Perception of English Suprasegmental Features by Non-Native Chinese Learners

Shuang Zhang, Kun Li, Wai-Kit Lo and Helen Meng[1]
Human-Computer Communications Laboratory
Shun Hing Institute of Advanced Engineering
The Chinese University of Hong Kong
{zhangs, kli, wklo, hmmeng}@se.cuhk.edu.hk

Abstract

This study is an initial attempt to assess the knowledge and perception of English suprasegmental features by non-native (Chinese) learners. The suprasegmental features covered are: lexical stress, utterance-level stress, intonation and phrasing, as well as prosodic disambiguation. Our findings suggest the need to enrich pronunciation training in terms of knowledge and production of English suprasegmental features. Learners have particular difficulty with stress patterns of long polysyllabic words, unreduced function words, intonation of Wh-questions and continuation phrases, as well as prosodic disambiguation for semantic interpretation. Our findings also show that the learners are capable of perceiving acoustic realizations of the suprasegmental features, which brings performance improvements between the knowledge test and perceptual test. This validates the value of developing speech technologies that can support perceptual and productive training of English suprasegmental features on a computer-aided language learning (CALL) platform.

Index Terms: English suprasegmental, perceptual test, language learning

1. Introduction

The long-term goal of this project is to develop speech technologies that assist second language (L2) acquisition of English by adult Chinese learners, focusing specifically on suprasegmental phonology (i.e. prosody). English is the *lingua franca* of our world. It is of prime importance that we acquire communicative competence in English. 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. The process of second language acquisition is interfered by well-established perceptions of sounds and articulations in the primary language (L1). Chinese and English have stark contrasts linguistically. We often observe notable L1 (i.e. Chinese) interferences with L2 (i.e. English) speech in phonetics (i.e. segmental phonology) as well as prosodics (i.e. suprasegmental phonology). While both impede the intelligibility of L2 speech, perceptual studies suggest that suprasegmentals may have a stronger effect [2]. The interferences are ingrained with age and hamper acquisition of proficiency, especially for adult L2 learners. Improvements require persistent and individualized perceptual and productive training. Recent advancements in speech technologies have opened up new possibilities in computer-aided language learning [3]. Major thrusts lie in applying automatic speech recognition to the learner's non-native speech and devising algorithms for computer-aided pronunciation training (CAPT). Existing works predominantly address phonetic deviances in L2 speech (cf. native speech), e.g. [4]. While there is growing appreciation of suprasegmental training for language learners, few existing studies have investigated L2 prosodic deviances in non-native English uttered by adult Chinese learners. This work is an initial attempt to understand the perception of English suprasegmental phenomena by non-native Chinese learners, which will guide our subsequent efforts in developing speech technologies that support pronunciation training in English suprasegmental phonology.

Our focus is on suprasegmental features that relate to the communicative functions of *highlighting* and *phrasing* [5], which may be applied at both the lexical and utterance levels to convey linguistic and paralinguistic information. Lexical stress can encode the part-of-speech of a word. 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 disambiguation between continuation versus termination, for conveying the syntactic structure of an utterance that corresponds to different semantic meanings, and for communicating speech acts and relevant discourse or emotive functions.

2. Scope

The scope of our study include several categories of associations between English suprasegmental features with linguistic and information structures [5]. They include:

(i) Lexical stress – covering the primary, secondary and unstressed syllables of polysyllabic words, as well as reduced versus unreduced function words.

(ii) Utterance-level stress – covering the narrow focus in an utterance that relates to *sentential context and discourse* information.

(iii) Intonation and phrasing – relating to continuation or termination, as well as speech acts such as declarative statements, Wh-questions and Yes-No questions.

(iv) Prosodic disambiguation in semantically ambiguous sentences.

3. Organization of Perceptual Tests

We have designed a list of textual prompts and invited a native American English speaker to record with a natural speaking style. We designed a questionnaire that includes a list of questions relating to the suprasegmental categories laid out in Section 2. Each perceptual test is conducted in two phases:

- The first phase aims to elicit the subject's prior *knowledge* about the suprasegmental features, by writing down his/her answers on the questionnaire. The answer option "I don't know" is also presented for all the questions. Since no audio presentation is involved, this phase does not include the suprasegmental category of prosodic disambiguation (see Section 2).

- The second phase aims to elicit the subject's *perception* of suprasegmental realizations. The relevant speech recording is played for each question before the subject is

[1] Corresponding author: Please direct questions and comments to hmmeng@se.cuhk.edu.hk

asked to write down the answer. Again, the answer option "I don't know" is presented for all the questions. All suprasegmental categories in Section 2 are covered.

We have recruited a total of 58 native speakers of Putonghua (44 postgraduate and 14 undergraduate students from all majors across our university) to take the perceptual tests. This subject pool has received 11 years of English instruction on average.

4. Lexical Stress

4.1. Polysyllabic words

Our study of knowledge and perception of lexical stress by Chinese learners covers different stress patterns (with primary, secondary or no syllable stress) in polysyllabic words (between 3 to 6 syllables). Two-syllable words are omitted due to their simplicity. The words include:

- 3-syllable words:
  hospital, processing, tomorrow,
  department
- 4-syllable words:
  elevator, available, experience,
  transportation, misunderstand
- 5-syllable words:
  refrigerator, interchangeable transformational,
  anniversary, unacceptable, documentation,
  experimental, intellectually, unambiguously
- 6-syllable words:
  eligibility, characterization, intercontinental

For a given word, the subject is asked to mark '1' under the syllable with primary stress, '2' under the syllable with secondary stress, or check under "I don't know" if he/she does not know the stress position(s) for the word. An excerpt of the questionnaire is shown in Table 1.

Word:	ae	ro	plane	I don't know
aeroplane	1		2	

Table 1: Excerpt of the questionnaire relating to the study of lexical stress.

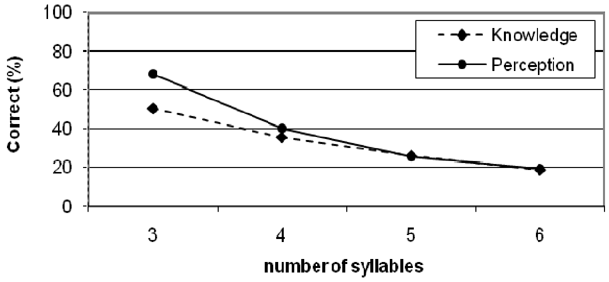

Figure 1. Average stress identification accuracies for words with different syllable lengths.

In evaluation, a word is considered correct if its entire stress pattern is correct. Results from the *knowledge* test show that 31% of the words are labeled correctly. Results from the *perceptual* test rose to 36%. Average stress identification accuracies for words with different syllable lengths are shown in Figure 1. We observe that:

- Stress identification accuracy decreases dramatically as the syllable length of the word increases, possibly because many more stress patterns are possible for longer words.
- After listening to the audio, subjects are able to perceive the word stress in order to improve stress identification accuracies, especially for shorter words (with 3 to 4 syllables). Such improvement is not observed for longer words (with 5 to 6 syllables), possibly due to many more possible stress patterns.

4.1.1. Common error patterns

(i) *Words with a single stressed syllable:* All the three-syllable words and two of the four-syllable words fall into this category. Subjects generally perform well for these (55% based on knowledge and 66% base on perception, as shown in Figure 1). The word that is particularly problematic is "processing", for which 66%of the instances were labeled with the wrong pattern based on knowledge, but subjects are able to perceive the correct pattern from the speech audio. Examples are shown in Table 2.

Word	Knowledge		Perception	
hospital	● – –	67%	● – –	69%
● – –	– ● –	21%	– ● –	17%
processing	● – –	14%	● – –	71%
● – –	– ● –	66%	● – ○	12%
			● ○ –	10%

Table 2: Stress patterns labeled by the subjects for three-syllable words. Correct stress patterns are shown in the leftmost column. '●' denotes primary stress, '○' secondary stress and '–' unstressed. Patterns labeled based on *knowledge* are shown in the middle column. The word "processing" is particularly problematic. Low frequency patterns are omitted. The right column shows how labeling accuracies change with *perception* of the speech recording. Correct stress patterns are in black and incorrect ones in grey.

(ii) *Words with both primary and secondary stress:* Long words tend to contain primary stress and secondary stress syllables. We observe that subjects can often distinguish between syllables that carry stress (especially primary stress) and syllables that do not. However, there is often confusion between the labeling of primary versus secondary stress. To a lesser extent, secondary stress syllables may sometimes be labeled as unstressed. Listening to the audio may not lead to improved performance in stress pattern identification. Examples are shown in Table 3.

Word	Knowledge		Perception	
elevator	● – ○ –	21%	● – ○ –	33%
● – ○ –	● – – –	22%	● – – –	36%
	○ – ● –	21%	– – ● –	9%
	– – ● –	19%	○ – ● –	7%
transformational	○ – ● – –	38%	○ – ● – –	26%
○ – ● – –	● – ○ – –	31%	● – ○ – –	60%
	– – ● – –	10%	● – – – –	3%
	● – – – –	9%	– – ● – –	2%
misunderstand	○ – – ●	5%	○ – – ●	5%
○ – – ●	– ● – –	28%	● – – ○	21%
	○ ● – –	19%	● ○ – –	12%
	● ○ – –	17%	– ● – –	12%
	● – – –	9%	○ ● – –	10%
	● – – ○	5%	– ● – ○	10%
	– ● – ○	5%	● – – –	9%
			– – – ●	7%
intercontinental	○ – ○ – ●–	2%	○ – ○– ● –	5%
○ – ○ – ● –	○ – ● – – –	22%	● – ○ – – –	22%
	● – ○ – – –	14%	○ – – – ● –	12%
	– – ● – – –	14%		

Table 3: Stress patterns elicited from the subjects, for words that have syllables with primary stress, secondary stress or no stress. Patterns with low occurrences are omitted. Correct stress patterns are in black and incorrect ones in grey.

4.2. Reduced / Unreduced Function Words

Function words serve grammatical functions in English and carry little lexical meaning. Although function words are normally reduced in English, there are certain sentential contexts in which function words are unreduced. In this test, we included four sentences (see Table 4) with 21 function words and asked our subjects to identify the reduced and unreduced function words, first based on their knowledge and subsequently based on perception.

If the party wasn't for Mary, then who was it *for*?
Jane saw a picture of the boy she was fond *of*.
John went to visit the woman he had written *to*.
He was invited to a costume party as a guest, but what did he dress *as*?

Table 4: Test materials for reduced/unreduced function words. Function words are in gray and unreduced function words are in italic.

Results (see Figure 2) show that most of the reduced function words (over 77%) can be identified correctly based on the subjects' knowledge. The performance improves to 87% with perception of the audio. Only 42% of the unreduced function words were identified correctly based on knowledge. But the performance improves significantly to 72% with speech perception, which suggests that the subjects are able to perceive unreduced function words. We should also note that the unreduced function words happen to be located at the sentence-end positions in our test materials. The declination effect may induce errors whereby a subject labels an unreduced function word as a reduced one.

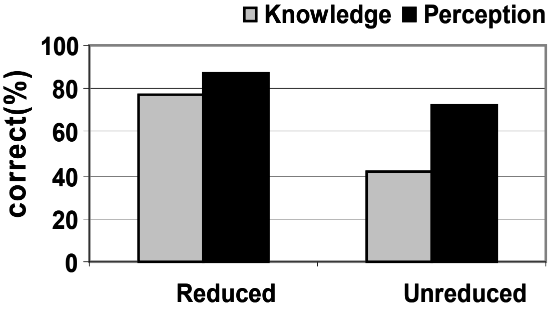

Figure 2. Performance on identification of reduced versus unreduced function words in the knowledge and perception tests.

5. Utterance-level Stress

Our study of utterance-level stress includes knowledge and perceptual tests on narrow focus.

5.1. Narrow focus

The test materials include eight sentences (examples in Table 5) with contextual information. In the knowledge test, subjects are presented with pairs of context and the sentence. They are then asked to circle the word that should carry emphasis, or select the answer option "I don't know". The perceptual test includes the same procedures, together with presentation of speech recordings.

Results show that for the knowledge test, subjects can identify the correct word with narrow focus for 86% of the sentences. This performance improves to 98.5% for the perceptual test. A possible reason for the good performance is that the contextual information helps our subjects interpret the sentence. Also, the performance improvement suggests that subjects are able to perceive emphasis well.

[Context] Can doctors give blood tests at this clinic? No. you should go to a hospital for blood test.
[Context] How will I carry all these boxes up to the fifth floor? You should take the elevator instead of the stairs.
[Context] Do you buy fruit at the farmer's market? No. I usually buy fruit at the supermarket because they stay open later.
[Context] have you been trained to do this job? No. But I think experience is more important than training.
[Context] Why can't I travel? You need documentation before you can travel.

Table 5: Examples of test materials for utterance-level stress. The words carrying narrow focus for the eight sentences are respectively: *hospital, elevator, supermarket, experience,* and *documentation*.

6. Intonation and Phrasing

To assess the subject's knowledge and perception in intonation and phrasing, we design the test materials that include declarative statements, Wh-questions, Yes-No questions and continuation rise. Eight locations are marked in the six sentences (see Table 6), where subjects are asked to indicate whether there should be a rising or falling intonation, or choose the answer option "I don't know".

Do you need any money___?
She returned to Hong Kong___.
Where is the nearest supermarket___?
Has Jane found an apartment___?
In December and January___, the sun rises at seven in the morning___.
If we are going to have a discussion___, we should have it this afternoon___.

Table 6: Test materials for intonation and phrasing. Subjects are asked to fill in the blanks, indicating rising (↗) or falling (↘) intonation, or select the answer option "I don't know".

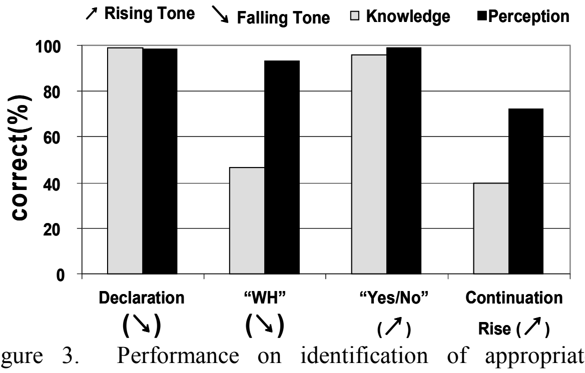

Figure 3. Performance on identification of appropriate intonation for declarative statements, Wh-questions, Yes-No questions and continuation phrases.

Results are shown in Figure 3. We observe that:

- Subjects are generally unaware that Wh-questions should carry a falling intonation and indicated the correct answer for only 47% of the sentences. However, they are able to perceive the correct intonation from the audio, which brings the accuracy to over 93%.
- Subjects are also unaware that phrasal continuation should be accompanied with a rising intonation. Correct labeling based on knowledge was obtained for only 40% of the sentences. This is comparable to random guessing, given the limited answer options. However, subjects are able to

perceive the correct intonation from the audio and the accuracy improves to 72%.

- Performance is very high for the remaining categories, i.e. declarative statements and Yes-No questions.

7. Prosodic Disambiguation

This task is different from the others in that we only included the perceptual test. Subjects are presented with sentence text without punctuation. Each of the six sentences has two possible semantic interpretations, which are provided to the subjects as indicated (see Table 7). Upon hearing the speech recording, each subject is asked to select the appropriate interpretation for the sentence, or select the option "I don't know". The prosodic realization in the speech recording serves to disambiguate between the possible semantic interpretations for each sentence. The subjects need to base their decision on both their knowledge and perception of prosodic disambiguation.

1.	*[Context 1] Fred and John are arguing. They both want Mary to be on their team.* The fight is over Mary.
	[Context 2] Mary doesn't know why everyone else has already left the boxing arena. The fight is over, Mary.
2.	*[Context 1] I'm not sure if I should let Peter into my English class.* He is a good boy, isn't he?
	[*Context 2] Peter always helps the younger children with their homework.* He is a good boy, isn't he?
3.	*[Context 1] Whenever May goes, everyone stops and talks to her.* She knows everyone, doesn't she?
	[Context 2] Should I introduce May to the team? I think she has met everyone before. She knows everyone, doesn't she?
4.	*[Context 1] The feeling of the couple on the marriage.* They are married happily.
	[Context 2] The feeling of the speaker on the marriage. They are married, happily.
5.	*[Context 1] The speaker is scared because John is not here* He is not here, I'm afraid.
	[Context 2] The speaker is sorry that John is not here. He is not here; I'm afraid.
6.	*[Context 1] A profit is made by those who sold something quickly.* Those who sold quickly, made a profit.
	[Context 2] A profit is made quickly by those who sold something. Those who sold, quickly made a profit.

Table 7: Test materials for prosodic disambiguation. Each test sentence is semantically ambiguous and possible interpretations are indicated.

Results are shown in Figure 4. Subjects rarely select the option "I don't know" option. Given that there are primarily two answer options, random guessing should give accuracies in the vicinity of 50%. The accuracies range from 4% to 83%. This suggests that our subjects may generally be unaware of how suprasgemental features (such as pausing and intonational phrasing) are used for semantic disambiguation, or they are unable to perceive the relevant prosodic realizations. The first and last sentences may suffer less, perhaps because they involve a straightforward association between pausing and phrasing.

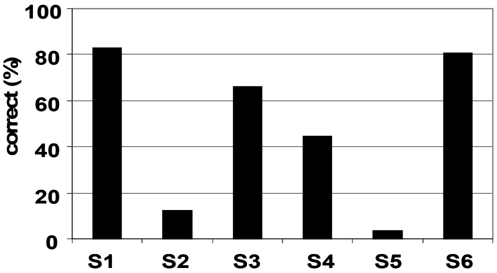

Figure 4. Performance on prosodic disambiguation across six example sentences (S1 to S6, as listed in Table 7).

8. Conclusions

This study is an initial attempt to assess the knowledge and perception of English suprasegmental features by non-native (Chinese) learners. The suprasegmental features covered are: lexical stress, utterance-level stress, intonation and phrasing, as well as prosodic disambiguation. Our findings suggest the need to enrich pronunciation training in terms of knowledge and production of English suprasegmental features. Learners have particular difficulty with stress patterns of long polysyllabic words, unreduced function words, intonation of Wh-questions and continuation phrases, as well as prosodic disambiguation for semantic interpretation. Our findings also show that the learners are capable of perceiving acoustic realizations of the suprasegmental features, which brings performance improvements between the knowledge test and perceptual test. This validates the value of developing speech technologies that can support perceptual and productive training of English suprasegmental features [5-8] on a computer-aided language learning (CALL) platform.

9. Acknowledgments

This project is partially supported by the General Research Fund (CUHK416108) from the HKSAR Government Research Grants Council. We wish to thank Miss Alissa Harrison, Dr. Pauline Lee of Independent Learning Centre, instructors of the English Language Teaching Unit of CUHK, Professor C-Y Tseng of Academia Sinica, Miss Tanya Visceglia of Ming Chuan University and Professor Mariko Kondo of Waseda University for their support of this project.

10. References

[1] Asia Economic News, 20 February, 2006. 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] Eskenazi, M. "An Overview of Spoken Language Technology for Education," Speech Communication, 2009.

[4] Meng, H., Lo, Y. Y., Wang, L. and Lau W. Y., "Deriving Salient Learners' Mispronunciations from Cross-Language Phonological Comparisons", Proc. of ASRU2007.

[5] Meng, H., Tseng, C., Kondo, M., Harrison, A. and Viscelgia, T., "Studying L2 Suprasegmental Features in Asian Englishes: A Position Paper", Proc. Interspeech 2009.

[6] Hincks, R., "Speech synthesis for teaching lexical stress," TMH-QPSR 2002, vol. 44, pp. 153-156.

[7] Sundstrom, A., "Automatic Prosody Modification as a means for Foreign Language Pronunciation Training," Proceedings of ESCA ETRW STiLL, 1998, pp. 49-52.

[8] Delmonte, R., "Prosodic Modeling for Automatic Language Tutors," Proceedings of ESCA ETRW STiLL 1998, pp. 57-60.