is influenced by well-established perceptions of sounds and articulatory motions in the primary language (L1). We propose to design and collect a corpus to support systematic analysis of L2 suprasegmental features. We lay out a set of carefully selected textual environments that illustrate how suprasegmental features convey information including part-of-speech, syntax, focus, speech acts and semantics. We intend to use these textual environments for collecting speech data in a variety of Asian Englishes from non-native English speakers. Analyses of such corpora should lead to research findings that have important implications for language education, as well as speech technology development for computer-aided language learning (CALL) applications.

**Index Terms:** L2 suprasegmental features, language education, prosody, stress, intonation.

## 1. Introduction

English is the lingua franca of our world. It is the second language (L2) most actively studied across Asia, as well as an official language or working language used in education, government, media, etc. in many regions. Hence, acquiring communicative competence in English is of prime importance. It has been estimated [1] that by 2010 there will be 2 billion English learners worldwide, and the proportion in Asia alone will exceed the number of native speakers. Second language learning, specifically pronunciation learning, involves correct perception and production of sounds in the target language. The learning process tends to be influenced by well established perceptions of sounds and articulatory motions in the primary language (L1). This cross-linguistic influence is often referred as language transfer. Negative transfer of L1 features causes inaccuracies and errors in L2 speech productions, which impedes intelligibility. Consequently, the study of L2 speech productions (i.e. the “interlanguage” of learners who have not acquired native-like proficiency) is of great interest to phoneticians, linguists, language educators, as well as technologists engaged in the development of CALL (computer-aided language learning) systems.

Language transfer occurs at both the segmental (i.e. phonetic) and suprasegmental (i.e. prosodic) levels. Furthermore, cross-linguistic transfer effects become fossilized with age and present specific challenges for adult L2 learners. Perceptual studies such as [2] indicate that while segmental and suprasegmental features all have significant influence on expert judgments regarding speaking proficiency, the latter has stronger effect. Segmental features are important

for correct word pronunciations. Suprasegmental features encode rich information structure [3] that helps the listener locate emphasized words, phrase boundaries, speech acts (e.g. statements, questions, continuations, etc.) [4, 5, 6], as well as the speaker’s attitudes and emotions. Hence, negative transfer in suprasegmental features may significantly impede intelligibility and comprehensibility of L2 speech and hamper communication.

Little is known about the nature of L2 suprasegmental features and their implications in communication and pedagogy. It is important to identify the typology of suprasegmental features which are communicatively significant. For L2 suprasegmental features that impede communication, either due to negative transfer or other reasons, we aim to generate corrective feedback to facilitate learner improvement. Our understanding of L2 suprasegmental features lags far behind that of segmental features, because of the following challenges [7]:

- High variability in acoustic features: suprasegmental phenomena are manifested as their acoustic correlates in speech, but the acoustic signal is modulated by many factors, including individual speaker variations, lexical and sentential contexts, communicative semantics, non-linguistic factors such as emotion, etc. It is difficult to derive an overall structural model that captures the relations between an acoustic feature and its underlying suprasegmental feature.
- Lack of consensus about communicative significance of suprasegmental features: Suprasegmentals can be used to express pragmatic meanings which are relative to the speaker or the discourse, e.g. a speaker may change their pitch to express disinterest in the topic. While this is a significant function of suprasegmentals, it is currently difficult to reach consensus on their meaning due to the lack of objective measures.

In order to circumvent the challenges described above, we propose an approach that is grounded on the communicative functions of suprasegmental features. We plan to record speech data from non-native English speakers, or more specifically, English learners’ speech. We designed specific sentential contexts and discourse situations, where L2 suprasegmental features may distort the intended syntactic structure or semantic interpretation of the utterance. We believe that systematic analysis of such corpora will help us understand the structural connection(s) that govern L2 suprasegmental features in relation with intelligibility and interpretation. We plan to focus our analysis on observations common across Asian Englishes. The analysis may also involve comparison with native (L1) productions, as well as “standard” non-native (L2) productions. This plan is part of a joint effort in Asia, known as the AESOP (Asian English Speech cOrpus Project) [8] initiative. To the best of our knowledge, this is one of the first data collection efforts that aims to create considerably sized corpora that captures L2

phenomena across Asian Englishes. This position paper presents our planned methodology for corpora design and collection to support investigations at the suprasegmental level.

## 2. Communicative Functions of Suprasegmental Features

We focus on two of the main communicative functions supported by suprasegmental features in English, namely, *highlighting* and *phrasing* [3]. Highlighting relates primarily with stress, which may be applied to the lexical and utterance levels. A stressed syllable in English is realized acoustically with higher intensity, longer segment duration, as well as stronger (or unreduced) articulation reflected in the spectral quality. Such prominence may be further increased in an accented syllable<sup>1</sup> at the utterance level, by means of pitch movements near or around the syllable. Phrasing relates to chunking in speech, where phrase breaks are delimited by pauses, rhythm changes or pitch movements that are not associated with highlighting.

Both highlighting and phrasing serve to communicate linguistic and paralinguistic information. Lexical stress can encode the part-of-speech of a word. Also, stress changes may occur for different inflectional forms of a given word. Utterance-level stress can mark the intended focus, which helps convey the information structure of a discourse by distinguishing between given versus new information, or background versus foreground information. Phrasing is important for conveying the syntactic structure of an utterance, which includes disambiguation between continuation versus termination, or among possible syntactic structures that may correspond to different semantic meanings, as well as discourse functions. Pitch movements in an utterance, or intonation, can further convey continuation versus termination, or speech acts, e.g. making a statement, asking a question, etc. Intonation also serves an emotive function by conveying the speaker's mood or attitude.

It is important to increase the learner's awareness of the rich informational content that is encoded in suprasegmental features. This will enable analytical perception and accurate production of spoken realizations, which should facilitate effective communication. If a suprasegmental feature does not match the intended meaning of an utterance, it should also be rectified. However, offering suprasegmental training for the language learner is not an easy task. While instructors may be able to pinpoint individual L2 suprasegmental features and offer guidance [7], it is difficult to formulate generalizable rules that can guide accurate productions. This is due to the high variability in suprasegmental features, and the lack of consensus of L2 suprasegmental features with regards to their communicative significance. This will be elaborated in the next section.

## 3. L1 Transfer in L2 Productions

As an initial step in our planned investigation of L2 suprasegmental features, we focus on L1 transfer on L2 productions [9]. Asian Englishes encompass L2 productions from a variety of mother tongues, such as Chinese, Japanese, Korean, Malay, Thai, Vietnamese, to name a few. As mentioned earlier, English contains stressed and unstressed

syllables and is often known as a stress-timed language with quasi-uniform durations between consecutive stressed syllables. English pitch dynamics also have specific linguistic and paralinguistic roles. Contrastively, the suprasegmental system of an Asian L1 language may be notably dissimilar. For example, Chinese is syllable-timed where syllables have quasi-uniform durations; which exhibits a very different rhythm compared with English. Furthermore, Chinese is a tonal language where each syllable carries a tone and the tone pattern in the pronunciation of a word determines its meaning. Hence, pitch dynamics in Chinese need to convey tonal information. As another example, Japanese is mora-timed, where a mora is a V or CV unit. There is a distinction in vowel length, such that CVV has double the duration of CV. Japanese is also a pitch accent language, where different words may have the same syllables and differ only in their pitch patterns.

The suprasegmental system of a language is realized with multiple acoustic features (i.e. intensity, duration, spectral quality and pitch) in a coordinated manner. Hence, the stark contrast between the suprasegmental systems of L1 and L2 present difficulties for language learning. Negative transfer of L1 features in L2 productions lead to inaccuracies and errors. It has been generally observed in Asian Englishes that L2 learners have difficulty with appropriate placement and realization of stress at both the word and utterance levels, as well as appropriate control of intonation variations in specific utterance contexts. For example, while English signals stress with all four cues to produce a stress-timed rhythm, it has been noted that L2 English produced by native Chinese speakers tends to preserve syllable-timed rhythm and uses the fundamental frequency as the dominant cue to signal stress. Furthermore, vowels in unstressed syllables are not sufficiently reduced because of the syllable-timed rhythm [10]. Native Japanese speakers also have problems in learning English lexical stress, with pitch being less of a problem than other prosodic features such as duration and vowel quality [11, 12]. Therefore, in comparing stress realizations between native and L2 productions, we should devote special attention towards transfer effects from specific L1 characteristics, e.g. tonal and syllable-timed rhythm, or pitch-accented and mora-timed rhythm, etc. We need to garner stronger empirical support for various phenomena in L2 suprasegmental features based on systematic, corpus-based analysis of Asian Englishes. To this end, we have carefully defined a set of basic linguistic and information structures for the purpose of eliciting L2 English speech. We intend to collect speech data for a variety of Asian Englishes using these structures. Details are provided in the next section.

## 4. Suprasegmental Features in Linguistic and Information Structures

This section presents several categories of associations between English suprasegmental features with linguistic and information structures. These include words, phrases, sentences and discourse contexts. We believe that these are major categories of importance to L2 learning, where special attention should be devoted in neutralizing negative transfer effects from L1 features. These categories constitute the corpus design for joint data collection under the AESOP [8] initiative. Illustrative examples are provided for each category.

---

<sup>1</sup> A syllable carrying pitch accent.

#### 4.1 Lexical and Compound Stress

Learners often have difficulty with stress placement, especially in polysyllabic words. In the corpus design, we plan to emphasize the relations between lexical stress and part-of-speech (e.g. 'record as a verb and re'cord as a noun), stress shifting for certain inflectional forms (e.g. 'nation and natio'nality), as well as the profile of stress patterns that are common in polysyllabic words. For example, in three-syllable words, the syllable stress patterns include:

- Initial stress, e.g. 'wonderful;
- Medial stress, e.g. a'partment; and
- Final stress, e.g. over'night)

In four-syllable words, the stress may occur in:

- the first syllable, e.g. 'elevator;
- the second syllable, e.g. a'vailable;
- the third syllable, e.g. infor'mation; and
- the final syllable, e.g. misunder'stand.

We plan to cover two- to five-syllable words.

We will also study reduced versus unreduced productions of function words, e.g. pronouns, prepositions and auxiliary verbs. While function words are normally reduced in English, there are certain sentential contexts in which function words are unreduced. Examples include:

*I can (reduced) run faster than you can (unreduced).*

*He went to a fancy dress party as (reduced) a guest, but what did he dress as (unreduced)?*

Additionally, we will include stress patterns of compound words and phrases, such as *orange juice*, *flash flood warning*, *John F. Kennedy*, etc. Stress placement is an indicator of the syntactic structure of compounding. For example, the noun phrase:

*"a [German language] teacher"*

should be distinguished from the adjectival phrase:

*"a German [language teacher]"*

through stress placement. Similar examples include:

*"a [lighthouse] keeper"* versus  
*"a light [housekeeper]"*

or

*"a dark [room curtain]"* versus  
*"a [darkroom] curtain"*

The words and compound words will be embedded in a carrier phrase for recording, in order to better control variations due to contextual factors.

#### 4.2 Utterance-level Stress and Focus

We will design text prompts based on sentential context and discourse information, to elicit speech productions that place utterance-level stress (mainly narrow focus) in the appropriate locations. Examples of narrow focus placement based on sentential contexts are:

*We need to finish the project **overnight**, not over the weekend.*

*You should buy food at the **supermarket** and not at the convenience store, because it will be much cheaper.*

*Mary's flight arrives at six in the **morning**, not six in the evening.*

An example from [Selkirk, 1995] illustrates narrow focus placement in different renditions of "Mary bought a book about caterpillars", when the sentence is used to respond to a series of questions that scaffold the contexts:

Question: *What did Mary do?*

Answer: *Mary **bought** a book about caterpillars.*

Question: *What did Mary buy?*

Answer: *Mary bought a **book** about caterpillars.*

Question: *What kind of book did Mary buy?*

Answer: *Mary bought a book about **caterpillars**.*

#### 4.3 Intonation and Phrasing

Intonation is realized as pitch variations and carries rich information about phrasing in an utterance. The latter may be referred to as an intonational phrase. A common example occurs in describing a list of objects, where a rising intonation (denoted by ↗) signifies continuation; and a falling intonation (denoted by ↘) signifies termination, e.g.

*He bought strawberries↗, pineapples↗, bananas↗ and apples↘.*

Furthermore, there are sentences for which a change in intonation will change the phrasing and the meaning. The following are some ambiguous examples, where we use the punctuation (/) to indicate the intonational phrase boundaries:

*"The fight is over, // Fred"*

*"The fight is over Fred //"* (referenced from [9])

where the intonational phrasing indicates adverbial attachment

*"They are married happily //"*

*"They are married, // happily"*

where the intonational phrasing indicates adverbial attachment. The former utterance states that they are happy, while the latter states that the speaker is happy about their marriage.

*"He is not here, // I'm afraid."*

*"He is not here, I'm afraid //"*

where the former utterance expresses fear but the latter expresses a slight regret, because he is not around.

*"Those who sold quickly // made a profit"*

*"Those who sold // quickly made a profit"*

where the intonational phrase indicates the relative clause. The former utterance states that a profit was made by those who sold quickly, but the latter utterance states that a profit was quickly made by those who sold.

*"I stopped singing //"*

*"I stopped // singing"*

where the intonational phrasing indicates the main verb. The former utterance states that I had been singing, while the latter utterance states that I stood still and was singing.

*"He [washed and brushed] his hair"*

*"He washed // and brushed his hair"*

where the intonational phrasing indicates the verb-object relations. The former utterance states that he washed his hair as well as brushed his hair, while the latter states that he washed (himself) and brushed his hair.

#### 4.4 Speech Acts

Intonation also marks the speech act of an utterance, such as whether it is a statement, request, Wh-question, Yes-No question, etc. The following are some examples of speech acts and their associated intonations:

- A declarative statement has a falling intonation, e.g. “*She returned to Hong Kong.* ↘ ”
- A Wh-question has a falling intonation, e.g. “*Where is the elevator?* ↘ ” or “*Who can give me the information?* ↘ ”
- A Yes-No question has a rising intonation, e.g. “*Would you like a glass of white wine?* ↗ ” or “*Has Jane found an apartment?* ↗ ”

### 5. Summary and Conclusions

This position paper underscores the importance of suprasegmental training in secondary language (L2) acquisition. Suprasegmental features include lexical stress, sentential stress and intonation. They are manifested in acoustic cues such as intensity, duration, fundamental frequency and spectral quality. They also convey information about linguistic and information structures, e.g. the part-of-speech of a word, syntactic structure of a compound word, narrow focus of an utterance, syntactic structure of an utterance (which is associated with its meaning), speech acts, etc. Hence, L2 learners must harness appropriate suprasegmental productions for effective communication. However, the process of L2 acquisition tends to be influenced by well-established perceptions and productions of sounds in the primary language (L1), which is referred as language transfer. Negative transfer of L1 features to L2 lead to inaccurate or erroneous productions. There is a paucity of research concerning L2 acquisition of suprasegmental phonology. We need to conduct systematic analyses of L2 suprasegmental features based on sizeable corpora, in order understand the nature of language transfer, to identify the typology of suprasegmental features with communicative significance, with the hope of generating corrective feedback that facilitates learning. We have laid out a carefully selected set of textual environments, with illustrative examples showing how suprasegmental features relate with important linguistic and information structures. We intend to use these textual environments for collecting speech data in a variety of Asian Englishes from non-native English speakers. We believe that analyses of such corpora should lead to research findings that have important implications for language education, as well as speech technology development for computer-aided language learning (CALL) applications. Additionally, the corpora should also be a valuable resource for the study of speech and language phenomena in World Englishes.

### 6. Acknowledgements

The AESOP consortium is initiated by Professor Yoshinori Sagisaka of Waseda University. Other collaborators include Professor Michiko Nakano of Waseda University, Dr. Chai Wutiwiwatchai of the National Electronics and Computer Technology Center in Thailand, Professor Sudaporn Luksaneeyanawin, Professor Tavicha Phadvivulaya and

Professor Kulaporn Hiranburana of the Chulalongkorn University, Dr. Wai-Kit Lo of The Chinese University of Hong Kong, Dr. Lan Wang of the CAS-CUHK Shenzhen Institute of Advanced Integration Technologies, and Dr. Sakriani Sakti and Dr. Dawa Idomuco from ATR. This project is partially supported by a General Research Fund (project no. CUHK416108) from the Research Grants Council, Hong Kong SAR Government.

### 7. References

- [1] Asia Economic News, 20 Feb., 2006 [http://findarticles.com/p/articles/mi\\_m0WDP/is\\_2006\\_Feb\\_20/ai\\_n16086425/](http://findarticles.com/p/articles/mi_m0WDP/is_2006_Feb_20/ai_n16086425/)
- [2] Anderson-Hsieh, J., Johnson, R. and Koehler, K., “The Relationship between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody and Syllable Structure,” *Language Learning*, 42:4, 1992.
- [3] Grice, M. and Bauman, S., “An Introduction to Intonation – Functions and Models,” in *Non-native Prosody: Phonetic Description and Teaching Practice*, Trouvain, J. and Ulrike, G. (eds), pp. 25-52, Mouton de Gruyter, 2007.
- [4] Crystal, D. *Clinical linguistics*, New York, Harper Press, 1981.
- [5] Price, P., Ostendorf, M., Shattuck-Hufnagel, S. and Fong, C. “The Use of Prosody in Syntactic Disambiguation,” *Journal of the Acoustical Society of America*, 90(6), pp. 2956-2970, 1991.
- [6] Hirschberg, J., “Communication and Prosody: Functional Aspects of Prosody,” *Speech Communication*, Volume 36, Issues 1-2, pp. 31-43, 2002.
- [7] Jilka, M., “Different Manifestations and Perceptions of Foreign Accent in Intonation,” in *Non-native Prosody: Phonetic Description and Teaching Practice*, Trouvain, J. and Ulrike, G. (eds), pp. 77-96, Mouton de Gruyter, 2007.
- [8] Sagisaka, Y., AESOP (Asian English Speech cOrpus Project) Project, Personal Communication, 2007.
- [9] Mennen, I., “Phonological and Phonetic Influences in Non-native Intonation,” in *Non-native Prosody: Phonetic Description and Teaching Practice*, J. Trouvain and U. Gut (Eds). Mouton de Gruyter, 2007.
- [10] Zhang, Y., Nissen, S.L. and Francis, A.L., “Acoustic Characteristics of English Lexical Stress produced by Native Mandarin Speakers,” *Journal of the Acoustical Society of America*, Vol. 123, no. 6, pp.4498-4513, June 2008.
- [11] Lee, B., Guion, S. and Harada, T., “Acoustic Analysis of the Production of Unstressed English Vowels by Early and Late Korean and Japanese Bilinguals”, *Studies in Second Language Acquisition*, Vol. 28, pp. 487-513, 2006
- [12] Mochizuki-Sudo, M. and Kiritani, S., “Production and Perception of Stress-related Durational Patterns in Japanese Learners of English,” *Journal of Phonetics*, Vol. 19, pp. 231-248 (1991).
- [13] Pennington, M. and Ellis, N., “Cantonese Speakers’ Memory for English Sentences with Prosodic Cues,” *The Modern Language Journal*, 86 (iii), 2000.