

# A Corpus-Based Online Pronunciation Learning System: The Pedagogical Applications of a Spoken Corpus

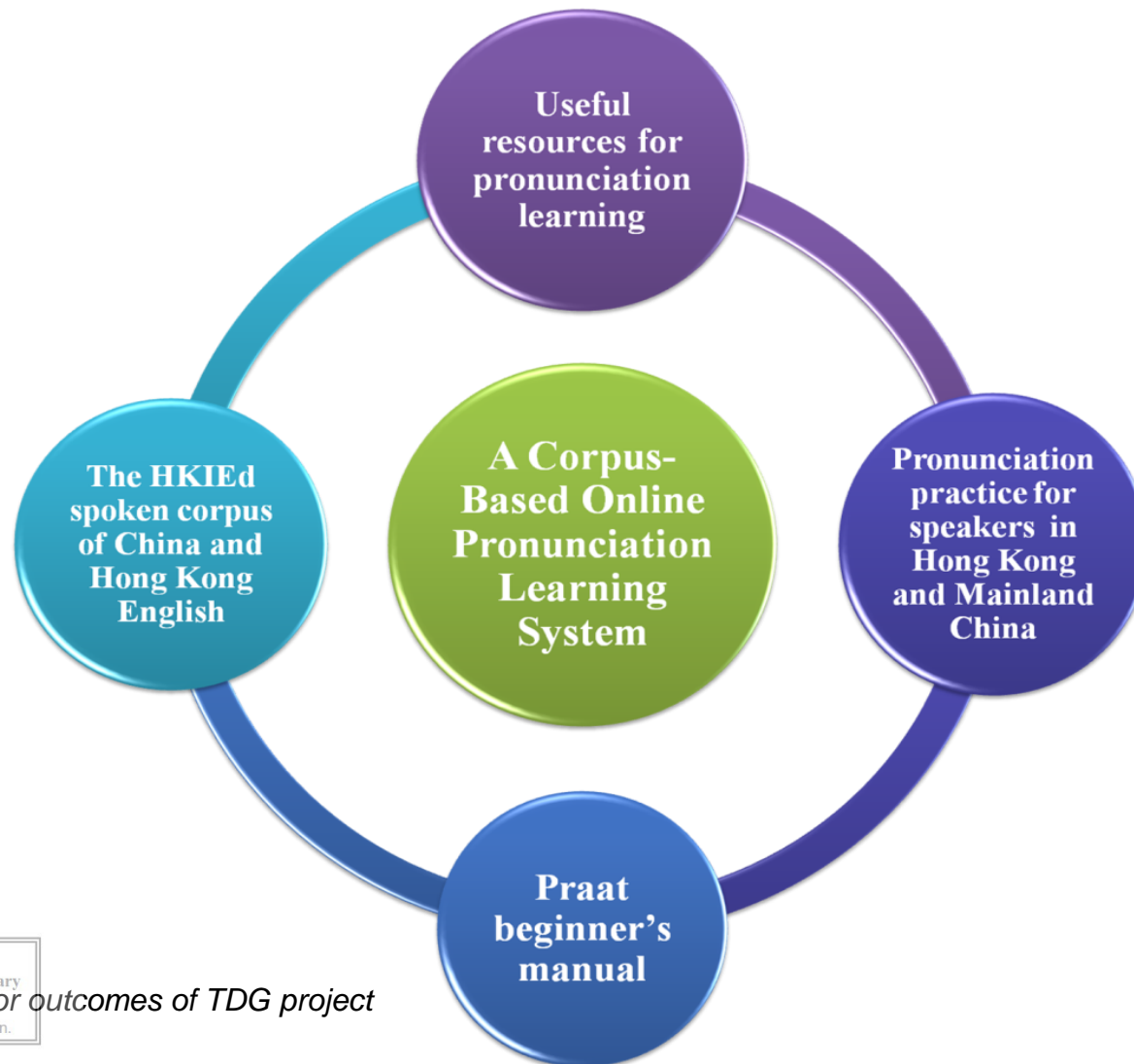
CHEN Hsueh Chu; CHAN WONG Pui Man;  
CHAN Ka Yin; CHEE Epin;  
WANG Lixun; WANG Qian

[hsuehchu@ied.edu.hk](mailto:hsuehchu@ied.edu.hk)



The Hong Kong Institute of Education

# A Corpus-Based Online Pronunciation Learning Website



# Our spoken corpus


A spoken corpus of the English of Hong Kong and Mainland Chinese learners has been established and analyzed in order to

- (a) identify Hong Kong and Mainland university students' recurrent difficulties in English pronunciation learning
- (b) suggest possible remedies that may reduce or eliminate such difficulties

# Our spoken corpus

This corpus contains roughly 14.5-hour recording data from two speech tasks

- 4 hours of reading-aloud
- 8 hours of interview
  - 22 Mainland students
  - 50 Hong Kong students



The spoken corpus is to identify Hong Kong and mainland university students' pronunciation features in...

➤ **Three** areas of segmental features

➤ Vowels, consonants and syllable structures

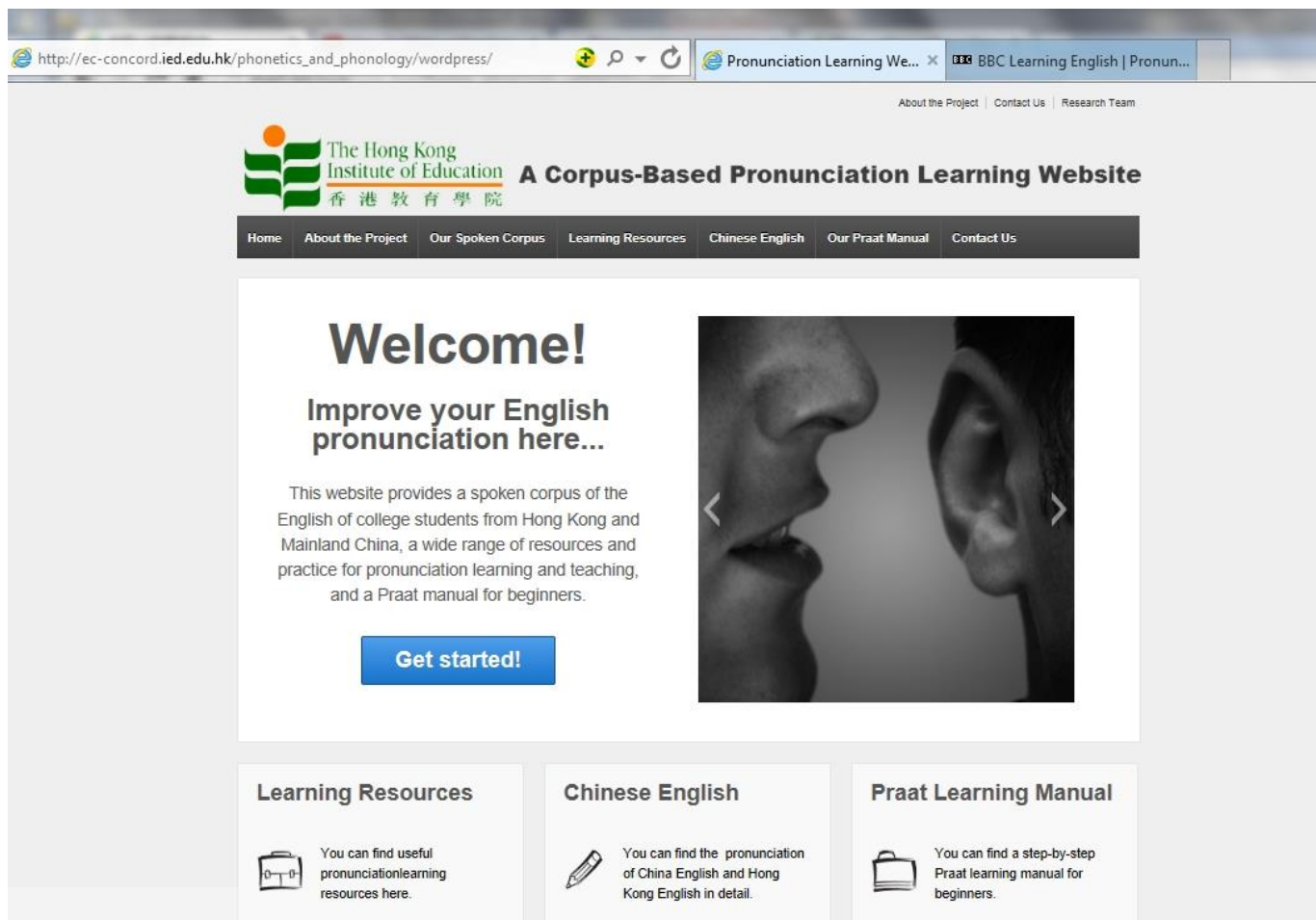
➤ **Four** areas of suprasegmental features

➤ word-level stress, pausing, linking and intonation

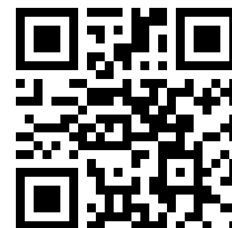
# Users can

1. browse the recordings by speakers' biographic information  
(e.g., the birthplace, the major, the age)
2. search the recordings by the pronunciation features they are interested in.

# A Corpus-Based Pronunciation Learning Website



The screenshot shows the homepage of the A Corpus-Based Pronunciation Learning Website. The browser address bar displays [http://ec-concord.ied.edu.hk/phonetics\\_and\\_phonology/wordpress/](http://ec-concord.ied.edu.hk/phonetics_and_phonology/wordpress/). The website header includes the logo of The Hong Kong Institute of Education (香港教育學院) and the title "A Corpus-Based Pronunciation Learning Website". A navigation menu contains links: Home, About the Project, Our Spoken Corpus, Learning Resources, Chinese English, Our Praat Manual, and Contact Us. The main content area features a "Welcome!" message, the text "Improve your English pronunciation here...", and a description: "This website provides a spoken corpus of the English of college students from Hong Kong and Mainland China, a wide range of resources and practice for pronunciation learning and teaching, and a Praat manual for beginners." A blue "Get started!" button is present. To the right is a black and white image of a person's profile, focusing on the mouth and ear. Below the main content are three sections: "Learning Resources" with a book icon and text "You can find useful pronunciation learning resources here.", "Chinese English" with a pencil icon and text "You can find the pronunciation of China English and Hong Kong English in detail.", and "Praat Learning Manual" with a book icon and text "You can find a step-by-step Praat learning manual for beginners."



# Main page of corpus website



<http://corpus.ied.edu.hk/phonetics/>







## Browse Raw Records

Hometown: [ANY] Gender: [Either] Age: [ANY] [Filter Speaker](#)

HK21	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK22	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK23	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK24	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK25	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK26	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK27	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK28	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK29	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK30	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK31	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK32	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK33	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK34	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK35	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK36	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK37	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>
HK38	<a href="#">Passage One</a>	<a href="#">Passage Two</a>	<a href="#">Interview Question</a>	<a href="#">Interview Answer</a>



## Search Passage

Suprasegmental (Passage One):

Step 1: [NONE SELECTION]

Segmental (Passage Two):

Step 1: Consonants

Step 2: Fricatives

Step 3:

- ☐ [NONE SELECTION]
- ☒ Dental fricatives / θ / & / ð /
- ☐ Labiodental fricatives / v / & / ʃ /
- ☐ Post-alveolar fricatives / ʒ / & / ʝ /
- ☐ Alveolar fricatives / s / & / z /
- ☐ Glottal fricative / h /

Step 4:

- ☒ [NONE SELECTION]
- ☐ / θ /->/ ʃ /
- ☐ / θ /->/ s /
- ☐ / θ /->/ ʈ /
- ☐ / ð /->/ n /

Passage Two

[Search Result]

[Speaker: ML1](#) [Speaker: ML2](#) [Speaker: ML3](#)  
[Speaker: ML4](#) [Speaker: ML5](#) [Speaker: HK1](#)  
[Speaker: HK2](#) [Speaker: HK3](#) [Speaker: HK6](#)  
[Speaker: HK7](#) [Speaker: HK9](#) [Speaker: ML7](#)  
[Speaker: ML8](#) [Speaker: ML10](#) [Speaker: ML11](#)  
[Speaker: HK4](#) [Speaker: HK5](#)



# Passage Two

Speaker: HK2

Hometown: Hong Kong

Gender: Female

Age: 20

English Level: (N/A)

The First Language:

Cantonese

The Second Language:

Mandarin

The Third Language:

English

Year of Learning English:

13 year(s)

Education Level: Bachelor

Major: Liberal Studies

Education

Notes: "HK" and "ML" in the speaker's name stand for "Hong Kong" and "Mainland" respectively.

Speaker's Record:



US Sample:



UK Sample:



Marking Notes of Passage 2

## The Boy Who Cried Wolf

There was once a poor shepherd boy who used to watch his flocks in **the** (/ð/→/d/) fields next to a dark forest near **the** (/ð/→/d/) foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran down to the village shouting 'Wolf, Wolf.' As soon as **they** (/ð/→/z/) heard him, the villagers all rushed from their homes, full of concern for his safety, and two of his cousins even stayed **with** (/θ/→/f/) him for a short while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful. However not long after, a wolf that had just escaped from the zoo was looking for a change from its usual diet of chicken and duck. So, overcoming its fear of being shot, it actually did come out from the forest and began to threaten the sheep. Racing down to the village, the boy of course cried out even louder than before. Unfortunately, as all **the** (/ð/→/d/) villagers were convinced **that** (/ð/→/z/) he was trying to fool **them** (/ð/→/d/) a **third** (/θ/→/f/) time, they told him, 'Go away and don't **bother** (/ð/→/d/) us again.' And so the wolf had a feast.

Notes

1. The words marked in **red** are those specific features spotted.
2. The words marked in **blue** are those mispronounced words which are different from the featured ones.

## Feature List

### Syllable structure change

Omission of consonant(s)

Omission of vowel(s)

Insertion of consonant(s)

### Consonants

Dental fricatives /θ/ & /ð/ ✓

Labiodental fricatives /v/ & /f/

Post-alveolar fricatives /ʒ/ & /ʃ/

### Vowels

/ɒ/

/ɪ/

/ɔ:/



The Hong Kong  
Institute of Education Library

For private study or research only.  
Not for publication or further reproduction.

English Level: (N/A)

The First Language:  
Cantonese

The Second Language:  
English

The Third Language: (N/A)  
Year of Learning English:  
10 year(s)

Education Level: Bachelor  
Major: Liberal Studies  
Education





Notes: "HK" and "ML" in the  
speaker's name stand for "Hong  
Kong" and "Mainland" respectively.


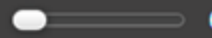



- ☒ Inappropriate pausing (IP)
- ☒ Absence of CV linking (AL)
- ☒ Inappropriate intonation (II)


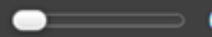



Full Passage Recording: 

Marking Criterion of Passage 1

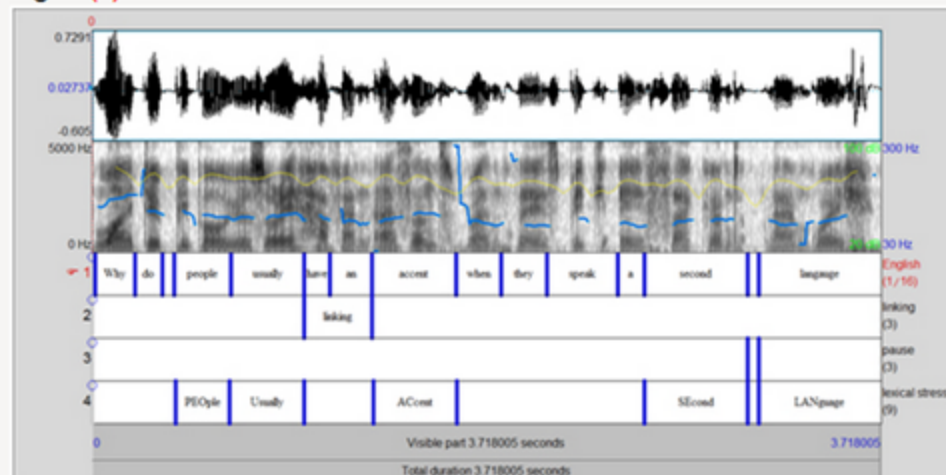
### Sentence 1:

Speaker's Record:   0:04  


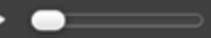


US Sample:   0:03   


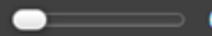



UK Sample:   0:04   


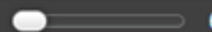



Why do PEOPle Usually have an ACcent (AL) when they speak a (AL) Second  
LANguage (II)?



### Sentence 2:

Speaker's Record:   0:08  

US Sample:   0:06   

UK Sample:   0:07   



Speaker: HK2

Hometown: Hong Kong

Gender: Female

Age: 20

English Level: (N/A)

The First Language:

Cantonese

The Second Language:

Mandarin

The Third Language:

English

Year of Learning English:

13 year(s)

Education Level: Bachelor

Major: Liberal Studies

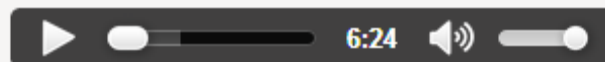
Education

Notes: "HK" and "ML" in the speaker's name stand for "Hong Kong" and "Mainland" respectively.

Remarks:

1. The sentences in black are from interviewee, the sentences in blue are from the interviewer.
2. Use "... " to indicate the fragmented sentences.
3. Use "(. . .)" to signify a word or a phrase that you cannot understand after trying several times.
4. We put the words that are guessed from the context in the brackets ( ).

Speaker's Record:



[1.Interviewer: What do you think of Singapore?]

HK2: Singapore is like... I think some...it's like Hong Kong, but it has...it has a lot of the rules, the forest, and people need to...need to followed, and so for example, we cannot eat the gums in the Singapore, I think it is not very good for me, because that I...I am the one, I love freedom, and I don't want others controls me and manage me. I think that the Singapore's government do a lot to controls the people's life, but I think that the Singapore is a good place, because that he is a place to Universal (Studio), the place to watch the... I don't know how to explain. Yes, Universal Studio, that place is very beautiful and awesome. I love the Universal Studio very much. Even though I didn't go to Singapores because..., but my friends had went to Singapore, they posted a lot of the photos about the Universal Studio, then I really really want to go there, just take a lot of the photos, yeah.

[2.Interviewer: First, tell me something about your family?]

HK2: My family is a big family; I have one sister...two...I have three sisters, and one brothers and my mother and my father. And in total my family has 7 people in there and that yes



The Hong Kong  
Institute of Education Library

For private study or research only.  
Not for publication or further reproduction.

# Main page of corpus website



<http://corpus.ied.edu.hk/phonetics/>



# How to make use of this corpus-based learning website?

## For Graduate students:

A workshop on Praat applications was held at HKIEd campus to help students use the Praat in analysing their English pronunciation data.



## For Undergraduates:

Students of HKIEd are encouraged to use this system during or after class for self-access.

## For Teachers

This website provides abundant pronunciation teaching resources and pronunciation practice for teachers and learners.

# THE APPLICATIONS OF THE CORPUS:

## THE STUDY OF THE INTELLIGIBILITY IN ESL CLASSROOM



# Aims of the study

- To identify the pronunciation aspects that are crucial for intelligible pronunciation in actual L2 classroom from in-service teachers' point of view;
- To investigate how teachers help their students successfully understand L2 classroom input through teachers' self reflections;
- To explore the most frequently-used pronunciation teaching strategies teachers used to rectify students' pronunciation.



# Questionnaire Survey

➤ We made use of the features found in our corpus and compile a questionnaire to investigate

1. the **intelligibility**,
2. **pronunciation adjustment strategies**, and
3. **pronunciation teaching** in L2 classroom

# Participants

We have collected **47** questionnaires from in-service primary and secondary school teachers in Hong Kong and Mainland China.

- **25** are from Hong Kong
- **22** are from mainland China

# Major findings

The pronunciation features of your students that may impede the intelligibility in classrooms

## HK teachers

- “Voiced TH/ð/ is pronounced as/d/” (M=3.60),
- “Syllable-timed rhythm”(M=3.56)
- “Deletion of final /l/” (M=3.48)

## ML teachers

- “Syllable-timed rhythm” (M=3.27)
- “Inappropriate sentence stress” (M=3.23)
- “Absence of linking” (M=3.23)

# The most common pronunciation-related adjustment strategies

## HK teachers

- “Contraction avoidance” (M=3.88)
- “Sentence stress emphasis” (M=3.88)
- “Repetition” (M=3.84)

## ML teachers

- “Repetition” (M=4.45),
- “Speech rate modification” (M=4.18)
- “Segmental-level enunciation” (M=4.18)

# The most frequently-taught aspects of pronunciation in your language lesson

## HK teachers


- “Consonant clusters”(M=3.28)
- “Word stress” (M=3.24)
- “Sentence stress” (M=3.16)
- “Pausing” (M=3.16)

## ML teachers

- “Voiced or voiceless” (M=3.95)
- “Word stress” (M=3.95)
- “Intonation” (M=3.86)
- “Sentence stress” (M=3.82)

# The most common pronunciation teaching strategies

- The result reveals that ML teachers apply significantly *more* pronunciation teaching strategies than their HK counterparts.
- Both ML and HK teachers deemed “**Modeling the correct pronunciation**” and “**Reading aloud**” as their most frequently-used teaching strategies in their classroom.

- 
- Teachers' responses reveal that the features compiled by the researchers are quite common to their students in both ML and HK.
  - Most of the teachers do not think the features impede the intelligibility in classroom so much as they have been very familiar with their students accents.

But how about

HK teachers vs. ML students?

ML teachers vs. HK students?

HK teachers vs. Minority students in HK?

# Next Stage

- The existing corpus will be further expanded to 20 hours by collecting more authentic recordings in a natural setting,  
  
e.g. students' group discussion, classroom interaction



# Next Stage

## **Next project:**

Corpus-based feature analyses of Hong Kong/Mainland university students' spoken English

# Next Stage

- The aim of this project is to develop the research team's expertise in corpus construction and linguistic analysis at **discourse**, **syntax** and **phonology** levels.

It is hoped that the multi-facet analysis of this spoken corpus will make valuable contributions to the field of corpora and language learning.

# Q & A





Thank You!

