

## AESOP-ILAS Corpora Specifications

### I. Abstract

1. **Name:** AESOP-ILAS (Asian English Speech cOrpus Project - Institute of Linguistics, Academia Sinica) Corpora
2. **Language:** L2 English by Taiwan Mandarin speakers, corresponding speech by L1 American speakers, and some corresponding L1 Mandarin by Taiwan speakers
3. **Aims:** Prosodic features of Taiwan L2 English
4. **Content:** AESOP-ILAS 1 and AESOP-ILAS 2
5. **Time of Recording:** (1) AESOP-ILAS 1: 2009.07-2010.09  
(2) AESOP-ILAS 2: 2011.11-2012.07
6. **Data Type:** Microphone speech(Sennheiser PC155 headset microphone; See [Appendix 1](#) for further information)
7. **Number of Speakers:** 540 in total; the number of male and female speakers is approximately equal in both groups
  - (1) AESOP-ILAS 1: 12 native speakers of North American English (L1) and 488 native speakers of Taiwan Mandarin (L2)
  - (2) AESOP-ILAS 2: 10 native speakers of North American English (L1) and 30 native speakers of Taiwan Mandarin (L2)
8. **Data Size:** 14 GB (approximately 812 hours) in total; including sound files, corresponding text, and automatic HTK aligned files for phone boundaries, some of which have been manually adjusted
  - (1) AESOP-ILAS 1: 8.64 GB; 500 hours (approximately 1 hour per speaker)
  - (2) AESOP-ILAS 2: 5.42 GB; 312 hours (approximately 7.8 hours per speaker)
9. **Data Contents:**
  - (1) AESOP-ILAS 1: 6 elicited read speech tasks, 1 fully aided computer-prompted dialogue task, and 1 partially aided picture description task
  - (2) AESOP-ILAS 2: 4 elicited read speech tasks (including one Taiwan Mandarin task) and 1 fully aided computer-prompted dialogue task

## II. Recording Setup

### 1. Technical specifications:

- (1) Signal File Format: \*.wav
- (2) Sampling Rate: 16kHz
- (3) Bit Rate: 16-bit
- (4) Number of Channels: mono channel recording

2. **Recording Software:** TWNAESOP Recording Tool (designed based on the *CUHK-SIAT recording tool.exe*, which was developed by Chinese University of Hong Kong in collaboration with Shenzhen Institutes of Advanced Technology, with modifications made to meet the requirements of the AESOP project.)

### 3. Acoustical Environment:

- (1) AESOP-ILAS 1: quiet room, such as a classroom, a lab, or an instructor's office
- (2) AESOP-ILAS 2: The Phonetics Lab, Institute of Linguistics, Academia Sinica  
The participants wore a head-mounted microphone positioned about 1.5 to 2 cm below the mouth and about 1 cm away from the face in order to reduce the amount of breathing noise introduced into sound recordings.

## III. Introduction

The AESOP-ILAS speech corpus is especially designed for the Taiwan division of the multinational research project AESOP (Asian English Speech Corpus Project), featuring L2 English speech by native speakers of Taiwan Mandarin. The aim of AESOP is to build up an open resource of English speech corpus representing the varieties of English spoken in Asia. It was agreed in 2008 that each participating member will use the same recording platform to collect a common set of core data so that a set of core properties common to all varieties of Asian English, as well as to features that are particular to individual varieties can be developed.

AESOP-ILAS is funded by the Chiang Ching-kuo Foundation for International Scholarly Exchange (DB002-D-08. 2009.7.1-2012.12.31.). It aims to build up a corpus of the English spoken in Taiwan as an open resource and to investigate a wide range of communicative phonetic and prosodic features in Taiwan English at the segmental, lexical, phrasal, and discourse levels, rather than focusing on specific

and individual phenomena.

The principal investigator of this project is Dr. Chiu-yu TSENG, Distinguished Research Fellow and Director of the Institute of Linguistics, Academia Sinica, who specializes in phonetic interdisciplinary research, corpus and computational linguistics, especially corpus based prosody analysis and modeling. The three co-investigators are Professor Tanya VISCEGLIA of National Yang Ming University, a native English speaker who specializes in cross-linguistic analysis of Taiwan Mandarin and English phonetic features, as well as characteristics of L2 prosody, Professor Helen MENG of The Chinese University of Hong Kong, who specializes in Cantonese, Mandarin, and English speech characteristics, as well as development of automatic speech recognition (ASR) and other speech technologies, and Professor Yoshinori SAGISAKA of Waseda University, the convener of AESOP, who specializes in computational linguistics and cross-linguistic modeling of prosody.

The AESOP-ILAS speech corpora are 14 GB in total, featuring approximately 812 hours of sound files. The corpora are separated into 2 parts: AESOP-ILAS 1 and AESOP-ILAS 2.

AESOP-ILAS 1, which is the core properties of AESOP, focuses on eliciting production of a large range of segmental and suprasegmental characteristics. It is 8.64 GB (500 hours) and includes L1 English speech data by 12 American English native speakers and L2 English speech by 488 Taiwan Mandarin speakers. The recording time of each speaker is approximately 1 hour. However, while the controls used for the core design included the phonetics aspects of English segmental and supra-segmental in 1 hour of recording time per speaker, the collected data did not include sufficient phonotactic variations or a wide range of prosodic and pragmatic aspects of L2 English as well. In order to collect more speech data to remedy data sparsity, we have designed a different set of materials, that is, AESOP-ILAS 2. We review some of the major corpora including TIMIT, IViE, UME-ERJ and CASSAESOP. We also discussed with the CASSAESOP team as well as the Japan AESOP team at Waseda University and reached a consensus to collect some common data of mutual interests. We believe the AESOP-ILAS 2 can better solve the above mentioned problems. It is 5.42 GB (312 hours) and includes L1 English speech data by 10 American English speakers and L2 English speech data by 30 Taiwan Mandarin speakers. The recording time of each L1 speaker is approximately 5.25 hours and 8.7 hours for each L2 speaker.

AESOP-ILAS is released in April, 2015 for use of academic research, including sound files, corresponding text, and automatic HTK aligned files for phone boundaries, some of which have been manually adjusted. The corpora should be useful for research and development in language teaching, language modeling, phonetic research and applications to speech synthesis and recognition. The copyrights of AESOP-ILAS Corpora belong to Academia Sinica and it is authorized to ACLCLP to release and distribute.

#### IV. Data Contents

##### A. AESOP-ILAS 1

AESOP-ILAS 1 consists of 8 recorded tasks: 6 elicited read speech tasks, 1 fully aided computer-prompted dialogue task, and 1 partially aided picture description task. Total recording time for each speaker is approximately 1 hour. Descriptions of the 8 tasks appear below:

- (1) Task 1: Target Words in Carrier Sentences  
20 sentences: Each sentence includes 1 of 20 target words (1-4 syllable words of varying stress patterns selected from the CMU dictionary), each embedded in the simple carrier sentence “I said X five/ ten times.”
- (2) Task 2: Target Words at Phrase Boundaries  
15 sentences: The 20 designed target words in Task 1 are embedded in three different phrase boundary positions: sentence-final position of yes-no questions and wh- questions, and at the first phrase boundary in 2-phrase sentences.
- (3) Task 3: Target Words in Contrastive Stress Positions  
20 sentences: The 20 designed target words in Task 1 are embedded in contrastive stress or narrow-focus positions.
- (4) Task 4: Production of Reduced and Unreduced Function Words  
5 sentences: Function words are embedded in stressed and unstressed contexts.
- (5) Task 5: Prosodic Disambiguation  
5 sets of 2 sentences: Speakers are required to use prosody to disambiguate two lexically similar sentences.

(6) Task 6: Reading Passage: “The North Wind and the Sun”

This passage is recommended by International Phonetic Association (IPA) for the purpose of eliciting all phonemic contrasts in English from non-native speakers or speakers of regional dialects. It consists of 144 syllables, 113 words, 5 sentences (8 main clauses and 5 subordinate clauses), and 3 paragraphs.

(7) Task 7: Computer-Prompted Dialogues

17 sets of sentences, which constitute a dialogue between a reservation agent at an airline and a customer who is trying to make an airline reservation over the phone.

(8) Task 8: Picture Description Task

Speakers must answer 5 questions related to a picture of a man with a grocery list who is shopping at a supermarket. [\* This section has not been transcribed.]

## B. AESOP-ILAS 2

### 1. Corpora Review:

In order to design the contents of AESOP-ILAS 2, we reviewed some of the major corpora including TIMIT, IViE, UME-ERJ and CASSAESOP. A brief introduction to the major corpora is listed below in Table 1.

Table 1: Referred major corpora

Corpus Name	Institution	Year of construction	Data contents
TIMIT	MIT	1993	A. 460 phonetically-compact sentences B. 2 calibration sentences C. 1890 randomly selected sentences
IViE	University of Oxford	1998	A. 22 phonetically-controlled sentences with different grammatical structures sentences B. A passage of “The Cinderella Fairy Tale” C. A retold version of the Cinderella fairy tale D. A map task E. Discussion on a given topic
UME-ERJ	Tokyo University	2000	A. 8 sentence sets (S1-S8) 1. 460 TIMIT-based phonetically-balanced sentences

			<ol style="list-style-type: none"> <li>2. 32 sentences including phoneme sequences difficult for Japanese to pronounce correctly</li> <li>3. 100 sentences designed for test set</li> <li>4. 94 sentences with various intonation patterns</li> <li>5. 120 sentences with various accent and rhythm patterns</li> </ol> <p>B. 5 words sets (W1-W5)</p> <ol style="list-style-type: none"> <li>1. 302 minimal-pair words</li> <li>2. 300 phonemically balanced words</li> <li>3. 109 words with various accent patterns</li> </ol>
CASSAESOP	CASS	2010	<p>A. AESOP core sentences: the same as AESOP-ILAS 1</p> <p>B. The text pieces from UME-ERJ</p> <p>C. A passage of “The Cinderella Fairy Tale” from IViE</p> <p>D. CASS-E: 415 English sentences to cover a large variety of intonation variations</p> <p>E. CASS-C: 124 Chinese tokens of various syllables, words, sentences, dialogues and paragraphs</p>

## 2. Data Design:

AESOP-ILAS 2 consists of 5 recorded tasks: 4 elicited read speech tasks (including one Taiwan Mandarin task) and 1 fully aided computer-prompted dialogue task. Total recording time for each speaker is approximately 7.8 hours. Descriptions of the 5 tasks appear below:

(1) Task 1: Phonotactic Variations from Phonetically-rich Isolated Words

*The syllabified CMU Pronouncing Dictionary (version: 0.6, 2009) and The Word Frequency List* were used as source material. We removed unsuitable and elsewhere-appeared words and finally got a list of 5400 words. These words were assigned to 10 L1 speakers (each covered 600 words) and 30 L2 speakers (each covered 200 words) respectively.

(2) Task 2: Broad and Narrow Focus Sentences

176 broad and narrow focus sentences were selected from CASSAESOP (CASS-E).

(3) Task 3: A Long Reading Passage: “The Cinderella Fairy Tale”

This long passage of “The Cinderella Fairy Tale” was chosen by IViE corpus and is considerably longer than the “The North Wind and the Sun” in AESOP-ILAS 1. It consists of more varied intonations and discourse prosody in read narrative. The text contains a total of 759 words (1,000 syllables) in 14 paragraphs of 93 independent clauses and 49 dependent clauses.

(4) Task 4: DCT (Discourse Completion Tasks): Situational Dialogue

The DCT is designed by Professor Michiko NAKANO’s research group in Waseda University. It’s a well-established and widely adopted tool to investigate speech acts such as request and refusal performances in second language learning. A total of 86 2-speaker A-B dialogues were selected and tailored to be socially and culturally Taiwan appropriate.

(5) Task 5: A Short Taiwan Mandarin Reading Passage: “The North Wind and the Sun”

According to the 2011 AESOP Steering Committee decision, “The North Wind and the Sun” chosen by International Phonetic Association (IPA) was designated to be the commonly-shared L1 core data. It’s especially designed for the purpose of L1 influence on L2 language such as language transfer and tone borrowing.

The data collected from Task 2 to Task 4 could serve to conduct cross-linguistic comparison with the data collected from other AESOP corpora as well as IViE project.

## V. Collected Data

### A. AESOP-ILAS 1

Recording of AESOP-ILAS 1 was conducted at ten institutions in Taiwan, which aided in the process of data collection. Total number of speakers, gender distribution, and corpus size are given in Table 2.

Table 2: AESOP-ILAS 1 collected data

L1/ L2 English	# of speakers	Gender (M/F)	Corpus Size (GB)
L1 English	12	6/6	0.25
L2 English (recorded from 10 research institutes or universities*)	488	231/257	8.33
Total	500	237/263	8.58

**\*Institutions:**

Institute of Linguistics, Academia Sinica  
 Institute of Information Science, Academia Sinica  
 National Taiwan University  
 National Tsing Hua University  
 Graduate Institute of Foreign Literatures &  
 Linguistics, National Chiao Tung University

Institute of Communications Engineering,  
 National Chiao Tung University  
 National Chung Cheng University  
 National Taiwan Normal University  
 National Taipei University of Technology  
 Ming Chuan University

**B. AESOP-ILAS 2**

In total, Corpus of AESOP-ILAS 2 contains 5,400 phonetically-rich isolated words, 176 broad focus and narrow focus sentences, 1 long passage of “The Cinderella Fairy Tale” , 86 discourse completion tasks (DCT), and 1 Taiwan Mandarin short passage of “The North Wind and the Sun” . Since the recording time needed for AESOP-ILAS 2 is far more than that for AESOP-ILAS 1, each speaker was recorded in several separate sessions in order to avoid repetition fatigue. A brief summary of data content of AESOP-ILAS 2 appears in Table 3 below.

**Table 3: AESOP-ILAS 2 collected data**

Set	Token	# of L1/L2 speaker	Size/Speaker (L1/L2)	Total (L1/L2)
Phonetically-rich isolated words	5,400	10/30	600/200 (60/20 overlapped words)	6,000/6,000
Broad focus and narrow focus sentences	176		176/88	1,760/2,640
The Cinderella Fairy Tale	1		1/1	10/30



DCT Situational Dialogue	86 sets, 492 sentences		492/492	4920/14760
Taiwan Mandarin “The North Wind and the Sun”	1	0/30	0/1	0/30

## VI. Corpus Structure

1. **Structure:** Recordings are stored in separate subdirectories.

2. **Terminology:**

- (1) L1: Native North American English speech  
L2: Taiwan Mandarin English speech
- (2) Speaker coding scheme:  
Gender (F: Female; M: Male) + Native language (1: L1; 2: L2) + Speaker serial number (001-999)
- (3) adjust: manually adjusted HTK aligned files (not all files are manually adjusted)
- (4) corpus: Texts (can be opened with NotePad, WordPad)
- (5) phn: automatic HTK aligned files for phone boundaries
- (6) wav: Sound files

3. **File Name:**

- (1) AESOP-ILAS 1: Record #(of Recording session)\_# (of files).wav (corpus)
- (2) AESOP-ILAS 2: Recording session\_# (of files).wav (corpus)

## VII. Speaker Background Information

At the beginning of each recording session, the speaker was asked to enter some personal information in the log-in window, including gender, native language, education level, academic discipline, and L2 speakers’ self-evaluated English proficiency (shown in Table 4 below).

Table 4: Speaker background information

Gender	Native Language	Education Level	Academic Discipline	self-evaluated English proficiency
Male	English	Ph.D.	Arts	Excellent

Female	Taiwan Mandarin	Master	Business Adm.	Good
	Taiwanese	College	Education	Average
	Hakka	High School	Engineering	Poor
	Austronesia	Primary School	Humanities	
	Other	Other	Law	
			Medicine	
			Sciences	
			Social Sciences	
			Other	

**A. AESOP-ILAS 1**

**1. Gender:**

	<b>L1 (# of speakers)</b>	<b>L2 (# of speakers)</b>
<b>Male</b>	6 (50%)	231 (47%)
<b>Female</b>	6 (50%)	257 (53%)

**2. Native Language:**

	<b># of speakers</b>
<b>English</b>	12 (2%)
<b>Taiwan Mandarin</b>	385 (77%)
<b>Taiwanese</b>	93 (19%)
<b>Hakka</b>	10 (2%)

**3. Education Level:**

	<b>L1 (# of speakers)</b>	<b>L2 (# of speakers)</b>
<b>Ph.D.</b>	3 (25%)	14 (2%)
<b>Master' s degree</b>	6 (50%)	237 (49%)
<b>College</b>	3 (25%)	237 (49%)

**4. Academic Discipline:**

	<b>L1 (# of speakers)</b>	<b>L2 (# of speakers)</b>
<b>Art</b>	0 (0%)	13 (3%)

<b>Business Administration</b>	1 (8%)	44 (9%)
<b>Education</b>	4 (34%)	5 (1%)
<b>Engineering</b>	0 (0%)	178 (36%)
<b>Humanities</b>	4 (34%)	114 (23%)
<b>Law</b>	0 (0%)	3 (1%)
<b>Medicine</b>	0 (0%)	12 (2%)
<b>Sciences</b>	2 (16%)	77 (16%)
<b>Social sciences</b>	0 (0%)	9 (2%)
<b>Other</b>	1 (8%)	33 (7%)

#### 5. L2 Speakers' Self-evaluated English Proficiency:

	<b># of speakers</b>
<b>Excellent</b>	11 (2%)
<b>Good</b>	117 (24%)
<b>Average</b>	293 (60%)
<b>Poor</b>	67 (14%)

#### B. AESOP-ILAS 2

##### 1. Gender:

	<b>L1 (# of speakers)</b>	<b>L2 (# of speakers)</b>
<b>Male</b>	5 (50%)	15 (50%)
<b>Female</b>	5 (50%)	15 (50%)

##### 2. Native Language:

	<b># of speakers</b>
<b>English</b>	10 (25%)
<b>Taiwan Mandarin</b>	29 (73%)
<b>Taiwanese</b>	1 (3%)

##### 3. Education Level:

	<b>L1 (# of speakers)</b>	<b>L2 (# of speakers)</b>
--	---------------------------	---------------------------

<b>Ph.D.</b>	2 (20%)	2 (7%)
<b>Master' s degree</b>	3 (30%)	21 (70%)
<b>College</b>	5 (50%)	7 (23%)

#### 4. Academic Discipline:

	<b>L1 (# of speakers)</b>	<b>L2 (# of speakers)</b>
<b>Art</b>	1 (10%)	8 (27%)
<b>Business Administration</b>	1 (10%)	1 (3%)
<b>Education</b>	2 (20%)	3 (10%)
<b>Engineering</b>	1 (10%)	8 (27%)
<b>Humanities</b>	2 (20%)	2 (7%)
<b>Law</b>	0 (0%)	0 (0%)
<b>Medicine</b>	0 (0%)	0 (0%)
<b>Sciences</b>	0 (0%)	4 (13%)
<b>Social sciences</b>	1 (10%)	4 (13%)
<b>Other</b>	2 (20%)	0 (0%)

#### 5. L2 Speakers' Self-evaluated English Proficiency:

	<b># of speakers</b>
<b>Excellent</b>	11 (27%)
<b>Good</b>	14 (35%)
<b>Average</b>	15 (38%)
<b>Poor</b>	0 (0%)

## VIII. Appendix

### ● Appendix 1: Microphone specifications

#### Sennheiser PC155 Product Features and Technical Details

## PC155 The ultimate gaming tool

- ▶ **XL, velour padded ear caps** block out surrounding noise and give you ultimate wearing comfort for heavy duty gaming
- ▶ With the unique **3D, 360-degree sound experience**, sound comes at you from all sides including above and below
- ▶ The **USB adapter with integrated sound card** lets you take your own personal extreme surround sound with you wherever you play



Our best pro-gamers headset, the PC 155 USB lets you play, watch, talk or listen almost anytime, almost anywhere. It will keep you one step ahead of any attackers and right in the middle of the action.

- ▶ Adjustable headband for the perfect fit
- ▶ Adjustable, bendable, pivotable, noise-canceling microphone improves the clarity of your communication, and gives you more comfort and flexibility
- ▶ In-line volume control and mic mute switch lets you decide volume intensity and have a side conversation in privacy
- ▶ Can be used with CD/DVD/MP3 players and for IP-based communications via a detachable USB adapter that reveals 2 x 3.5 mm plugs
- ▶ Single-sided cable keeps chin and neck free so you can concentrate on your game
- ▶ Cable wind-up clip keeps excess cable wire - between headphones and adapter - out of the way when you play.

#### Technical data

##### Headphones

Frequency response .....	18 – 22,000 Hz
Impedance.....	32 Ω
SPL (at 1 KHz, 1 Vrms).....	114 dB
Cable length .....	3 m / 9.8 ft
Connector Plug .....	USB adapter and 2 x 3.5mm jack

##### Microphone

Frequency response .....	80 Hz – 15,000 Hz
Pick-up pattern.....	Uni-directional
Sensitivity .....	-38 dB
Impedance.....	-2 kΩ

#### Accessories

##### Included:

- ▶ Detachable USB adapter
- ▶ Wind-up cable clip

##### Optional:

- ▶ Replacement foam ring ear cushions
- ▶ Headband pad, also available in leatherette
- ▶ Microphone windscreen