

Comparison of the inter-accent variation of articulation rate  
between English and English with imitated Chinese accent  
speech



Mr. Komchit Taweessablert

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

An Independent Study Submitted in Partial Fulfillment of the  
Requirements  
for the Degree of Master of Arts in Linguistics  
Department of Linguistics  
FACULTY OF ARTS  
Chulalongkorn University  
Academic Year 2022  
Copyright of Chulalongkorn University

การเปรียบเทียบการแปรของคำอัครการเปล่งเสียงระหว่างภาษาอังกฤษที่พูดแบบปกติกับ  
ภาษาอังกฤษที่พูดเลียนแบบสำเนียงจีน



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาอักษรศาสตรมหาบัณฑิต  
สาขาวิชาภาษาศาสตร์ ภาควิชาภาษาศาสตร์  
คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2565  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Independent Study Title      Comparison of the inter-accent variation of articulation  
rate between English and English with imitated Chinese  
accent speech  
By                                      Mr. Komchit Taweessablamlert  
Field of Study                      Linguistics  
Thesis Advisor                      Assistant Professor SUJINAT JITWIRIYANONT,  
Ph.D.

---

Accepted by the FACULTY OF ARTS, Chulalongkorn University in Partial  
Fulfillment of the Requirement for the Master of Arts

INDEPENDENT STUDY COMMITTEE

..... Chairman  
(Associate Professor PITTAYAWAT PITTAYAPORN,  
Ph.D.)  
..... Advisor  
(Assistant Professor SUJINAT JITWIRIYANONT,  
Ph.D.)  
..... Examiner  
(Assistant Professor PAVADEE SAISUWAN, Ph.D.)



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

คมจิต ทวีทรัพย์ลีลาเลิศ : การเปรียบเทียบการแปรของค่าอัตราการเปล่งเสียงระหว่างภาษาอังกฤษที่พูดแบบปกติกับ  
ภาษาอังกฤษที่พูดเลียนแบบสำเนียงจีน. ( Comparison of the inter-accent variation of  
articulation rate between English and English with imitated Chinese accent  
speech) อ.ที่ปรึกษาหลัก : ศศ. ดร.ศุจินันท์ จิตวิริยนนท์

งานวิจัยนี้ได้เปรียบเทียบอัตราการเปล่งเสียงของภาษาอังกฤษสองสำเนียงที่ผลิตโดย Nigel Ng นักแสดงเดี่ยว  
ไมโครโฟนและยูทูเบอร์ที่เป็นที่รู้จักทางสื่อสังคมออนไลน์ในขณะที่กำลังแสดงบทบาท Uncle Roger และขณะที่รับ  
บทบาทเป็นตัวเขาเอง ขณะที่กำลังแสดงบทบาท Uncle Roger คุณ Nigel Ng ผลิตถ้อยคำด้วยภาษาอังกฤษที่พูด  
เลียนแบบสำเนียงจีนเพื่อจับเน้นคุณลักษณะที่เป็นภาพจำของชาวเอเชีย และเพื่อสร้างความแตกต่างจากภาษาอังกฤษที่ค่อนข้าง  
ใกล้เคียงกับภาษาอังกฤษแบบมาตรฐาน วัตถุประสงค์ของงานวิจัยนี้มีขึ้นเพื่อตรวจสอบการแปรของอัตราการเปล่งเสียงระหว่าง  
ถ้อยคำสองสำเนียงที่ผลิตโดยผู้พูดคนเดียวกัน จากการวิเคราะห์อัตราการเปล่งเสียงของ Nigel Ng พบว่าทั้งสองสำเนียงมี  
อัตราการเปล่งเสียงที่ต่างกันแม้จะผลิตจากบุคคลเดียวกัน ข้อค้นพบของงานวิจัยนี้เป็นไปในทิศทางเดียวกันกับข้อค้นพบว่า  
สำเนียงภาษาที่มีผลกระทบต่อการแปรของอัตราการเปล่งเสียง อย่างไรก็ตามงานวิจัยนี้มีข้อค้นพบที่ขัดแย้งกับข้อค้นพบว่าอัตรา  
การเปล่งเสียงมีการแปรภายในบุคคลที่ต่ำในวรรณกรรมที่ได้พบทวนไป



สาขาวิชา           ภาษาศาสตร์  
ปีการศึกษา       2565

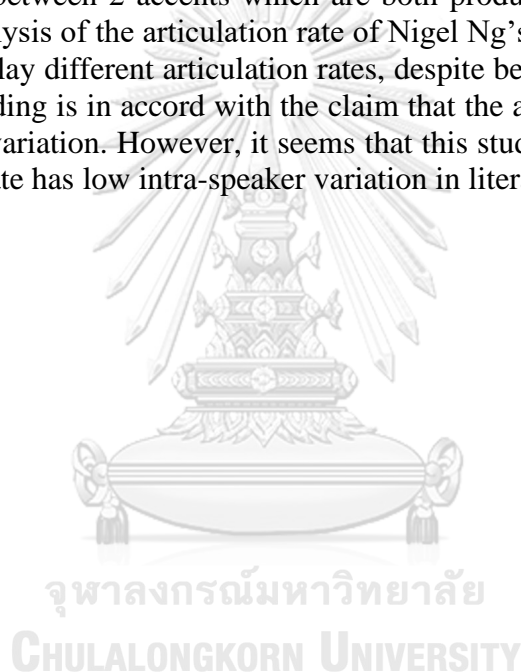
ลายมือชื่อนิสิต .....  
ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

# # 6380007322 : MAJOR LINGUISTICS

KEYWORD articulation rate, English, Chinese accent, inter-accent variation  
D:

Komchit Taweasablamert : Comparison of the inter-accent variation of articulation rate between English and English with imitated Chinese accent speech. Advisor: Asst. Prof. SUJINAT JITWIRIYANONT, Ph.D.

This paper compares the articulation rate of 2 English speeches with different accents produced by Nigel Ng, a well-known stand-up comedian and youtuber, while he is taking the role of Uncle Roger and while he is not. Taking the role of Uncle Roger, Nigel Ng produces utterances with imitated Chinese accent to underline stereotypical Asian characteristics, and to contrast with his own “neutralish English accent”. The objective is to investigate the variation of articulation rate between 2 accents which are both produced by only one speaker. Based on the analysis of the articulation rate of Nigel Ng’s speeches, it is found that both accents display different articulation rates, despite being produced by the same speaker. This finding is in accord with the claim that the accent has an effect on the articulation rate variation. However, it seems that this study contrasts the claim that the articulation rate has low intra-speaker variation in literatures.



Field of Study: Linguistics

Student's Signature

Academic Year: 2022

Advisor's Signature

Year:

.....

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deep reconnaissance to my advisor Asst. Prof. Sujinat Jitwiriyanont, Ph.D., for his continuous support of my research, for his patience, motivation and guidance. Without his support, the research cannot be conducted to the right direction and to such an established conclusion.

My gratitude is as well extended to the committees: Assoc. Prof. Pittayawat Pittayaporn, Ph. D. and Asst. Prof. Pavadee Saisuwan, Ph. D., for their support and their inspiring questions that lead this research to its more refined state. My sincere gratitude must also be extended to all the professors and staffs at the department of Linguistics, Faculty of Arts, Chulalongkorn University, for their immense knowledge that inspired and developed my research, and for their support all along my academic years.

I am also thankful to all fellow Asst. Prof. Sujinat's advisees for their support all along this meaningful semester. Most importantly, my greatest thank must be delivered toward Rattanasuwan for his suggestion on various statistical tests which provided me a vast knowledge about statistics. My thank must also be extended to all my friends at the faculty and all my friends who are not at the faculty for all their sentimental and emotional support especially during the hard time in this graduate's journey.

My thanks must as well be expressed to Nigel Ng whose videos were chosen to be the data for this experiment. My gratitude is expressed toward him for his videos that inspired me when I had to come up with new research topic, for his comedic videos that brought me joy during the stressful data analysis period and for his speech as a mean to deepen my linguistic knowledge that I could never achieve without this research.

Komchit Taweasablamlert

## TABLE OF CONTENTS

	<b>Page</b>
ABSTRACT (THAI) .....	iii
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
1. Introduction.....	7
2. Background.....	7
2.1. <i>Articulation rate</i> .....	7
2.2. <i>Variation on articulation rate</i> .....	9
2.3. <i>Chinese accent English</i> .....	11
2.4. <i>“Uncle Roger”</i> .....	12
2.5. <i>Research Question, Hypothesis and Prediction</i> .....	14
3. Methodology.....	14
3.1. <i>Articulation rate calculation</i> .....	14
3.2. <i>Data collection</i> .....	15
3.3. <i>Statistical test</i> .....	17
4. Result .....	17
4.1. <i>Accent type</i> .....	17
4.2. <i>Interlocutor condition</i> .....	19
5. Discussion.....	20
6. Conclusion .....	24
REFERENCES .....	25
VITA.....	29

## 1. Introduction

Going through the internet, we can see various Asian origin comedians overtly expressing themselves in Asian accents. Among them, Nigel Ng is one of the most well-known comedians and youtubers. Taking the role of Uncle Roger, Nigel Ng produces utterances with imitated Chinese accent. This particular accent occurs to underline the Asian characteristic and to contrast with his “neutralish English accent”. This is an opportunity to investigate the articulation rate variation between the accents that are produced by the same speaker. Literatures show that the articulation rate has low intra-speaker variation. Yet, they also show variations between various factors such as language proficiency or accent. By examining the articulation rate of both the imitated Chinese accent and the normal accent of Nigel Ng, we can focus on investigating the inter-accent variation of articulation rate, and eliminate the effect of the inter-speaker variation. In order to observe the inter-accent variation, the speeches are collected from videos published on Nigel Ng’s YouTube channels. Those videos are extracted into short clips representing the 2 accents. The articulation rates are locally calculated and statistically compared.

## 2. Background

### 2.1. Articulation rate

In forensic phonetic, one of the most challenging tasks is speaker discrimination task. It is to compare the voice in the recording to the voice of the suspect and to verify whether the suspect is the same person as the unknown speaker. In practice, many acoustic cues are less effective due to the quality of the recording. Yet, duration is one of the least affected parameters. There are 2 duration-based cues that are often used in speaker discrimination task: *speaking rate* and *articulation rate*. Speaking rate (SR) is a ratio between number of linguistic units in utterance and the duration of whole utterance including pauses. Articulation rate (AR), on the other hand, is a ratio between number of units in utterance and the duration of the utterance without pauses (Chen & Robb, 2004; Gold, 2018; Jessen, 2007; Trouvain, 2003).



To calculate the articulation rate, researchers must make 3 methodological decisions (Jessen, 2007). First, the kind of linguistic unit that will be counted must be decided. The frequently used units are the segment, the syllable and the word, which return the term of segment per second, syllable per second and word per second, respectively. And, the most commonly used unit in several study is the syllable.

Second, the method of counting the linguistic units must also be decided. There are 2 ways to count the units: by their intended form or by their realized form. The commonly used choice is to count the linguistic units by the realized form.

Third, the size and also the kind of the speech interval are to be decided. In case of speech rate, the size of interval is the whole recording. Articulation rate, however, needs to define the size of the interval and how to define the interval.

To choose the kind of the interval, we must consider the type of articulation rate that we need to analyze. There are 2 type of articulation rates in literature: *global articulation rate* and *local articulation rate*. Global articulation rate takes the duration of the whole recording excluding every pauses. Thus, there is only one value of articulation rate for each recording. Yet, speakers tend to change the speed during their speech. So, local articulation rate is introduced to solve the problem. Calculating local articulation rate, researchers need to split the recording into multiple short intervals. Then, there are multiple values of articulation rate for each recording, and global articulation can be derived by making an average of local articulation rates in each recording (Jessen, 2007; Trouvain, 2003).

In order to calculate local articulation rate, the interval is also to be defined. Trouvain (2003) states 2 ways to define an interval: interpause stretch – an interval of fluent speech surrounded by pauses – and intonation phrase – an utterance produced within one cycle of exhalation surrounded by pauses. Jessen (2007) later examines Trouvain's approach. Jessen shows that interpause stretch is the most objective, given that it needs to set the threshold for pauses. But it may be flawed by the variation of tempo between person. Otherwise, intonation phrase can give the most uniform duration, but the linguistic or phonetic criteria may be varied among different researchers. Due to the flaws in Trouvain's approach, Jessen then proposes a novel

method called “memory stretch”. The length of this stretch is defined by the number of linguistic units in a fluent speech that can be held in short-term memory. He claims that this method returns simpler and more pragmatic articulation rate. However, this method also needs extra experimental task on defining the number of units that can be easily held in short-term memory, which may also be varied between language.

In this study, the local articulation rates from the videos of Nigle Ng are analyzed. The local articulation rates are defined in term of syllable per second. The linguistic units are counted by their actual pronunciation. This is to prevent the different realization that may occur between accent or occur according to the speaker’s intention (e.g., ‘meme’ is intended to be pronounce as /mim/, but in one of his videos, he deliberately says /mimi/ to make the video comical). The intervals are delimited by interpause approach. As the experiment is conducted on speeches produced by one person, speaker variation on pause length is less concerned. The pause-threshold is set at 250 msec or whenever an audible inhalation is present. Moreover, every interval must contain at least 4 syllables. Only fluent speeches are collected.

## *2.2 Variation on articulation rate*

Researchers have examined the variation of articulation rate. Studies shows that articulation rate has lower intra-speaker variation than inter-speaker variation in German (Künzel, 1997 in Gold, 2018; Jessen, 2007) and Chinese (Cao & Wang, 2011). This may prove that articulation rate has better performance in speaker discrimination task. Jessen (2007) points out that the discrimination power of articulation arises from the fact that the articulation rate reflects particular habits of speakers, which speakers do not manage to conceal. On the other hand, various studies found that articulation rate is related to brain’s motor function, such as the drop of articulation rate under the influence of alcohol or cerebellar activities and motor cortex activities (*see* Jessen, 2007). Jessen argues that these anatomical and physiological foundations only reflect the habit and intention of speakers and they can only explain certain intra-speaker variation of articulation rate. He also raises that these evidences do not show that anatomical foundation could determine the

articulation rate of different persons as direct as, given by the author, how the vocal tract length can determine the formant frequencies of different persons. Also, Gold (2018) reexamines the discrimination power of articulation rate by mean of Log Likelihood Ratios Model. Though the model predicts better result on same speaker comparison than different speaker comparison, the within person variability is larger than between person variability. Gold (2018) suggests that the model might “allocate higher degrees of similarity if two speakers have similar degrees of (high) within speaker variation” and claims that the articulation rate might not have that much power in speaker discrimination task.

Apart from speaker variation, the relation between articulation rate and social aspects along with geographical contexts is examined. Studies prove that both social and geographical contexts do have an effect on the articulation rate.

Ordin & Polyanskaya (2015) investigate the speech rhythm of L2 learners. They record 51 German learners and 48 French learners of English. The result reveals that English learners speak slower than the native speakers. It also reveals that the speech rhythm differs in function of the language proficiency. The speakers with higher L2 proficiency speak faster than the speakers with lower L2 proficiency. Although Ordin and Polyanskaya (2015) did not examine directly the articulation rate, the result can infer that L2 learners with lower proficiency speaks with lower articulation rate (speak slower) than L2 learners with higher proficiency and the native speakers.

Also, the native speakers from different regions have varied articulation rate. Avanzi et al. (2012) compare the articulation rate of 4 varieties of French: Parisian French, Swiss French spoken in Neuchâtel, French spoken by Swiss German speaker in Bern and French spoken by Swiss German speaker in Zurich. 16 elderly speakers are recorded (2 males and 2 females each variety). The result reveals that Swiss speakers show lower articulation rate than Parisian speakers. The similar result also found in American English. Jacewicz et al. (2009) examined the articulation rate of 47 speakers who live in Wisconsin (Northern America) and 47 speakers who live in North Carolina (Southern America). They found that Southern speakers articulate faster than the Northern speaker. Coats (2020) also found the different articulation

rate from different regions of the United States. Analyzing data from YouTube channels of state and local government or civic organization, Coats is able to reveal that the speakers in American South speak slower than speakers in American North.

Additionally, articulation rate can be affected by pragmatic aspect. Cao & Wang (2011) also comment that even the intra-speaker variation of articulation rate is relatively small, compared to inter-speaker variation, the intra-speaker variation may be larger if styles of samples are different. Both Jacewicz et al. (2009) and Avanzi et al. (2012) found that reading style shows lower articulation rate than conversation style. As the analysis is run on YouTube videos, which are mostly spontaneous and conversation style, the condition of interlocutor is to be concerned. Scarborough et al. (2007) test effect of interlocutor conditions: real or imagined interlocutor and native or foreign interlocutor. They asked 10 native American speakers to produce utterances in 4 conditions. First, they were asked to give a direction to real native speaker of English who sat face-to-face with the participants. Second, they were asked to give a direction to real foreigner. Third, they were asked to give a direction to an imagined native speaker of English. Lastly, they were asked to give a direction to an imagined foreigner. The result reveals the effect of both the language background of interlocutor and the authenticity of interlocutor. Speakers tend to produce utterances with larger vowel space, longer duration and lower rate of speech (it is not stated whether this rate was calculated with pauses (speech rate) or without pauses (articulation rate)), when they speak to foreigner and to an imagined interlocutor. Scarborough et al. suggest that the real interaction between speakers speed up the tempo of speech. Taking the anatomical view of speech tempo, it is possible that imagining an interlocutor takes a functional load and slows down motor function, thus, the slower articulation rate.

### *2.3. Chinese accent English*

Chinese accent English is an accent of English produced by the Chinese native speakers who learn English as a Second Language. This accent is occurred due to the negative transfer from Chinese to English. The primary differences that could affect the tempo of English are timing and syllable structure. Chinese is syllable-timed language and monosyllabic language with predominant consonant + vowel structure,

whereas English is stress-timed language and polysyllabic language with more diverse syllable structure (Chen & Robb, 2004).

In effort to examine the characteristics of English that is produced by Chinese native speaker, Ding et al. (2021) observe English utterances that are produced by 20 speakers from each of 3 groups: namely, Chinese native speakers with high English proficiency, Chinese native speakers with low English proficiency and English native. They found that Chinese accent English has different F0 pattern from native English. English produced by Chinese speakers display F0 peaks that are more related to lexical items than native English speakers whose F0 pattern is more intonation-stress related. Chen & Robb (2004) investigated further the tempo of English produced by Chinese speakers. Comparing the utterances of 40 native Mandarin speakers producing American English to the utterances of native American English. The researchers found that English produce by Chinese speaker displays different articulation rate from English produced by native speaker. Chinese speakers tend to speak slower, thus, the lower articulation rate than native speaker.

We have observed that articulation rate can be varied by several factors such as accent or interlocutor condition. However, studies of articulation rate variation that depends on social or geographical factors are performed with multiple representative groups of participants. It is hard to deny that inter-speaker variation may also have an effect on the other types of variation such as inter-accent variation. Then, we decide to focus our study of inter-accent variation on only one person, to investigate the effect of accent on the articulation rate without the influence of inter-speaker variation.

#### 2.4. *“Uncle Roger”*

To observe the effect of accent on articulation rate of the same speaker, we choose the investigate the articulation rate of Nigel Ng’s speeches. Nigel is a Malaysian stand-up comedian based in London, England, and youtuber. He was born in Cantonese and English-speaking family in Kuala Lumpur, Malaysia. It is not clear whether he speaks Cantonese with family but he mostly speaks English in professional life. In one of his interviews, he claims that he speaks “neutralish English accent” and that he speaks “quite fast [...] for Asian standards” (LADbible TV,

2022). By “neutralish English accent”, it is hard to determine whether his accent is neutral toward General American English or British Received Pronunciation. This is because he has been both in the United States during his bachelor years and in the United Kingdom at present. However, as we observe the speeches of Nigel Ng, his accent shows both General American accent, with evidences such as rhotic ‘r’ at the end of word ‘water’ [‘wɑ.r̥ə] and flapped ‘t’ in the middle of word ‘little’ [‘li.r̥əl], and Received Pronunciation, with evidence such as the back [ɑ] in word ‘podcast’ [‘pɒd.kɑst].

In social media, Nigel Ng is largely known for his persona “*Uncle Roger*”. This persona portrays an image of a stereotypical middle-aged Asian “uncles” who, according to Nigel Ng, are “a little bit sassy” and “a bit know-it-all” but “kind” and who “stuck in their way” (LADbible TV, 2022). In linguistic view, the image of an Asian uncle is underlined by pronounced Cantonese accent or *Chinglish* which are widely use in Malaysia (Brenda, 2022; Li, 2020). This *Chinglish* is reflected in various cues: lack of auxiliary verb or linking verb, lack of plural suffix ‘s’, lack of third person pronoun or the catchy word “haiya” which is a Cantonese interjection used by Malaysian Chinese speakers (Li, 2020). The *Uncle Roger* persona also reflects some phonological features which are matched with the phonological feature of Chinese accent or *Mock Asian* according to (Chun, 2022): reduplication of word and alveolarization of voiceless interdental fricative [θ] to [s]. But in case of *Uncle Roger*, this alveolarization occurs to both voiced and voiceless along with de-frication ([ð] → [d] and [θ] → [t]). However, the tempo cue is not found in literatures of Asian stylization.

As *Uncle Roger* is a character of Nigel Ng, the effect of accent on the articulation rate variation in the same person can be investigated by comparing the articulation rate of two personas’ speeches. Nigel Ng’s speech represents his neutral English accent (further mentioned as “normal English accent”) and *Uncle Roger*’s speech represents his pronounced Cantonese English accent (further mentioned as “imitated Chinese accent”).

### 2.5. Research Question, Hypothesis and Prediction

This study aims to investigate the effect of accent on the articulation rate of one person. Thus, the main question is: does articulation rate of normal English accent speech differ from that of imitated Chinese accent English speech which are both produced by the same speaker?

Our first hypothesis ( $H_0$ ) is that the accent has an effect on the articulation rate even though the speeches are produced by the same person. With reference to the previous findings, we predict that the imitated Chinese accent English displays significantly lower articulation rate than the normal English speech.

We also speculate an alternative hypothesis ( $H_1$ ). We hypothesize that the intra-speaker stability of articulation rate has enough power to overcome the effect of accent. To support this hypothesis, we expect that the imitated Chinese accent speech should not display significantly lower articulation rate than the normal English speech.

These hypotheses are tested with the speeches extracted from YouTube videos. The methodology of the experiment is to be described in *section 3*. Then, the result of the experiment is reported in *section 4* and later discussed in *section 5*. Finally, the *section 6* will conclude our experiment and the main finding of our present study.

## 3. Methodology

### 3.1. Articulation rate calculation

To calculate the articulation rate, the three methodological decisions are made according to Jessen (2007).

First, linguistic units are counted by syllable. Therefore, the articulation rate in this study is defined by syllable per second. For each syllabic nucleus, a unit is counted. In case of word with multiple vowels clustering together, diphthong and triphthong that exist in English inventory are counted as one unit, otherwise, the unit is counted segment-wise. Second, the units will be counted by the actual

pronunciation. This is to prevent the difference of realization that rises from accent or from the speaker intention. Third, local articulation rates are calculated and each interval are delimited by interpause-stretch. The pause-threshold is set at 250 msec or whenever an audible inhalation is present.

Moreover, every utterance must contain at least 4 syllables. This threshold is set in respect to Jessen (2007) to avoid the inclusion of very short intervals that could affect the overall articulation rate with the phrase-final lengthening. Only fluent speeches are collected. Utterances with stuttering and hesitation are filtered out.

### 3.2. Data collection

The utterances are extracted form Nigel Ng's YouTube channels: *mrnigelng* and *HAIYAA with Nigel Ng*. 10 videos are selected. These 10 videos are posted on YouTube form 10<sup>th</sup> December, 2021 to 23<sup>rd</sup> May, 2022. They are then extracted to 12 clips with length of 3 or 4 minutes according to the amount of his utterances in the clips. Among these 12 clips, 6 clips represent Uncle Roger's imitated Chinese accent and other 6 Nigel Ng's normal speech. Among 6 clips of each accent, 3 clips represent real interlocutor condition, given that the utterances are produced when at least one interlocutor present in the setting (all of utterances are counted, whether the speaker converses with an interlocutor or not). Other 3 represent imagined interlocutor condition given that the utterances are produced without the presence of interlocutor. It is safe to assume that the speaker might as well talk to the camera which we considered as an imagined interlocutor.

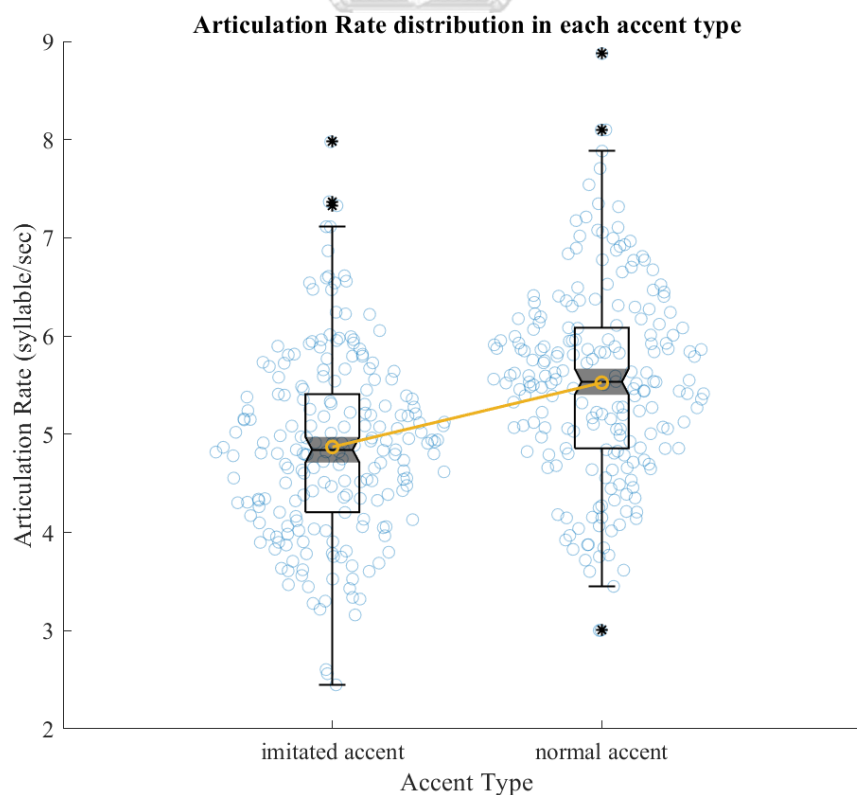
	Imitated Chinese	Normal English	Total
Imagined interlocutor	3	3	6
Real interlocutor	3	3	6
Total	6	6	12

*Table 1* Number of clips representing each type of speech

Each clip is segmented into intervals in Praat software. Syllables are counted by hand and a Praat script is written to extract the duration of each interval. Local articulation rates are calculated by dividing the number of syllables by the duration of interval with help of Microsoft Excel software. Therefore, 426 utterances are



collected: 210 imitated Chinese accent utterances against 216 normal English accent utterances and 215 utterances with interlocutor against 211 utterances without interlocutor. Each clip has the average of 35.5 utterances (range = [19, 51], std = 10.77). For all the data, the observed articulation rate has the average of 5.21 syllables/sec (range = [2.45, 8.88], std = 1.03). The imitated Chinese accent utterances have the average articulation rate of 4.87 syllable/sec (range = [2.45, 7.98], std = 0.93) and the normal English utterances have the average articulation rate of 5.54 syllable/sec (range = [3.00, 8.88], std = 1.00). The distribution of articulation rate is presented in violin plot (*Fig.1*). The box plot represents the first quartile of data (lower side of the box) and third quartile of data (upper side of the box). The horizontal line inside the box represents the median. The shaded area of the box is the notch representing the most likely expected values for median. The blue swarm scatter plot in the background represents the distribution of collected speech articulation rate. The data points horizontally spread out according to Kernel Density Estimation; the more horizontally spread out, the more data are clustered around a value of articulation rate.



*Fig. 1* Distribution of articulation rate in imitated Chinese accent and normal English accent

	Imitated Chinese	Normal English	Total
Without interlocutor	100	111	211
With interlocutor	110	105	215
Total	210	216	426

*Table 2* Number of intervals representing each type of speech

### 3.3. Statistical test

A statistical test is performed on the articulation rate of both accent of Nigel Ng. As both the imitated Chinese accent and normal English accent are produced by the same speaker, a statistical test for dependent sample is chosen. However, we were not able to find the suitable test for dependent sample with unequal sample size. The paired *t*-test is then chosen in this present study. Using such statistical test, we are obliged to balance the size of two groups. We randomly remove 6 intervals from the normal English data. We recheck that removing 6 random intervals does not affect the shape of the normal English data. The shape of balanced normal English data does not visibly differ from the former normal English data. Also, the descriptive statistic of the balanced data does not deviate much from the former data (average articulation rate = 5.53 syllable/sec, range = [3.00, 8.88], std = 0.99; former data see 3.2. *Data collection*)

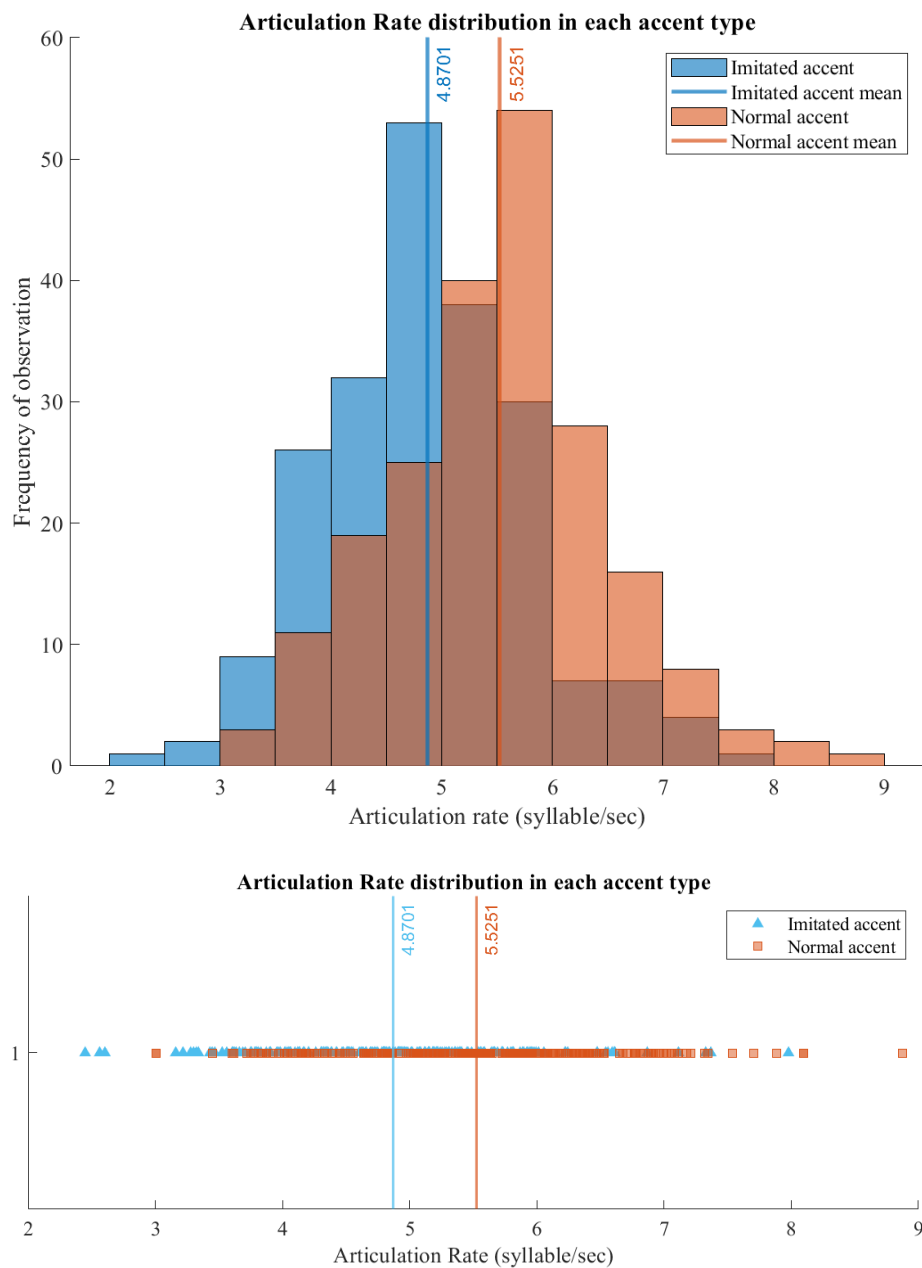
The paired *t*-test is performed with help of MATLAB software. Aiming to test whether the articulation rate of imitated Chinese accent English is lower than the articulation rate of normal English, we conduct a one tailed *t*-test. The following function is used to perform the statistical test: `ttest(imitated Chinese AR, normal English AR, "Tail", "left")`

## 4. Result

### 4.1. Accent type

A one tailed paired *t*-test is performed to compare the articulation rate in imitated Chinese accent speech and normal English speech. The test reveals that the imitated Chinese accent speech (mean = 4.87, std = 0.93) has a significantly lower articulation rate than the normal English accent (mean = 5.53, std = 0.99);  $t(209) = -7.47, p < 0.001$ .

The result reveals that accent has an effect on the articulation rate of Nigel Ng’s speech. The imitated Chinese accent speech shows significantly lower articulation rate than the normal English accent speech. It supports our first hypothesis ( $H_0$ ) that the accent has an effect on the articulation rate even though the speeches are produced by the same person. The result may also confirm the previous findings that English produced by Chinese speaker (Chinese accent) has lower than English produced by Native speaker (normal English accent).



*Fig. 2* The overlap of the articulation rate distribution in both accents

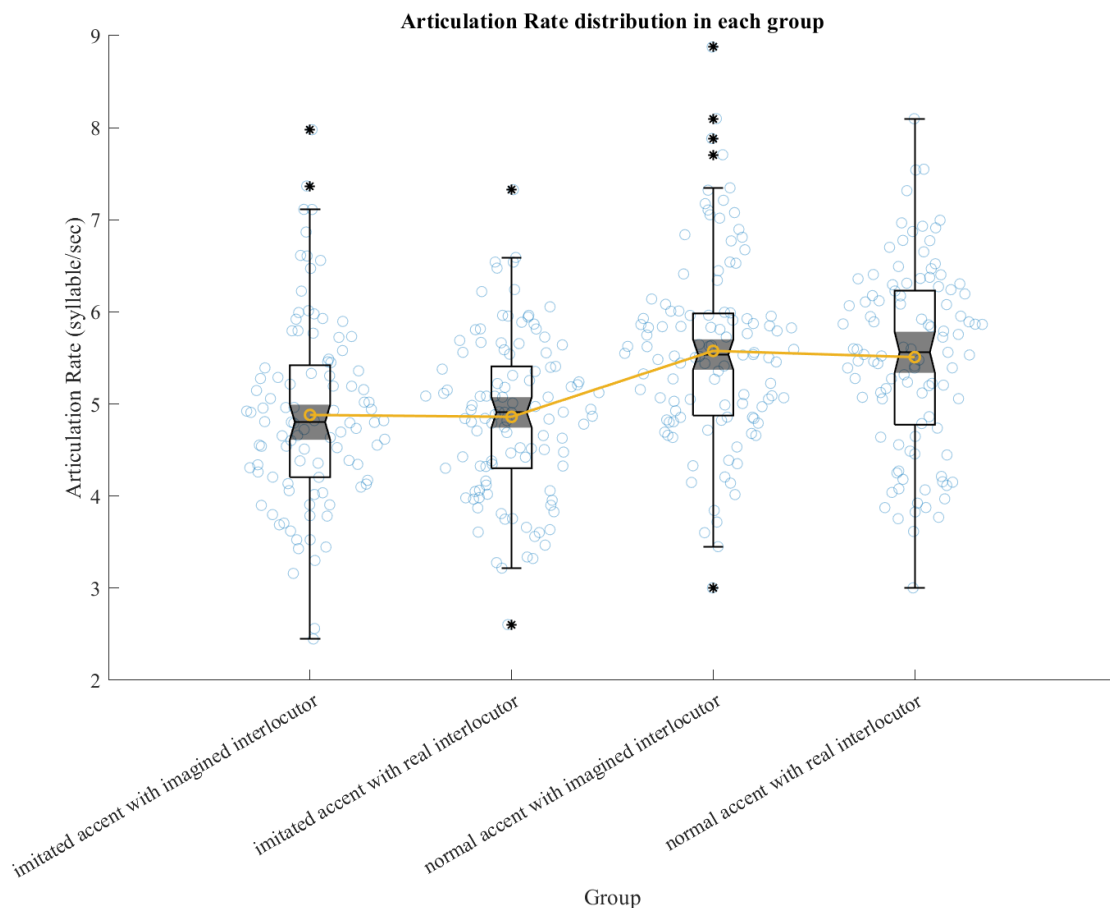
Although, the inter-accent variation of articulation rate is proved to have an effect even in the same person scenario, it is not clear that this finding contrasts the claim that articulation rate has low intra-speaker variation. Observing the distribution of articulation rate in both accents, we can see that there is a large overlapping area between the two accents (see *Fig. 2*). It seems highly counterintuitive that such a largely overlapping data shows a significant difference between two groups. This may arise from the choice of statistical test or the design of data collection; because our data violates at least one of tests' assumptions. We also tried to run a non-parametric test for dependent sample (i.e., Wilcoxon Signed-rank test) and it still returned the same result with  $p$ -value  $< 0.001$ . To improve the experiment, the more proper statistical test should be used to examine the difference of speeches that are produced by the same speaker.

#### 4.2. Interlocutor condition

The interlocutor condition is also controlled in the experiment. Unfortunately, without a proper statistical test, we cannot statistically test the effect of the interlocutor condition on the articulation rate. We neither use the same solution as we did to compare between the accent type. Doing so, too many data points would be randomly removed to construct the data with the size of 100 samples in each group; 10 samples from imitated accent with real interlocutor group, 11 samples from normal accent with imagined interlocutor group and 5 samples from normal accent with real interlocutor group must be randomly removed. However, as we visualize the distribution of articulation rate in function of accent and the interlocutor condition (*Fig. 3*), it shows that the means of articulation rate between interlocutor conditions barely differ in same accent: 4.86 syllables/sec with real interlocutor and 4.88 syllables/sec with imagined interlocutor in imitated Chinese accent, and, 5.50 syllables/sec with real interlocutor and 5.55 syllables/sec with imagined interlocutor in normal English accent.

Even the means between interlocutor conditions are almost the same, it is shown that, in both accents, the speaker articulates slightly faster with imagined interlocutor than with real interlocutor. This observation appears to contrast the

finding of Scarborough et al. (2007). On the other hand, the difference of articulation rate between interlocutor conditions is exactly the same pattern across the two accents. Regardless the interlocutor conditions, imitated Chinese accent speeches are more slowly articulated (have lower articulation rate) than normal English accent speeches.



*Fig. 3* Distribution of articulation rate in function of the accent and the interlocutor condition

## 5. Discussion

The accent is observed to have a significant effect on the articulation rate of Nigel Ng. As he produces speech in Uncle Roger's imitated Chinese accent, he articulates slower than his normal English accent. This finding is in accord with the literatures; English speech produced by Chinese speakers tends to have lower articulation rate than the speech produced by native English speaker. However, there are still some points that should be further discussed.

We have suggested that the result of the present study cannot strongly prove the inter-accent influence on the intra-speaker articulation rate variation due to the improper choice of statistical test or the data collection. To further investigate this influence, we suggest to run the experiment on multiple speakers. With multiple speakers who can speak more than one accent, it should be more effective to run a statistical test for dependent sample, such as paired *t*-test that is currently used in this study. Moreover, it may reveal the pattern of variation that is happened across the speakers. It could be possible to find whether the distribution of two accents is more separated or completely overlap. Then, we could precisely prove the effect of either the inter-accent variation or intra-speaker variation in the case of speakers with more than one accent.

Another potential confound in this experiment is the type of video. The type of video is not controlled seeing that some types are the least published by the youtuber and some types are mainly produced with either character of Uncle Roger or Nigel Ng. Among the 12 clips, 8 clips are podcast type, 3 clips are reaction type and 1 clip is vlog type. Reaction videos and vlog video are filmed in imitated Chinese accent. Podcasts videos are predominantly filmed in normal English accent. The type of video

CLIP	ACCENT	TYPE
CLIP1	Imitated Chinese	reaction
CLIP2	Imitated Chinese	reaction
CLIP3	Imitated Chinese	reaction
CLIP4	Imitated Chinese	vlog
CLIP5	Imitated Chinese	podcast
CLIP6	Imitated Chinese	podcast
CLIP7	Normal English	podcast
CLIP8	Normal English	podcast
CLIP9	Normal English	podcast
CLIP10	Normal English	podcast
CLIP11	Normal English	podcast
CLIP12	Normal English	podcast

*Table 3* Distribution of Type of video

can be considered as a confound because it is highly related to speech style; vlog is conversational, reaction is a mixture of conversation style with or without interlocutor

and, sometimes, reading style, and podcast is mostly conversation style with or without interlocutor.

Regarding interlocutor condition, the result contrasts the finding of Scarborough et al. (2007). They found that speakers tend to articulate faster when they speak to the real interlocutor comparing to the imagined interlocutor. But, the criteria of with or without another person in the setting might not able to capture the effect of authenticity of interlocutor as Scarborough et al. (2007) did. First, the researchers asked participants to imagine an interlocutor without any point of focus. The recordings however show that Nigel Ng regularly interacts with the camera which can be a good point of focus for the speaker (see *Pic.1*). Second, reaction videos, which are imagined interlocutor condition, could be considered as more prepared than podcast videos, which are mostly real interlocutor condition. We can observe in outtakes added at the end of the reaction videos that even the videos are not scripted, Nigel Ng sometimes speaks the same sentences several times to make the utterance smoother. But podcast videos seem to be more integral with less cutting. This preparedness might be a confound that return the different observation of interlocutor condition in the present study from the finding of Scarborough et al. (2007).



*Pic. 1* Interaction with camera (from CLIP1)

Another matter to be concerned is that there are some extremely long intervals. For example, the longest interval has the length of 11.97 second and contains 51 syllables. This absurdly long interval occurs due to the cutting of video.

Such interval may be made up from different takes of the same sentence or from several cuts of multiple phrases. This cutting is to make a cohesive utterance. However, these long intervals may flaw our interpretation of data. They may be made up from different takes, in which the creator possibly speaks with subtly different articulation rate. If not, the creator may drop out some words or some part with hesitation or stuttering, which should be filtered out from the data. Even the articulation rates of these extremely long intervals are clustered around means, it would be best for studies in the future to concern about this matter. In defining the size of interval for local articulation rate, not only audio cue such as pause or audible inhalation, but the visual cue should also be concerned, for the fact that cutting in video is more easily perceived by watching than listening to the video.

Furthermore, it must be noted that the speech of Uncle Roger should not be considered as naturally produced, thus, the name of “imitated Chinese Accent”. This accent could be considered as a stylization that projects to the stereotypical Asian accent. The Asian stylization in *Uncle Roger*'s performance can be consider as a racializing style which explicitly marks the racial difference from mainstream language. The purpose of racializing stylization is to construct the ‘self’ and ‘other’ in sociolinguistic context (Tsiplakou & Ioannidou, 2012). The *Uncle Roger* style is not to mocking the absent participant in order to alienate the others as mentioned in Mason Carris (2011) and Tsiplakou & Ioannidou (2012) but rather to validate the speaker's community membership (Chun, 2022). As we observed the articulation rate, the Asian validation is reflected in speech tempo of Nigel Ng. He accentuates the *Chinglish* as much that the overall articulation rate of *Uncle Roger* is reduced in the same pattern as the English produced by Chinese speakers.

Finally, as the present study is conducted on only one person's speech, it may also reflect the intra-speaker variation of the articulation rate. Regardless the large overlapping area between two accents, the articulation rates are uniformly clustered around the mean of each accent. It appears that self-validating stylization may influence the articulation rate in the same speaker. Yet, the interlocutor condition seems to not has the influence on the intra-speaker variation. It is safe to assume that intra-speaker articulation rate variation is influenced by various factors at different



power. The speech style may have low effect on the intra-speaker variation and the racializing stylization such as imitated accent may have high effect on the intra-speaker variation of articulation rate.

## 6. Conclusion

The articulation rate of Nigel Ng, a Malaysian stand-up comedian and youtuber based in the United Kingdom, is investigated. The speaker produces utterances in 2 different accents: imitated Chinese accent and normal English accent. It is statistically showed that his imitated Chinese accent displays lower articulation rate than his normal English accent. The difference between two accents is in accord with the previous finding on the effect of accent on the articulation rate. Noted that the articulation rate between accents is largely overlapped, our finding may only be an observation that supports the idea of an influence of accent on the articulation rate in same person scenario. The intra-speaker variation, which is considered low, can be affected by various factors at different power. The influence of interlocutor condition may have low influence on the intra-speaker variation whereas the accent stylization may have higher influence on the intra-speaker variation. As we investigated the accents, it seems that, in order to project himself to an Asian identity, the speaker may be able to manipulate his speech tempo to contrast *Uncle Roger* persona from his self as Nigel Ng.

## REFERENCES



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

## REFERENCES

- Avanzi, M., Dubosson, P., & Schwab, S. (2012). *Effects of Dialectal Origin on Articulation Rate in French*.
- Brenda. (2022, October 27). *Nigel Ng Bio, Age, Net Worth, Wife, Partner, Family, Siblings, Heigh*. BiographyScoop. <https://biographyscoop.com/nigel-ng/>
- Cao, H., & Wang, Y. (2011). A forensic aspect of articulation rate variation in Chinese. *Proceedings of ICPhS*.
- Chen, Y., & Robb, M. P. (2004). A study of speaking rate in Mandarin speakers of Anterican English. *Asia Pacific Journal of Speech, Language and Hearing*, 9(3), 173–188. <https://doi.org/10.1179/136132804805575813>
- Chun, E. W. (2022). Ideologies of legitimate mockery: Margaret Cho’s revoicings of mock Asian. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 263–289. <https://doi.org/10.1075/prag.14.2-3.10chu>
- Coats, S. (2020). Articulation Rate in American English in a Corpus of YouTube Videos. *Language and Speech*, 63(4), 799–831. <https://doi.org/10.1177/0023830919894720>
- Ding, H., Lin, B., & Wang, L. (2021). *F0 Patterns of L2 English Speech by Mandarin Chinese Learners* (p. 3938). <https://doi.org/10.21437/Interspeech.2021-581>
- Gold, E. (2018). Articulation rate as a speaker discriminant in British English: Interspeech 2018. *Interspeech 2018*, 1828–1832. <https://doi.org/10.21437/Interspeech.2018-1384>
- Jacewicz, E., Fox, R. A., O’Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, 21(2), 233–256. <https://doi.org/10.1017/S0954394509990093>

- Jessen, M. (2007). Forensic reference data on articulation rate in German. *Science & Justice*, 47(2), 50–67. <https://doi.org/10.1016/j.scijus.2007.03.003>
- LADbible TV (Director). (2022). *How I Came Up With “Uncle Roger” Character | Minutes With | @LADbible*.  
<https://www.youtube.com/watch?v=oHKqcE59AnE>
- Li, P. (2020, November 5). Uncle Roger: A Stereotype for a Middle-aged Asian Man? *Educational Sociolinguistics*. <https://bild-lida.ca/educationalsociolinguistics/uncategorized/uncle-roger-a-stereotype-for-a-middle-aged-asian-man/>
- Mason Carris, L. (2011). La voz gringa: Latino stylization of linguistic (in)authenticity as social critique. *Discourse & Society*, 22(4), 474–490.  
<https://doi.org/10.1177/0957926510395835>
- Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, 138(2), 533–544.  
<https://doi.org/10.1121/1.4923359>
- Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *The Journal of the Acoustical Society of America*, 121(5), 3044–3044.  
<https://doi.org/10.1121/1.4781735>
- Trouvain, J. (2003). *Tempo variation in speech production. Implications for speech synthesis*.
- Tsiplakou, S., & Ioannidou, E. (2012). Stylizing stylization: The case of Aigia Fuxia. *Multilingua*, 31(2–3), 277–299. <https://doi.org/10.1515/multi-2012-0013>



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

## VITA

**NAME** Komchit Taweasablamlert

**DATE OF BIRTH** 11 September 1998

**PLACE OF BIRTH** Pisanuloke, Thailand

**INSTITUTIONS ATTENDED** Faculty of Arts, Chulalongkorn University (B.A., Honors Program)

**HOME ADDRESS** 64/1 M.4 Huaykapi Meuang Chonburi 20130



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY