Phonetic Aspects of Content Design in AESOP (Asian English Speech cOrpus Project)

Tanya Visceglia1, Chiu-yu Tseng2, Mariko Kondo3, Helen Meng4 and Yoshinori Sagisaka5
Department of Applied English Ming Chuan University1
Institute of Linguistics, Academia Sinica2
SILS&LSSRL,Waseda University, Japan3
Human-Computer Communications Laboratory, Chinese University of Hong Kong4
GITI & Lg Sp Science Res Lab,Waseda University, Japan5
tvisceglia@gmail.com, cytling@sinica.edu.tw, mkondo@waseda.jp, hmmeng@se.cuhk.edu.hk, ysagisaka@gmail.com

Abstract

This research is part of the ongoing multinational collaboration "Asian English Speech cOrpus Project" (AESOP), whose aim is to build up an Asian English speech corpus representing the varieties of English spoken in Asia. AESOP is an international consortium of linguists, speech scientists, psychologists and educators from Japan, Taiwan, Hong Kong, China, Thailand, Indonesia and Mongolia. Its primary aim is to collect and compare Asian English speech corpora from the countries listed above in order to derive a set of core properties common to all varieties of Asian English, as well as to discover features that are particular to individual varieties. Each research team will use a common recording setup and share an experimental task set, and will develop a common, open-ended annotation system. Moreover, AESOP-collected corpora will be an open resource, available to the research community at large. The initial stage of the phonetics aspect of this project will be devoted to designing spoken-language tasks which will elicit production of a large range of English segmental and suprasegmental characteristics. These data will be used to generate a catalogue of acoustic characteristics particular to individual varieties of Asian English, which will then be compared with the data collected by other AESOP members in order to determine areas of overlap between L1 and L2 English as well as differences among varieties of Asian English.

1. Introduction

As English continues to grow in importance as a language for international communication throughout the world, the face of English itself is continuously changing. The blending of English with local languages and dialects in countries and regions such as Greater China, India, Malaysia and the Philippines has given rise to a wide variety of world Englishes, which exhibit rich variation in pronunciation, lexicon and grammar. Consequently, considerable interest has emerged in researching the topic of Accent and Language Identification and Recognition [1,2]. English is also being studied and spoken as a second language in more countries than ever before. Thus, a comprehensive understanding of the variation present in the dialects of English spoken in the world today is a fundamental issue for the development of English language education as well as spoken language science and technology.

Asia is home to the largest number of English learners and speakers in the world; it has been claimed that combining native and non-native speakers, India now has more people who speak or understand English than any other country in the world. Following India is the People's Republic of China [3,4]. Thus, research in Asian English dialects from a multidisciplinary perspective is urgently needed to address issues in communication, learning and technology. Research in linguistics can catalogue and analyze the range of variation present in Asian English dialects; research in speech science can implement linguistic findings into the development of language pedagogy, as well as into ICT tools and environments tailored to the requirements of Asian speaker populations.

The current research is part of the ongoing multinational collaboration "Asian English Speech cOrpus Project" (AESOP) whose aim is to build up an English speech corpus which represents the varieties of English spoken in Asia. AESOP is an international consortium of linguists, psychologists, speech scientists, technologists and educators from Japan, Taiwan, Hong Kong, China, Thailand, Indonesia and Mongolia. Its primary aim is to collect and compare English speech corpora from the countries listed above using a consistent set of core materials in order to derive a set of phonetic properties common to all varieties of Asian English, as well as to discover phonetic features that are particular to individual

dialects.

Each research team will use a common recording setup and experimental task set, and will develop a common, open-ended annotation system and platform for analysis and tool development. Each of the AESOP collaborators is independently funded, so each team is free to tailor the experimental materials and collect additional data to address language-specific and discipline-specific phenomena. Moreover, AESOP-collected corpora will be an open resource, available to the international research community free of charge.

2. Methodology

The initial stage of the phonetics aspect of this project will be devoted to designing spoken-language tasks which elicit production of a large range of English segmental and suprasegmental characteristics which, based on previous research [6], are predicted to be present in L2 speech. These include: (1) word-level features such as segmental and tonal borrowing; (2) L1-specific differences in rhythm and timing; (3) phrase boundary phenomena such as declarative falls and interrogative rises and (4) form, timing and location of pitch accents, which are used to create phrasal and sentential prominence (broad and narrow focus).

Materials designed to elicit these features include the following: 2-, 3- and 4-syllable target words of all possible stress patterns embedded in carrier sentences (for the purpose of baseline comparison) and in the following prosodic contexts: (1) at phrase boundaries in yes-no questions, wh-questions and declarative sentences and (2) in narrow-focus positions.

Two additional sets of experimental sentences have been designed to elicit production of (1) function words in stressed and unstressed positions and (2) prosodic disambiguation of syntactic structures. L2 English speakers will also be required to produce strings of alphabetic letters and numbers, and to produce a passage of read speech, the content of which was developed to include the full range of English phonemes.

2.1. Target words

We have developed a list of target word candidates for each possible stress type present in English 2-, 3- and 4-syllable words, which have been excerpted from the CMU Dictionary database [5]. Words of five syllables or more have been excluded to avoid the possible confounds of secondary and tertiary stress[1]. Selection of target words from this candidate list was based on lexical familiarity (piloted), overall frequency and stress type (based on the analyses of the CMU dictionary shown in Appendix A) and semantic versatility (to facilitate construction of experimental sentences).

We chose to include words representing a range of stress patterns and syllabicities based on our prediction that the realization of lexical stress will differ between L1 and L2 English speakers. A list of target words categorized according to syllabicity and stress type appears in Appendix B. It includes the following types: (1) 2-syllable initial stress[2], (2) 3-syllable initial stress, (3) 3-syllable medial stress, (4) 3-syllable final stress, (5) 4-syllable initial stress, (6) 4-syllable medial 1 stress, (7) 4-syllable medial 2 stress, (8) 4-syllable final stress, (9) left-headed compounds (e.g *orange* juice), (10) right-headed compounds (e.g. after*noon*).

English is a stress-timed language, one consequence of which is that stressed syllables in individual words tend to be louder, higher in pitch and longer in duration than unstressed syllables are. Moreover, the vowels in English unstressed syllables are often reduced to schwa. However, in syllable-timed languages such as Guoyu (Taiwan Mandarin) and Cantonese, the distinction between stressed and unstressed syllables is marked by reduction of syllable duration and intensity rather than by vowel reduction. Japanese is a pitch accent language and uses the mora as a timing unit. Similarly, in pitch accent languages, the presence of stress or accent does not affect vowel duration or quality. Thus, L1 speakers of syllable-timed and mora-timed languages often have trouble with stress assignment in multi-syllabic English words and tend to use inappropriate cues to differentiate stressed and unstressed syllables [7,8].

2.1.1. Target words in carrier sentences

Two representatives from each syllable and stress condition were placed in controlled experimental conditions. The same tokens were also placed in a fixed, neutral context for baseline comparisons of inherent duration and formant values. Speakers will read a list of identical carrier sentences: "I said the word XX five times". Each of these sentences contains one target word appearing in a broad-focused position two syllables removed from any phrase boundary.

2.1.2. Target words at phrase boundaries

Previous research in L2 prosody suggests that L2 speakers realize prosodic phrase boundaries

[1] Comprehensive lists of each token type have been made available to all collaborators.

[2] The 2-2 (2-syllable final stress) type has been excluded from tasks 1-3 because this type is expected to yield very similar data to that of 3-3 and 4-4.

differently from L1 speakers [9,10]. To further investigate this phenomenon, we have designed materials with target words embedded in four prosodic boundary positions: the final fall of a wh-question, the final rise of a yes-no question, the continuation rise found in multiple-clause sentences, and the final fall in declarative sentences. To realize these prosodic boundaries, L1 English speakers usually anchor the nuclear (most prominent) pitch accent to the last prominent syllable in an intonation phrase, from which they begin their rise or fall to a phrase boundary. Our design will elicit L2 productions of the acoustic features associated with phrase and sentence boundaries. An example sentence is given below, and the full set of experimental sentences can be found in Appendix C.

Target word: overnight
Boundary type: yes-no question
"Can packages be shipped overnight"

2.1.3. Target words in narrow focus

Differences have also been found between L1 and L2 English speakers' production of the pitch accent used to mark narrow focus in English [11]. We have placed each of the target words in a narrow-focus context in order to elicit these data from speakers. An example sentence is given below, and the full set of experimental sentences can be found in Appendix D.

> Target word: overnight
> We have to finish the project ***overnight***, not over the weekend.

2.2. Stressed and unstressed function words

Another consequence of stress timing in English is that function words, such as pronouns, prepositions and auxiliary verbs are usually not made prominent, since they carry a minimal semantic load. Thus, L1 speakers often reduce the vowels in function words and may even delete them in spontaneous speech. We designed one set of sentences to elicit the same function words appearing in stressed and unstressed positions. An example is given below:

> Target word: can
> "I can (reduced kən) run faster than you can (canonical kæn)."

2.3. Prosodic disambiguation

There is evidence to suggest that L1 English speakers use prosody to disambiguate different syntactic structures in identical phonetic strings [12]. The strongest use of prosodic cues for this purpose has been found in differentiation of early and late closure sentences such as the following:

> When you learn // gradually you worry more.
> When you learn gradually // you worry more.

Our materials include a small set of syntactically ambiguous sentences whose ambiguity is resolved by prosody for the purpose of investigating whether L2 speakers will produce the prosodic cues that mark differences in boundary locations in sentences of this type.

2.4. Alphabetic strings and number sequences

L1 English speakers use phrase intonation when producing alphabetic letter strings, that is to say, when they are spelling out names or other words. Number sequences, such as telephone and credit card numbers, are also configured in fixed prosodic patterns [13]. Little data have yet been reported on L2 English speakers' production of alphabetic letter and number strings. However, they are of primary importance to the development of speech technology, as most computer interfaces require speakers to spell their names and addresses, or to provide their phone, identification or credit card numbers. To elicit spelling of letter and number strings, we have designed a series of questions, which requires speakers to spell the name and address of their sponsoring institution and to repeat a series of number strings that will appear on a screen.

2.5. The north wind

Following the tasks described above, each speaker will read Aesop's fable "The North Wind" aloud. This passage is recommended by the IPA for the purpose of eliciting all phonemic contrasts that occur in English. Full text of "The North Wind" appears in Appendix E. Another phonetically balanced passage in L1 will be included in the same task in order to compare L1 and L2 global speech rate, pitch range and prosodic cues for information structure.

2.6. Data collection and annotation

Recorded data will be analyzed as it is collected, so that materials can be revised to pinpoint emerging trends and patterns in the data. Experimental materials will be constructed in such a way that they can be widely distributed across university departments as well as among other AESOP collaborators. AESOP has agreed to design recording tasks to be carried out in naturalistic settings instead of soundproof chambers so that recording can easily be performed by research assistants in relatively quiet environments. The speaker population will consist primarily of

undergraduate and graduate students. Expected recording time is approximately one hour per session/speaker. In Taiwan, the first and second authors will collect data from approximately 210 male and 210 female speakers. Analysis of these data will generate a catalogue of phonetic characteristics particular to Taiwan English, which will then be compared with the data collected by other AESOP collaborators in order to determine areas of overlap and difference. AESOP members also plan to develop a common protocol for file formats and annotation.

3. Conclusion

A wide range of phenomena are encompassed in the description and analysis of Asian L2 English; the scope of this project includes pedagogical, scientific and technological areas of inquiry. This paper describes our initial step in developing a systematic understanding of the range of phonetic variation present in Asian L2 English. Collaborating members will collect spoken English data from L1 Japanese, Hong Kong Cantonese, Thai, Indonesian, Mongolian and Guoyu (Taiwan Mandarin) speakers. The resulting database will facilitate speech-related research over a wide range of academic disciplines and will provide a valuable resource for the development of ICT tools and environments tailored to the requirements of Asian speaker populations. Other research interests represented by the AESOP international collaboration project are open at this stage. We welcome feedback and participation from L2 researchers in all fields.

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 Phadvibulaya and Professor Kulaporn Hiranburana of the Chulalongkorn University, Dr. Wai-Kit Lo, Dr. Pauline Lee and Alissa Harrison 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.

References

[1]. Piat, M. D. Fohr, I. Illina. 2008. “Foreign accent identification based on prosodic parameters,” (759-762) Proceedings of Interspeech 2008 September 22-26, 2008, Brisbane, Australia

[2]. Nariai, T. & Tanaka, K. 2008. “A study of pitch patterns of Japanese English analyzed via comparative linguistic features of Japanese and English” (pp. 776-779) Proceedings of Interspeech 2008 September 22-26, 2008, Brisbane, Australia

[3]. Crystal, D. “Subcontinent Raises Its Voice” Guardian Weekly: Friday 19 November 2004.

[4]. Zhao, Y. and K.P. Campbell. 1995. "English in China". World Englishes 14 (3): 377–390.

[5]. http://www.speech.cs.cmu.edu/cgi-bin/cmudict

[6]. Anderson-Hsieh, J. Johnson, R. & Koehler, K. 1992. “The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable stucture,” *Language Learning* (42), 529-555.

[7]. Tajima, K., Port, R. & Dalby, J. 1997. “Effects of temporal correction on intelligibility of foreign-accented English,” Journal of Phonetics (25) 1-24

[8]. Jian, H.L. 2004. “On the syllable timing in Taiwan English” In Proceedings of Speech Prosody 2004 Nara, Japan. International Speech Communication Association.

[9]. Viger, T. 2007. “Fundamental Frequency in Mandarin and English: Comparing First- and Second-Language Speakers” Unpublished doctoral dissertation, CUNY Dissertations in Linguistics. Ph.D. Program in Linguistics: The City University of New York.

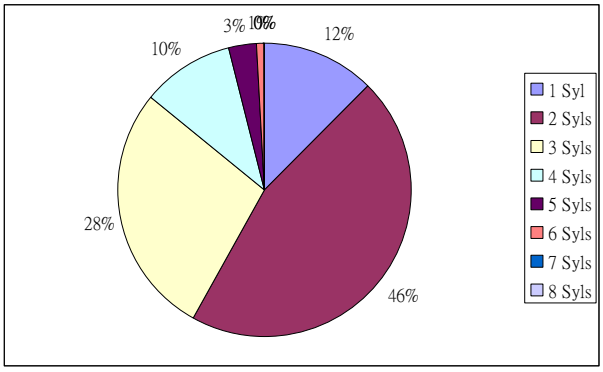
[10]. Wennerstrom, A. 1998. "Intonation as cohesion in academic discourse: a study of Chinese speakers of English" *Studies in Second Language Acquisition* (20), 1-25.

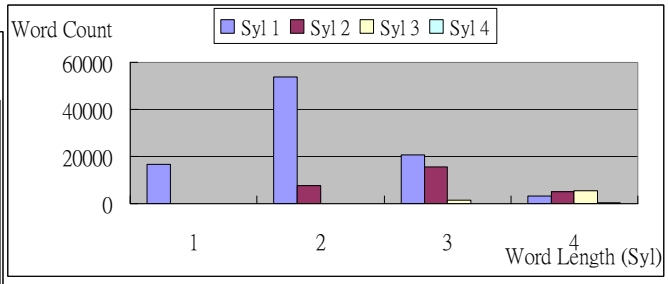
[11]. McGory, J. 1997. “The Acquisition of Intonation Patterns in English by Native Speakers of Korean and Mandarin.” Unpublished doctoral dissertation, Ohio State Dissertations in Linguistics. Department of Linguistics: The Ohio State University

[12]. Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956-2970.

[13]. Aylett, M. (2004). Merging data driven and rule based prosodic models for unit selection TTS. In Proceedings of 5th ISCA Speech Synthesis Workshop, Pittsburgh, PA, USA

Appendix A: CMU electronic dictionary analyses by frequency, syllabicity and stress type





Stress Distribution of English Words from 1- to 4-syl (96.22% of 133,693 words from CMU Electronic Dictionary

Appendix B: Target words by syllabicity, stress type and experimental condition

	2-1	3-1	3-2	3-3	4-1	4-2	4-3	4-4	LH	RH
Y-N (rise)	money	Wonderful	apartment	overnight						white wine
WH (fall)					elevator	available	information	misunderstand	Supermarket	
Cont.(rise)					January	experience	California	Vietnamese	Department store	
Decl. (fall)	morning	Video	tomorrow	Japanese						afternoon
Narrow focus	Money morning	wonderful Video	Apartment tomorrow	Overnight Japanese	Elevator January	Available Experience	Information California	Misunderstand Vietnamese	Supermarket department store	white wine afternoon

Appendix C: Target words at prosodic boundaries

Yes-no questions (IP rise)
1. 2-1 Do you need any money?
2. 3-1 Did he go to the hospital?
3. 3-2 Has Jane found an apartment?
4. 3-3 Can packages be shipped overnight?
5. RH Would you like a glass of white wine?

Wh-questions (IP fall)
1. 4-1 Where is the elevator?
2. 4-2 When will Bill be available?
3. 4-3 Who can give me the information?
4. 4-4 Why are these instructions so easy to misunderstand?
5. LH Where is the nearest supermarket?

2-clause declaratives

Continuation rise (iP rise)
1a. 4-1 Do you know that in December and January,
2a. 4-2 Although Fred didn't have any experience
3a. 4-3 When Sue left this evening for California
4a. 4-4 If you're interested in learning Vietnamese
5a. LH If you want to check out the new department store

10. You should buy food at the supermarket, not at the convenience store. It will be much cheaper

Final fall (IP fall)
1b. 2-1 the sun rises at seven in the morning.
2b. 3-1 he had no trouble learning how to make a video
3b. 3-2 she said she would call me tomorrow
4b. 3-3 I think it will be easier than Japanese
5b. RH we can go this afternoon

Appendix D: Target words in narrow focus

1. I don't think you stole the money, but you probably stole the car.
2. I said I want to go to the hospital, not the airport.
3. Bill is living in an apartment house, not a one-family house.
4. We have to finish the project overnight, not over the weekend.
5. Why are you drinking white wine instead of your usual whiskey?
6. You're going to the second floor. Why are you taking the elevator instead of the stairs?
7. These apartments are already rented. We want to see a list of available apartments.
8. I'm sorry, I didn't hear you. Should I go to the information desk or the service desk?
9. I didn't misunderstand your instructions; I chose not to follow them.

11. Does Scott usually take his vacation in January or in February?
12. Unlike most companies, we believe that

experience is more important than training.
13. Teresa is flying to California tomorrow, not to Texas.
14. I think you're wrong. That couple is speaking Vietnamese, not Thai.
15. Why are you waiting at the book store? We agreed to meet at the department store today.
16. Mary's flight arrives at six in the morning, not six in the evening.
17. We should have a video recording as well as an audio recording.
18. If we leave tomorrow instead of today, there will be less traffic.
19. Bill doesn't speak Japanese, but he does speak other Asian languages.
20. My schedule is full every morning. Can we meet in the afternoon this week?

Appendix E: The North Wind

The North Wind and the Sun were disputing which was the stronger when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shone out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two