國立臺灣大學文學院語言學研究所

碩士論文

Graduate Institute of Linguistics

College of Liberal Art

National Taiwan University

Master's Thesis

臺灣閩南語元音間 /t/ 變異音之聲學及構音研究

Acoustic and articulatory variations of intervocalic /t/ in Taiwan Southern Min

李旻宰

Minjae Lee

指導教授：邱振豪 博士

Advisor: Chenhao Chiu Ph.D.

中華民國 114 年 2 月

February, 2025

doi:10.6342/NTU202500546

Acknowledgements

衷心感謝恩師邱振豪教授的指導與鼓勵，

教授的細心與耐心幫助我順利完成碩士學業。

並衷心感謝在這段期間所有照顧我的教授和同學，

以及協助實驗的參與者。

내가 사랑하는 모든 이에게 감사의 말씀을 전합니다.

doi:10.6342/NTU202500546

摘要

許多語言蘊含有元音間輔音之交替現象。在臺灣閩南語的元音間 /t/ 之處理過程中也會出現元音間輔音的交替現象。在詞內，元音間的 /t/ 可能會保持原本的阻音 [t]，也有可能會變成閃音 [ɾ]，這種變化可能由於構詞規則的差異所致。本研究透過聲學與構音分析，探討相同構詞結構是否會引起相似的閃音化現象。本研究比較分析了三個元音間/t/的類型 ：詞內阻音 (A 類)、詞內閃音(B 類) 以及詞界間閃音(C 類)。

本研究招募了八位臺灣閩南語使用者進行實驗。實驗過程中，蒐集三個不同類型的發音及超音波舌位影像，並分析其聲學及構音上的特徵。聲學分析結果顯示，阻音之形態呈現典型的 [t] 或其同位異音 [ʔ]，而閃音的形態則呈現 [ɾ]的特徵。在三個類型中，閃音的形態具有相似的時長及強度，而阻音的聲學特徵較為多元。另外，考量閃音較短的時長，本研究採用超音波 M-Mode 進行構音分析，並使用 CoAtNet-2 提取超音波影像之特徵向量以進行超音波舌位影像的量化分析，進而比較其餘弦相似度。該分析結果顯示 [t] 及其同位異音各有不同構音特徵，形態相似語境下的閃音 (B 類及 C 類) 之間相似度高於不同環境下的 [t] 或 [ʔ] 的相似度。

這些發現證實了閃音化現象可能受到構詞條件的影響。研究結果也進一步提供了電腦視覺模型在分析超音波舌位影像時的實用性。

關鍵詞： 臺灣閩南語、元音間/t/、閃音、構詞條件、超音波舌位影像

doi:10.6342/NTU202500546

Abstract

Intervocalic consonant alternation is a characteristic feature of many languages, and Taiwan Southern Min demonstrates this phenomenon, particularly in its treatment of /t/. Within words, the intervocalic /t/ either maintains its original posture [t] or becomes a flap [ɾ], potentially due to morphological differences. This study investigates whether morphological structures trigger comparable flapping using acoustic and articulatory analyses. Three structures were analysed: non-flapping intervocalic /t/ within words (category A), flapping within words (category B), and flapping at word boundaries (category C).

Eight Taiwan Southern Min speakers were recruited for a production experiment, where they pronounced words and phrases from each category. Data collection included both acoustics and ultrasound tongue images. Acoustic analysis results showed that morphological structures forbidding flaps displayed the typical plosive [t] or its allophone [ʔ], while those allowing flaps exhibited characteristic [ɾ]. Consistent duration and intensity were observed for flaps whereas more variations were reported for the non-flapping category. On the other hand, ultrasound M-mode imaging was employed for articulatory analysis considering the brief nature of flaps. CoAtNet-2's pre-trained model was utilized to extract feature vectors and compare cosine similarity to address quantification challenges in ultrasound tongue imaging. Results showed characteristic patterns for [t] and its allophones, with flaps in morphologically identical contexts (categories B and C) showing higher similarity to each other than to [t] or [ʔ] in different contexts.

These findings confirm that articulatory gestures of flapping may be morphologically conditioned. The results also highlight the utility of deep learning-based computer vision models for objectively analyzing ultrasound tongue imaging.

Keywords: Taiwan Southern Min, intervocalic /t/, flap, morphological conditioning, ultrasound tongue images

doi:10.6342/NTU202500546

Contents

doi:10.6342/NTU202500546

doi:10.6342/NTU202500546

doi:10.6342/NTU202500546

doi:10.6342/NTU202500546

List of Figures

doi:10.6342/NTU202500546

doi:10.6342/NTU202500546

List of Tables

doi:10.6342/NTU202500546

Chapter 1 Introduction

This study examines the intervocalic consonant alternation in Taiwan Southern Min (hereafter TSM), a dialect of Southern Min spoken in Taiwan. The intervocalic consonant alternation is a significant linguistic phenomenon observed across multiple languages, with diverse manifestations in each language (Lavoie, 1996; Kirchner, 1998). One of the most prominent examples in American English is flapping (Kahn, 1976; Zue & Laferriere, 1979), where voiced and voiceless alveolar stops /t/ and /d/ are articulated as an alveolar tap [ɾ] in specific environments. This phenomenon has been systematically examined across various fields, including phonology (e.g., Kahn, 1976; Kiparsky, 1979), acoustics (e.g., Zue & Laferriere, 1979; Warner et al., 2009; Hwang & Kang, 2018), and articulatory linguistics (e.g., Price, 1981; Fukaya & Byrd, 2005; Derrick & Gick, 2011).

Previous research on the intervocalic /t/ alternation in TSM has primarily focused on phonological conditions, specifically the alternations of intervocalic /p/, /t/, and /k/. It explores these alternations alongside other stop consonant changes (Chiang, 1992; Wang & Liu, 2010), while acoustic and articulatory analyses have remained relatively underexplored. Although this systematic approach to phonological rules has significantly enhanced our understanding of TSM's overall phonological system, it has limitations in observing and analyzing the detailed patterns of individual phoneme variations, which necessitates further acoustic and articulatory study.

This study employs acoustic and articulatory analyses to explore the intervocalic /t/ in TSM. It evaluates the patterns of alternation in various morphological contexts, focusing on both intra-word and inter-word boundaries. The aim is to gain a thorough understanding of this prevalent yet underexplored phonological aspect in TSM.

doi:10.6342/NTU202500546

1.1 Background

The intervocalic /t/ alternation in TSM is common among native speakers; however, it has not been thoroughly examined from acoustic and articulatory perspectives. Previous studies often grouped it with other stops (Chiang, 1992; Wang & Liu, 2010), resulting in less detailed analysis specifically on /t/ alternations. Interestingly, research on English has shown that intervocalic /t/ can exhibit various forms, including not only the standard [t] but also flaps [ɾ] (e.g., Zue & Laferriere, 1979; De Jong, 1998), glottal stops [ʔ] (e.g., Roach, 1973; Huffman, 2005), and more. These findings raise questions about whether similar variability also exists in TSM and suggest the need for in-depth studies to identify and explore such patterns.

Flapping, the most noticeable phenomenon of intervocalic consonant alternation is well-documented in American English. The intervocalic /t/ flapping in American English has been examined from various perspectives. Regarding its occurrence, Kahn (1976) emphasized the essential role of stress, noting that alveolar stops /t/ and /d/ primarily occur between a stressed and an unstressed vowel, as well as when surrounded by unstressed vowels. Additionally, flapping can happen both within words and at word boundaries. Cho (2004) reported that while the influence of stress diminishes at word boundaries, other factors such as phonetic elements, semantic components, speech rate, and syllable/word position significantly influence the application of flapping rules. Patterson and Connine (2001) explored the relationship between word frequency and flapping, finding that words with a higher frequency of occurrence show a greater tendency toward flapping.

Current research often uses quantitative methods with reproducible experimental techniques, such as acoustic and articulatory analyses. From an acoustic viewpoint, Zue and Laferriere (1979) explored the acoustic differences between flapped /t/ and /d/,

doi:10.6342/NTU202500546

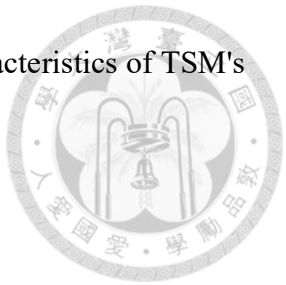
including variations in duration. Lavoie (2001) showed that flaps exhibit higher RMS amplitude measures than stops, based on research into various consonant and vowel characteristics in English and Spanish. Expanding upon these acoustic properties, Parker (2002) undertook a comprehensive investigation of sonority and concluded that when /t/ and /d/ weaken to become flaps, they gain greater intensity, resulting in increased sonority. Notably, the introduction of advanced acoustic analysis tools like Praat (Boersma & Weenink, 2005) has facilitated more detailed investigations of [ɾ], significantly enhancing the understanding of previously challenging aspects, such as variations in F0 and formants (Derrick & Schultz, 2013).

Additionally, ultrasound analysis of tongue movements proves to be very effective in revealing articulatory features, as it facilitates a precise examination of the tongue's position at specific instances, allowing for the analysis of variations across different phonetic environments (Stone, 2005). Furthermore, investigations of articulatory methods over time with M-mode ultrasound (Derrick & Gick, 2011) have shown significant benefits. For example, concerning the flap variation of intervocalic /t/, which is the primary focus of this study, the short duration of the sound makes it challenging to capture its articulatory features using B-mode ultrasound. However, M-mode allows monitoring tongue movement over time with precise boundary positions, highlighting its effectiveness utility.

The absence of quantitative analysis of TSM intervocalic /t/ alternation presents a significant opportunity. Given the extensive research on similar phenomena in English, applying these established techniques to TSM intervocalic /t/ could reveal unique alternation patterns that earlier systematic phonological studies have overlooked. The acoustic and articulatory analyses employed in English intervocalic /t/ alternation

doi:10.6342/NTU202500546

research can provide a solid basis for understanding the distinct characteristics of TSM's intervocalic /t/ alternations.



1.2 Research Objectives

This study employs analytical methodologies previously used in American English flapping research to examine how /t/ variants in intervocalic positions in TSM are realized according to morphological structure through acoustic and articulatory analyses. This research objectives are threefold:

1. Investigate the correlations between morphological context and flap occurrence in word-internal and cross-word boundary positions.

2. Analyse each variant's acoustic characteristics (duration, RMS amplitude).

3. Compare the postural characteristics of each variant through M-mode, which reflects temporal flow in ultrasound imaging.

The expected outcomes for each objective are as follows:

1. Morphological similarity is expected to induce similar flapping patterns within words and across word boundaries.

2. Variants in similar morphological contexts are expected to share analogous acoustic characteristics. When realized as [t], longer durations and lower RMS amplitudes are anticipated; shorter durations and higher RMS amplitudes are predicted when realized as a flap [ɾ].

3. Each variant is expected to exhibit characteristic postures, with similar morphological contexts displaying significant postural similarities.

doi:10.6342/NTU202500546

Additionally, this study will utilize 'CoAtNet-2' (Dai et al., 2021), a deep learning image processing model that combines the advantages of CNN and Transformers, to extract feature vectors from average posture images. Subsequently, cosine similarity will be applied to these feature vectors to quantitatively analyse the similarities between postures. This approach aims to provide a quantitative evaluation of posture similarities.

This research aims to gain new insights into TSM's intervocalic /t/ by utilizing various analytical methods previously employed in American English studies. It will facilitate the identification of TSM-specific characteristics.

1.3 Structure of the Dissertation

The dissertation is divided into five chapters, arranged as follows:

Chapter 2 reviews prior research on American English flapping, exploring the conditions that promote flapping and its characteristics through acoustic analyses and articulatory studies. It also surveys previous studies on alternating intervocalic /t/ TSM.

Chapter 3 presents the methodology, which includes the data collection, acoustic analysis with Praat, ultrasound imaging methods, and image analysis using the CoAtNet architecture.

Chapter 4 details the analytical findings, highlighting the effects of morphological context on flapping, the acoustic characteristics of variants, the articulatory patterns observed through ultrasound imaging, and the outcomes of similarity analysis.

Chapter 5 presents a comprehensive discussion of the findings and their implications. This chapter also addresses the current study's limitations and suggests potential directions for future research.

Chapter 6 wraps up the dissertation by summarizing the key findings and discussing their significance.

doi:10.6342/NTU202500546

Chapter 2 Literature Review

This chapter will examine cases that analyse intervocalic /t/ from various angles. First, it will overview phonological patterns and theories on flapping in English. Then, it will examine how the acoustic characteristics of flaps have been distinguished and review cases where flaps were studied from an articulatory perspective. The final part examines cases where intervocalic /t/ in TSM was investigated along with other stops.

2.1 Patterns of English Flapping

2.1.1. Flapping as a Phonological Rule

English flapping, a characteristic feature of American English, can be readily observed in words like *'butter'* and *'city'* (Kahn, 1976). This phenomenon occurs when /t/ is realized as an alveolar flap [ɾ], where the tongue tip quickly touches and releases from the alveolar ridge (Zue & Laferriere, 1979).

Kahn (1976) described flapping as a process influenced by syllable structure, occurring both within words (e.g., *butter*) and across word boundaries (e.g., *get it*). He established that flapping requires /t/ to be preceded by a non-consonantal sound followed by an unstressed vowel. By systematically describing how flapping interacts with syllable structure, Kahn emphasized the significance of syllables in English phonology.

Kiparsky (1979) introduced the larger prosodic unit of 'foot' to explain flapping. He attributed the weakened pronunciation of the second /t/ in *'potato'* to foot-internal prosodic structure formation. Kiparsky emphasized that English prosodic structure assignment occurs cyclically at each stage of word formation, demonstrating how prosodic structure assignment preserves and alters stress patterns. He modified Liberman and Prince's (1977) theory, which viewed stress and prosodic patterns as primarily governed by surface stress placement rules, arguing instead that prosodic structure

doi:10.6342/NTU202500546

assignment operates cyclically at each word formation and derivation stage rather than solely at the final stage.

These two scholars' insights align well with their explanations of English phonological features, such as flapping. Kahn focused on the conditions for flapping related to syllable boundaries and outlined specific phonological rules. In contrast, Kiparsky provided a systematic overview of English stress patterns, addressing flapping within the broader prosodic unit known as the 'foot.' This indicates that flapping arises from the interplay between syllable boundaries and prosodic hierarchy.

2.1.2. Refutation of Existing Theories

Challenges to phonological rules remained. Riehl (2003) examined whether American English flapping could be explained by paradigm uniformity (Steriade, 2000), which favors consistent morpheme realization across words sharing the same paradigm. According to this principle, base and inflected forms with identical paradigms should maintain the same realizations. However, most cases did not follow this pattern. The study demonstrated that paradigm uniformity does not govern flapping patterns by identifying inconsistent pronunciations within word pairs (e.g., *'negative'* and *'negativistic'*).

Shport et al. (2018) provided fresh insights into rare flapping situations by examining flaps before stressed vowels. In typical English, /t/ becomes a flap before unstressed vowels, but *'whatever'* is an exception. Here, the intervocalic /t/ becomes a flap even though it comes before a stressed vowel. The /t/ in *'whatever*,' stressed on the second syllable, lasts about 30ms, comparable to a standard flap duration. Conversely, the final /t/ in *'everwhat'* lasts roughly 60ms. This study argues that *'whatever'* defies conventional word-internal flapping patterns, indicating a need for new phonological rules that permit flaps before stressed vowels at morpheme boundaries. Nonetheless,

doi:10.6342/NTU202500546

given that *'whatever'* comprises two phonological words, it might qualify as a word-final flapping setting.

2.1.3. Other Effects Affecting Flapping

Moreover, subsequent research analysed morphological structures and individual characteristics in flapping environments. Patterson and Connine (2001) examined variations in the pronunciation of word-medial /t/ in American English corpora. Their key finding showed that while the word-medial /t/ is primarily pronounced as a flap, the ratio varies significantly based on word frequency. High-frequency words exhibited a flap pronunciation rate of 95.4%, whereas low-frequency words displayed only 76.1%. Furthermore, the morphological structure proved significant: monomorphemic words like *'water'* had a flap pronunciation rate of 95.9%, while complex words such as *'dirty'* had a rate of 62.8%. This provided evidence that both word frequency and morphological structure influence allophone variation.

Kilbourn-Ceron and Goldrick (2021) analysed speaker variation and identified connections between flapping patterns and speech planning dynamics. In two speech experiments, their study investigated the relationship between phonetic variation and word form encoding during pre-planning. It assessed isolated word pronunciations and variations influenced by subsequent words with specific sounds. Flaps—a context-sensitive phonetic variant of /t/—increased among English speakers when high-frequency words (e.g., *ocean*, *island*) were included in phrases or when response delays occurred, and decreased when low-frequency words (e.g., *acorn*, *okra*) followed. However, the impact of phonetic variation based on subsequent word frequency was less pronounced under delayed responses. Overall, the results demonstrated a notable increase in flap

doi:10.6342/NTU202500546

usage when conditions favored planning, highlighting that the extent of speakers' pre-planning fluctuates continuously over time.

2.2 Acoustic Properties of Flaps

2.2.1. Duration of Intervocalic /t/ and /d/

Duration is the most significant acoustic parameter for analyzing intervocalic /t/ variants. Considering acoustics, Zue and Laferriere (1979) conducted an acoustic analysis of word-medial /t/ and /d/ in American English, showing notable distinctions between flapped and non-flapped instances. The average duration for flapped sounds was brief—26ms for /t/ and 27ms for /d/. In contrast, non-flapped sounds occurring before a stressed syllable showed significantly longer average durations of 129ms for /t/ and 75ms for /d/. Additionally, flap duration was found to be longer after high vowels or diphthongs compared to other vowels.

In a comparative analysis of Japanese and American English, Price (1981) examined the acoustic variations between flapped /t/ and non-flapped /t/. The results revealed significant differences in their durations. It was found that the average articulation time for flapped /t/ was under 47 ms, primarily occurring between unstressed syllables. In contrast, non-flapped /t/ exhibited a longer and more distinct articulation, averaging over 78 ms, typically preceding stressed syllables.

2.2.2. RMS Amplitude and Intensity

In subsequent research, the analysis extended beyond duration to encompass consonant strength comparisons through intensity and RMS amplitude measurements. Lavoie (2001) examined the acoustic and articulatory features of consonantal weakening using electropalatography. The research aimed to investigate the acoustic and articulatory

doi:10.6342/NTU202500546

characteristics in both English and Spanish, emphasizing phonetically expressed consonantal strength. It was observed that variations in amplitude measurements likely stemmed from differences in sonority between flaps and stops. The results indicated that flaps typically have shorter durations than [t] and display intensity values (measured in RMS amplitude) that are higher compared to [t].

Parker (2002) noted that sonority is closely linked to acoustic intensity (dB) in linguistics, suggesting it can function as a measure of segment resonance. The primary experiment investigated the acoustic characteristics of English and Spanish across five key dimensions: intensity, the frequency of the first formant, total segment duration, peak intraoral air pressure, and combined oral and nasal airflow. The phonemes /t/ and /d/ were frequently pronounced as flaps in stressed syllable-final positions before an unstressed [ə]. This weakening process resulted in greater intensity for these phonemes than bilabial or velar sounds, illustrating increased sonority in post-stress environments. In Spanish, the flap ([ɾ]) consistently demonstrated higher intensity than [t], irrespective of its placement at the beginning or end of word syllables.

Warner et al. (2009) examined the acoustic cues that help listeners identify the flapped allophones of /t/ and /d/ as their weaker consonant forms in American English. The research concentrated on how intensity reduction, consonant duration, and changes in the fourth formant (F4) influence perception. It underscored the importance of a noticeable intensity dip that indicates tongue closure, highlighting that intensity and duration are essential markers for recognizing flapped consonants. The study assessed how the extent of intensity reduction impacted perception, suggesting that a more pronounced intensity dip increased the likelihood of the sound being recognized as a consonant.

doi:10.6342/NTU202500546

Herd et al. (2010) conducted a comprehensive acoustic and perceptual analysis of the flap phenomenon associated with /t/ and /d/ in American English. The results indicated that the vowels preceding /d/ were significantly longer than before /t/, and female speakers demonstrated a higher flapping rate than their male counterparts. Furthermore, speakers tended to use flaps more frequently when they were less aware of the /t/ and /d/ distinction, a factor that, unlike previous research, showed no correlation with word frequency or morphological complexity. While flapped tokens were recorded at a notably higher amplitude (68 dB) than unflapped tokens (64 dB), some overlap in intensity measurements between the two types persisted tokens.

2.2.3. Other Properties

Recent studies have demonstrated progressive development, utilizing tools like PRAAT to distinguish the acoustic characteristics of different flap variants or clarify acoustic differences in flapping between word-internal and word-boundary positions contexts.

Derrick and Schultz (2013) investigated the acoustic characteristics of flap variants in North American English. Eighteen native speakers participated in the study, where ultrasound imaging captured the midsagittal plane of the tongue, and PRAAT software was used to analyse fundamental frequency (F0) and formants. The research identified four flap variants: alveolar tap, down flap, up flap, and postalveolar tap, each showing notable differences in fundamental frequency and formant structure. Nearly non-rhotic vowels exhibited distinct acoustic variations across all variants, with F1 indicating tongue height and F2 representing anterior-posterior movement. The findings confirmed that flap variants are acoustically distinct, revealing differences in tongue movement pronunciation.

doi:10.6342/NTU202500546

Hwang and Kang (2018) studied flapping in word-internal and word-boundary contexts using the Buckeye Corpus. Their analysis identified similarities related to underlying voicing: 1) consistent flap duration with varying F2 profiles and 2) differences in vowel duration before the flaps. They proposed that flapping exemplifies 'incomplete neutralization,' marked by uniform flap duration alongside distinct acoustic signals. The factors affecting flapping can be classified as follows: 1) Lexical level: the pitch and duration of preceding vowels are crucial, and 2) Post-lexical level: flap duration is a significant signal. This suggests a strong influence of foot structure in word-internal contexts, while factors such as speech rate, frequency, and pauses influence flapping at word boundaries.

Building on prior research, it has been noted that flapping usually results in a shorter duration (Zue & Laferriere, 1979; Price, 1981). Regarding intensity, flaps exhibit higher values than stops (Lavoie, 2001; Parker, 2002; Warner et al., 2009). Additionally, flapping interacts with surrounding vowels (Zue & Laferriere, 1979). These acoustic properties vary depending on lexical and post-lexical contexts (Hwang & Kang, 2018). In this study, the acoustic characteristics of flaps are vital for determining whether intervocalic /t/ shifts to a flap, serving as a benchmark for comparing intervocalic /t/ variants within words to those at word boundaries. Based on previous research, it is essential to recognize that classifying TSM's intervocalic /t/ by its acoustic features allows for verification of whether flapping has occurred based on duration and intensity.

2.3 Articulatory Basis of Flaps

2.3.1 Methods of Verifying Articulatory Properties

Studies examining flapping from an articulatory perspective have progressed to integrate and validate the flapping conditions identified in prior research. In his 1998

doi:10.6342/NTU202500546

study, de Jong utilized X-ray microbeam experimental phonetic analysis to demonstrate that flaps can arise from gradual articulatory shifts rather than strict categorical rules. While Kahn's (1976) phonological conditions remain relevant, actual speech patterns exhibit greater complexity. Research indicates that flaps are optional in Kahn's phonological contexts, particularly at word boundaries. Moreover, flaps often result from increased articulatory interaction between vowels, with biomechanical constraints involving the tongue body and tip that disrupt the lenition process of the flap. This suggests that flapped stops and the subsequent unstressed syllables receive less articulatory emphasis, indicating that speakers may exert less effort when producing strong stops.

Furthermore, several research projects have examined the articulatory features of tap sounds using digital methods to precisely measure elements like tongue placement, movement patterns, and pronunciation accuracy. Price (1981) employed dynamic palatography to analyse the articulatory features of flapped /t/ versus non-flapped /t/. The research revealed that flapped /t/ and /d/ had varying contact points from 4.5 mm to 9 mm from the teeth, demonstrating less precision. In contrast, non-flapped /t/ and /d/ consistently made clear contact in the alveolar-palatal area at an anterior range of 4.5 mm to 6.75 mm. Additionally, the study noted differences in contact area: flapped /t/ had a limited range of tongue-palate contact and shorter contact duration, whereas non-flapped /t/ exhibited a broader contact area and a longer contact duration. These results suggest that non-flapped /t/ maintains stable and precise articulatory characteristics, while flapped /t/ reveals more significant variability and less distinct articulation features.

Fukaya and Byrd (2005) studied the flapping of the English word-final sounds /t/ and /d/ using an Electro-magnetic Midsagittal Articulometer (EMA). Tracking the tongue-tip movements of three speakers found that flaps were significantly shorter in

doi:10.6342/NTU202500546

duration—averaging 20 ms compared to 43 ms for regular /t/—and were primarily voiced. Additionally, voiceless sounds had longer durations, averaging 77 ms, particularly at phrase boundaries. The study identified three factors contributing to the brief duration of flaps: the diminished spatial trajectory of the tongue tip, faster articulatory actions, and overlap with the subsequent vowel. There were notable variations in articulatory characteristics among the speakers: while some lowered and adjusted their tongue tips, others began with their tongue tips already lowered. This observation implies that different articulatory gestures can achieve comparable acoustic results.

2.3.2 Research Using Ultrasound Imaging

Prior research has employed various methodologies. X-ray microbeam experimental phonetic analysis poses challenges in real-time measuring participants' articulatory movements. Electronic devices such as dynamic palatography and EMA are also placed in the mouth. These devices accurately detect precise tongue positions and calculate contact areas. However, they often disrupt participants' natural speech production due to the nature of invasiveness.

As Stone (2005) noted, ultrasound provides non-invasive, real-time imaging of tongue movements. The ultrasound probe captures reflection signals from the tongue's surface to create images, which helps in extracting tongue surface contours. This capability makes ultrasound ideal for detailed analysis of tongue movements in pronunciation research, allowing precise measurement and evaluation of variations across different pronunciation contexts. However, the quality of images can be affected by factors such as probe angle, tissue moisture, and the subject's physical characteristics. Proper positioning of the probe and fixation of the head are essential for accurately

doi:10.6342/NTU202500546

analyzing intricate tongue movements and structures. Nonetheless, ultrasound may have difficulty capturing high-speed tongue actions like flapping due to its limited temporal resolution.

Derrick and Gick (2011) studied variations of the English flap and tap sounds using ultrasound imaging and acoustic analysis with eighteen North American English speakers. Their findings indicated that flaps and taps can manifest as variants related to individual speaker traits, even within the same phonological contexts. The vertical tongue movement pattern associated with flaps was the most prevalent. While some speakers consistently used specific variants, others flexibly switched between them. This study demonstrated that flap and tap variations are not solely influenced by phonological context; instead, they can differ based on the speaker's individual choices, prompting a reevaluation of the boundaries between phonology and phonetics.

Derrick and Gick (2015) analysed the generation of multiple flaps in a single motor gesture while pronouncing *'Saturday.'* By utilizing ultrasound imaging with twenty-six native speakers of North American English, the study tracked tongue movements during the pronunciation process. The findings revealed that 84.5% of the participants produced three distinct sounds (Up-flap → Retroflex → Down-flap) in a seamless motion. Additionally, 3D model simulations of the oral structure, muscle movements, gravity, and elasticity confirmed these results. This research illustrated that natural forces like gravity and elasticity can effectively facilitate the production of sequential sounds without necessitating separate motor actions.

In summary, the articulatory features outlined above indicate that the flapped /t/, influenced by biomechanical limitations, displays different articulatory patterns compared to the non-flapped /t/ (de Jong, 1998). These patterns lead to durational differences resulting from spatial reduction, decreased articulatory time, and overlap with

doi:10.6342/NTU202500546

the following vowel (Fukaya & Byrd, 2005). Even for acoustically similar flaps, individual strategies for tongue movement can differ (Price, 1981; Fukaya & Byrd, 2005), and various possibilities for articulation methods exist (Derrick & Gick, 2011). Drawing from these earlier studies, the intervocalic /t/ in TSM is anticipated to showcase similar articulatory traits, making ultrasound analysis an effective tool for intuitively visualizing these distinct tongue movements. Considering the characteristics of flaps, such as brief duration and swift tongue motion, the M-mode analysis method proposed by Derrick and Gick (2015) appears particularly fitting for this research.

2.4 Consideration of intervocalic /t/ in TSM

In TSM, applying the conditions for intervocalic /t/ flapping in American English indicates that its phonological rules primarily focus on gemination and the voicing of stop consonants under specific phonological circumstances (Chiang 1992; Fon & Khoo 2025). These scholars contend that in TSM, the /t/ phoneme is subject to voicing and is realized as [l] because TSM lacks the phoneme /d/.

Chiang (1992) described 'affix-triggered processes' as phonological changes that occur when affixes attach to TSM words. When vowel-initial affixes merge with stems, the final consonants of the stems become geminated. Simultaneously, syllable-final obstruents begin to voice. Stops differ in three forms: aspirated voiceless, unaspirated voiceless, and voiced. However, due to the absence of the [d] phoneme in TSM's phonological system, syllable-final alveolar consonants shift to [l] through voicing. This transformation illustrates TSM's phonological traits and highlights the phonological connections and pronunciation alterations at syllable boundaries.

Fon and Khoo (2025) discussed the gemination phenomenon in TSM, where consonants at the end of a syllable replicate to the onset of the following syllable,

doi:10.6342/NTU202500546

primarily with certain suffixes. For instance, *kim-á* ("gold") is pronounced [kim.ma] and *a̍p-á* ("box") as [ab.ba]. This phenomenon also occurs with the nominalizing suffix -ê, as seen in *âng-ê* ("Something red") →[aŋ.ŋɐ]. Due to TSM's absence of /d/, the final /t/ transforms into [l] during voicing, illustrated by the phrase *tsi̍t-ê* ("one"), pronounced as [tsil.le]. This gemination exemplifies the distinctive linguistic aspects of TSM, blending phonological rules with morphological functions.

From a morphological perspective, research has compared the consonant alternation between TSM and Taiwanese Hakka. Wang and Liu (2010) explored how syllable-final consonants carry over to the initial syllable of the following word. They conducted two experiments: the first assessed how participants perceived liaison consonants as distinct onsets. In contrast, the second examined whether these consonants functioned as onsets when the order of bisyllabic words was reversed. The findings indicated that TSM speakers rarely recognized liaison consonants as onsets (14%), suggesting that phonological rules primarily govern them. In contrast, Hakka speakers, particularly those from the Sixian dialect, exhibited more excellent onset recognition rates than those from the Hailu dialect. This study highlights an early stage of morphologization for liaison consonants in both languages, offering valuable insights into language process change.

In conclusion, the intervocalic /t/ in TSM has been examined alongside other stop consonants. It has drawn attention for its gemination and voicing under certain phonological conditions (Chiang 1992; Fon & Khoo 2025). Importantly, TSM does not contain the phoneme /d/, which leads to the phenomenon where intervocalic /t/ is reported to become voiced and realized as [l].

Additionally, when comparing the flapping of intervocalic /t/ in English to TSM, it is essential to consider that TSM's intervocalic /t/ may also appear as a flap as geminated

doi:10.6342/NTU202500546

[l] does not appear to be always guaranteed. Flapping in English has been thoroughly studied from phonological, acoustic, and articulatory perspectives, providing a valuable framework for exploring similar occurrences in TSM. This framework allows for a deeper understanding of the variations in intervocalic /t/ in TSM.

Research focused specifically on TSM's intervocalic /t/ has been limited. Previous studies indicate that TSM's phonological constraints shape the distinctive realization patterns of intervocalic /t/ (despite Chiang 1992 and Fon & Khoo 2025 proposing its realization as [l]), suggesting that further independent exploration is necessary.

Therefore, building on the previous research above, this study aims to analyse TSM's intervocalic /t/ independently. I will use ultrasound imaging to identify articulatory features and measure the similarities among various postures. This will reveal its realization patterns based on morphological contexts within and across word boundaries while verifying each category's acoustic and articulatory properties.

doi:10.6342/NTU202500546

Chapter 3 Methods

To test the hypothesis that similar phonological changes occur in morphologically identical environments, I developed a two-track analysis approach using distinct data types:

1. Voice data to verify acoustic differences

2. Ultrasound data to examine articulatory variations

As illustrated in Figure 3.1, this dual-track analysis framework was developed to investigate how different morphological environments affect intervocalic /t/ in TSM.

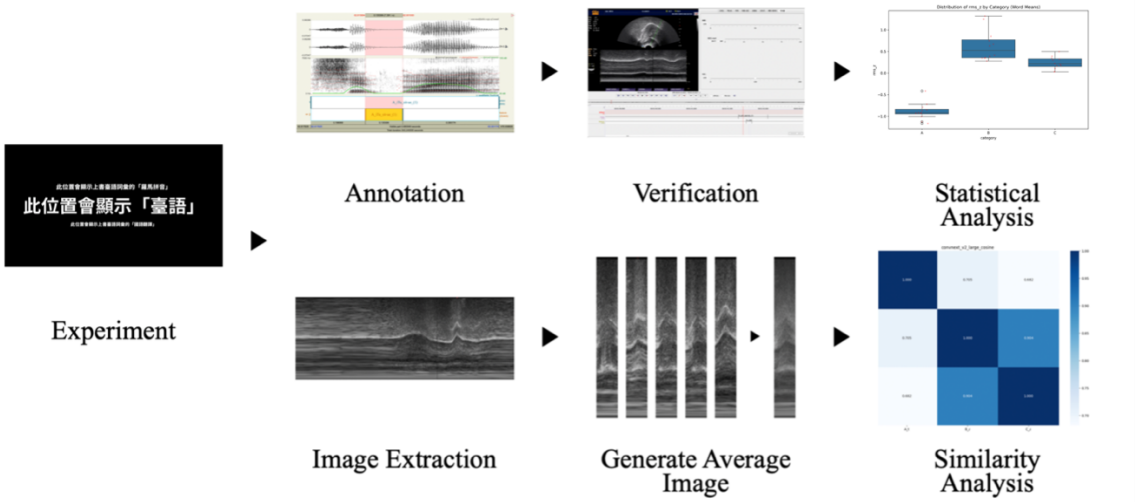


Figure 3.1: Analysis Procedures.

3.1. Experimental Design

3.1.1. Stimulus Materials

In this study, stimulus words were prepared across three categories reflecting morphologically distinct environments to systematically analyse /t/ variation.

doi:10.6342/NTU202500546

The first category consists of words composed solely of lexical morphemes. For example, *'t͡sʰut-un'* (出運; "one's luck turns around"), where 出 means "to go out, breakthrough" and 運 means "luck."

The second category comprises combinations of lexical and grammatical morphemes. A representative example is *'t͡sʰat-a'* (賊仔; "thief"), where 賊 means "thief" and the suffix 仔 functions as a grammatical morpheme indicating a person or object possessing that characteristic.

However, these two categories present a methodological limitation: they only allow observation of /t/ variation within words. Since morphological differences can occur at word boundaries, I additionally designed a third category. This category consists of phrases with a grammatical morpheme /e/ (的) between two disyllabic words. For instance, *'kiŋ-t͡shat e bo-a'* (警察的帽仔; "The hat of the police"), where 警察 means "police," 帽仔 means "hat," and 的 functions as a structural particle connecting these elements to mark attributive.

Table 3.1: Definitions and Examples of Stimulus Categories.

Category	Definition	Example
A	Lexical morphemes	t͡sʰut-un 出運
B	Lexical morpheme followed by a grammatical morpheme (Within a word)	t͡sʰat-a 賊仔
C	Lexical morpheme followed by a grammatical morpheme (Across word boundaries)	kiŋ-t͡shat e bo-a 警察的帽仔

doi:10.6342/NTU202500546

Introducing the third category enables comparative analysis to examine whether similar variations within words are observed at word boundaries when grammatical morphemes follow /t/. The preceding vowels before /t/ were controlled to be /a/, /e/, /i/, or /u/. Each category contained 12 words or phrases, as shown in Table 3.2.

Table 3.2: Distribution of Preceding Vowels Across Stimulus Categories.

Preceding Vowel	Category A	Category B	Category C	Total
a	pat-ui 別位	tʰat-a 窒仔	huat-tat e tɔ-t͡sʰi 發達的都市	9
	kuan-t͡sat iam 關節炎	t͡sʰat-a 賊仔	kiŋ-t͡shat e bo-a 警察的帽仔	
	peʔ-t͡sʰat-ue 白賊話	sip-pat-a 十八仔	sio-bat e kuan-he 相捌的關係	
e	ket-an 結案	ho-pʰet-a 雨撇仔	t͡sʰin-t͡shet e tʰai-tɔ 親切的態度	9
	ket-uɑn 結怨	kim-ket-a 金桔仔	lɑu-let e ke-a 鬧熱的街仔	
	t͡set-iok 節約	kuã-let-a 寒熱仔	piʔ-ket e tʰoŋ-kʰɔ 祕結的痛苦	
i	sit-ue 失話	t͡sui-pit-a 水筆仔	pi-bit e kɔ-su 秘密的故事	9
	pit-iɑu 必要	t͡sʰit-a 妗仔	pʰin-t͡sit e bun-te 品質的問題	
	lit-au 日後	paʔ-t͡sit-a 肉虱仔	sẽ-lit e i-gi 生日的意義	
u	put-an 不安	sut-a 術仔	pɑi-kut e t͡su-bi 排骨的滋味	9
	t͡sʰut-ĩ 出院	kʰut-a 窟仔	ui-kʰut e piɑu-t͡siŋ 委屈的表情	
	t͡sʰut-un 出運	but-a-hi 鮇仔魚	ki-sut e t͡sin-po 技術的進步	
Total	12	12	12	

doi:10.6342/NTU202500546

3.1.2. Participant Selection

Participants were recruited through online advertisements. The selection criteria were as follows: (1) adults aged 18-40 with normal vision and hearing, (2) Taiwanese nationals with at least one parent/guardian being a native TSM speaker, (3) individuals who primarily use TSM for household communication, and (4) individuals with no residence experience outside Taiwan exceeding 6 months before the age of 18.

Based on these criteria, eight native TSM speakers were recruited (eight individuals, evenly split between males and females, with an average age of 25.25). During recruitment, participants' TSM proficiency was self-assessed using a scale of 1-10, with an average score of 7.5 for 8 participants. Language proficiency was further evaluated through a short conversation, including self-introduction—only participants who met the required level proceeded with the experiment. The participants were naïve of the research hypotheses before the experiment. Analysis was conducted using data from all eight participants.

3.1.3. Experimental Procedures

The experiment was conducted in a sound-controlled laboratory environment. Participants were seated in front of a 13-inch monitor positioned at an appropriate desk height to ensure optimal stimulus viewing. While presenting stimuli as images would have been beneficial for natural speech production, some stimuli, such as idiomatic expressions (e.g., *sẽ-lit e i-gi*; "The meaning of one's birthday") or abstract concepts (e.g., *pit-iau*; “needs”), could not be easily represented visually. Consequently, the stimuli were displayed in Traditional Chinese characters (e.g., 水筆仔; "Kandelia obovate") at the center of the screen. To assist participants in pronouncing unfamiliar terms accurately, pronunciation guides (e.g., *tsuí-pit-á*) from the Ministry of Education Taiwan Dictionary

doi:10.6342/NTU202500546

were provided above the stimuli, with Mandarin Chinese explanations(e.g., 水生植物; “Pleuston”) in the dictionary below. Figure 3.2 illustrates this layout, ensuring participants could efficiently produce the instructed words or phrases.

Figure 3.2: Layout of Experimental Stimulus Presentation.

Before the main experiment, participants completed a practice session to prevent unnatural 'lab speech' and familiarize them with the experimental setup. During this session, they practiced two blocks of stimuli, arranged differently from the actual experiment, to encourage natural pronunciation patterns. Following the practice session, participants were fitted with an ultrasound probe and securely positioned beneath their chins using a stabilizer to minimize movements. The probe's position was carefully adjusted to ensure optimal visualization of tongue movements during speech production.

The main experiment consisted of six blocks, each containing 36 randomly ordered stimuli displayed for 3 seconds per stimulus. A three-minute rest period was provided after every two blocks to minimize participant fatigue. The total experimental duration was approximately 30 minutes, and participants were compensated at a rate of NTD 200 per hour, with proportional payment for shorter durations.

doi:10.6342/NTU202500546

3.2. Data Collection Methods

3.2.1. Apparatus

Acoustic data were captured using a microphone and digitized through a USBPre 2 for computer storage. Ultrasound imaging was captured by an OPUS5100 equipped with PV65, and data was saved to the computer system via an AverMedia Extremecap U3.

3.2.2. Data Collection Procedure

Ultrasound data were recorded in B/M mode, which enables simultaneous observation of overall tongue morphology in B-mode and detailed temporal tracking of tongue movements in M-mode. This dual-mode functionality provides a comprehensive view of articulatory motion's spatial and dynamic aspects. The data were recorded at 40 frames per second, with B-mode capturing tongue midsagittal images and M-mode tracking tongue movements over time.

The experimental procedure consisted of an initial calibration phase, during which participants produced nine rapid [ta] repetitions as reference points designed for ultrasound and acoustic signal synchronization. This synchronization process involved aligning the tongue release moment with the onset of acoustic signals using Adobe Premiere Pro 2024, with the initial [ta] sequences serving as temporal reference points. The main experiment followed, comprising six blocks of 36 stimulus repetitions each (216 utterances in total per participant).

While the ultrasound system did not feature the three parallel M-mode lines described by Derrick and Gick (2015), it employed a single scanning line that moved in a wiper-like motion from a central point. This scanning line was positioned individually for each participant at the highest point of tongue-tip contact with the alveolar ridge

doi:10.6342/NTU202500546

during [ta] articulation, as shown in Figure 3.3. This scanning line position was selected as the reference point because it provided optimal capture of tongue movement during intervocalic /t/ production.

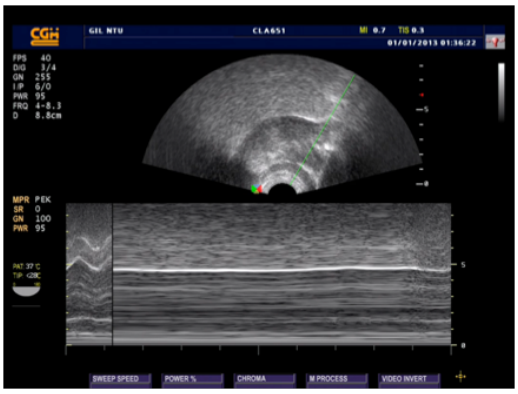


Figure 3.3: Representative B/M-Mode with the Activated Scanning Line (Green).

3.3. Analysis Methods

3.3.1. Acoustic Analysis (PRAAT)

After data collection, segments were labeled and annotated using PRAAT. First, the collected acoustic files were loaded into PRAAT, creating a TextGrid file. Two tiers were made in the TextGrid. Word and /t/ utterance intervals were measured according to the segmentation methodology of Herd et al. (2010). In the first tier, the utterance intervals of all tokens were labeled as separate intervals. Word duration was measured from the visible start to the end of the waveform and spectrogram. The second tier was used to analyse /t/ utterance intervals. The utterance boundaries were defined from the endpoint of the preceding vowel to the starting point of the following vowel, and acoustic boundaries were identified by carefully observing the periodic changes in waveforms and spectrograms. During the analysis, /t/ utterances showed various characteristics; the postures of each token were classified based on the researcher's auditory judgment. Various allophone realization types were observed during this process, including

doi:10.6342/NTU202500546

unreleased stops, flaps, glottal stops, and deletions. These postures and boundaries were meticulously recorded in detail in the second tier (Figure 3.4).

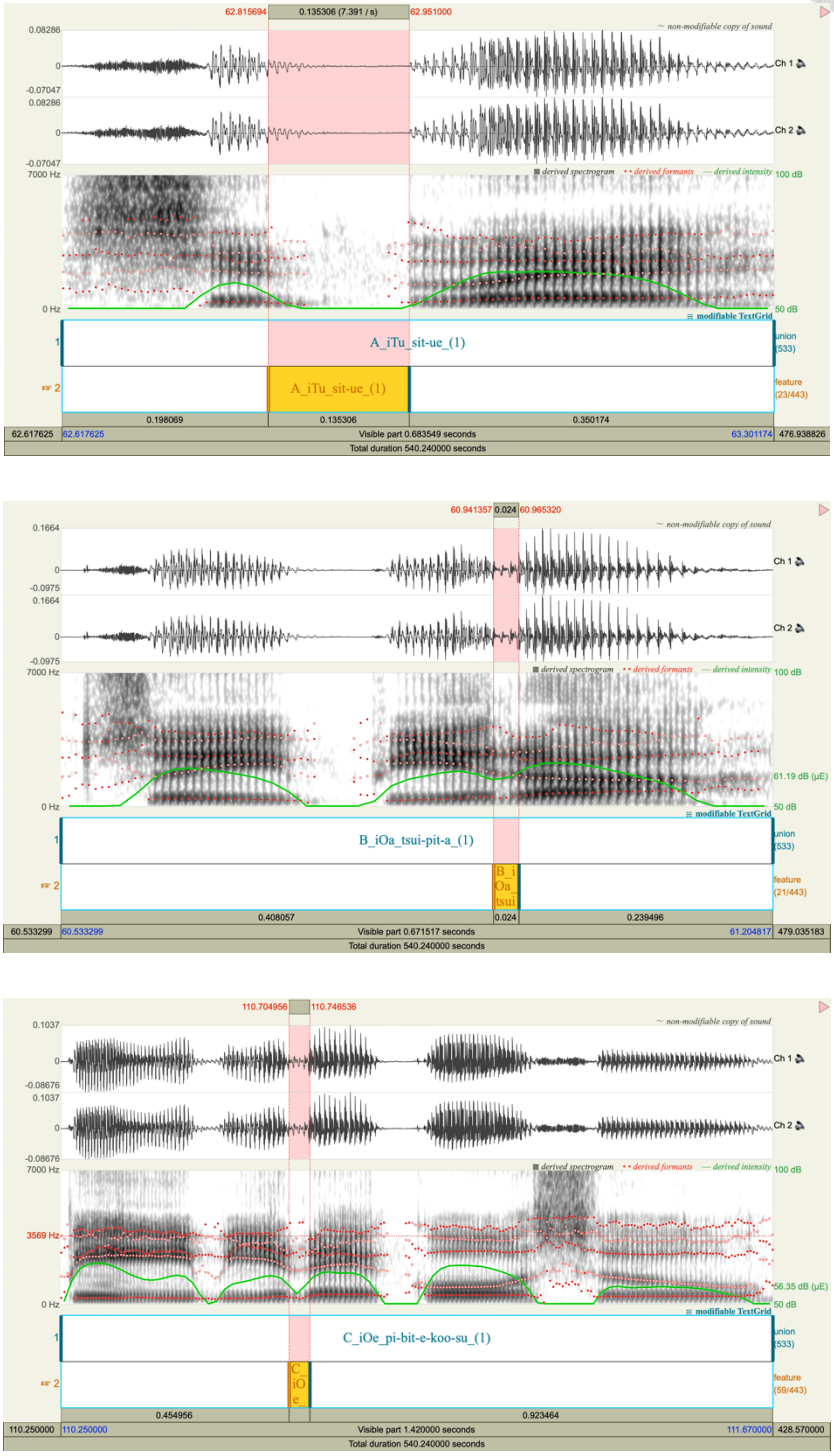
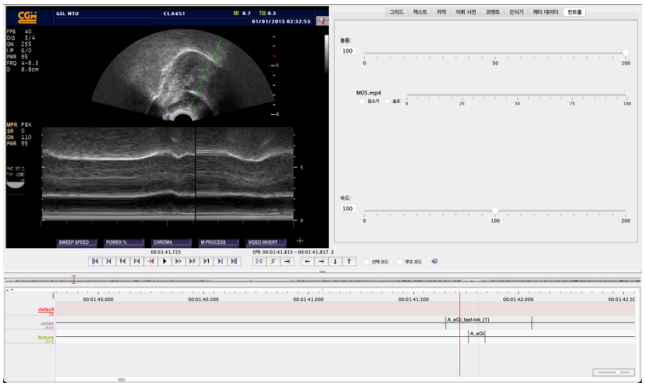


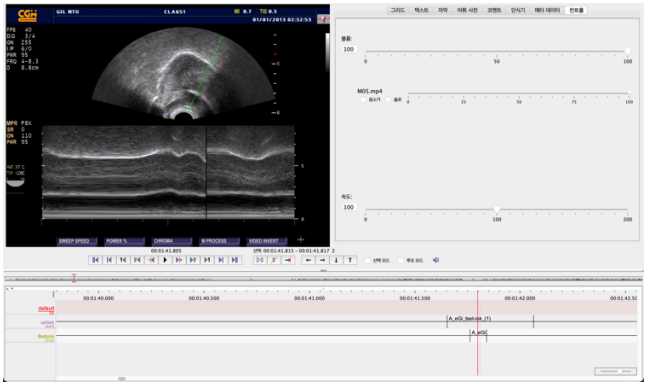
Figure 3.4: Examples of /t/ Allophones (Category A's [t], Category B and C's Flap [ɾ]).

doi:10.6342/NTU202500546

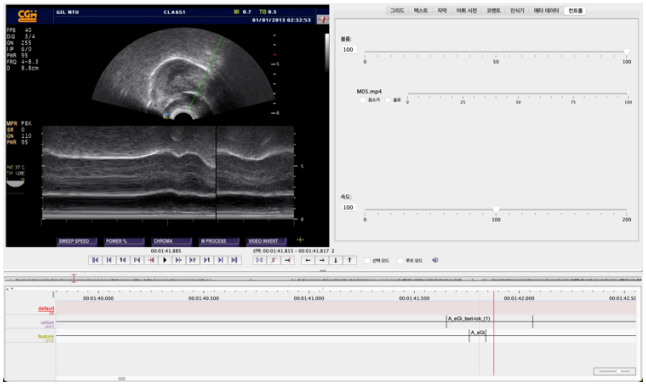
In cases where it was difficult to distinguish the utterance postures of /t/, the ELAN program (Wittenburg et al., 2006) was utilized. First, the ultrasound images recorded during data collection and the TextGrid created based on acoustic analysis were loaded into ELAN. Then, based on the midpoint of the relevant interval in the TextGrid, a total of five frames - two before and after the midpoint - were visually examined. The utterance postures were confirmed by considering both the direction of tongue movements and the corresponding acoustics (Figure 3.5).



Target frame -2



Target frame



Target frame + 2

Figure 3.5: ELAN-Based Posture Verification Procedure.

doi:10.6342/NTU202500546

For all /t/ utterance interval labels in tiers 1 and 2, a consistent naming convention was followed in the order of (Category)_(PV)(Posture)(FV)_(Token)_((Trial)), where PV refers to Preceding Vowel and FV refers to Following Vowel, to facilitate subsequent analyses. However, in cases of deletion [D], since the /t/ utterance interval could not be specified, annotations were recorded only in the first tier. The distribution ratio was analysed for these intervals without conducting acoustic and articulatory analyses.

To comprehensively analyse the phonetic characteristics of /t/ segments, two key acoustic measures were examined:

1. /t/ duration: This is a primary indicator for distinguishing flapping/tapping from plosives, as shorter durations are typically associated with the former (Zue & Laferriere, 1979). The duration of /t/ was calculated as the interval between the end of the preceding vowel and the beginning of the following vowel (Herd et al., 2010).

2. RMS amplitude of intensity: The RMS amplitude of intensity is an acoustic measure that quantifies the energy levels of /t/ segments. It reliably indicates flapping or tapping phenomena (Lavoie, 2001).

Box plots visualized the statistical results for duration and intensity across categories A, B, and C.

3.3.2. Articulation Analysis

As mentioned earlier, B-mode was used when it was difficult to distinguish specific postures during acoustic analysis. M-mode images were used to identify how each posture differed in articulation. Meanwhile, M-mode provided an intuitive

doi:10.6342/NTU202500546

understanding of tongue movement, direction, and speed, enabling the identification of similarities and unique articulatory postures across groups.

This study used preprocessed ultrasound M-mode images to analyse articulation. To acquire initial images, TextGrid files and video files were synchronized. The second tier of the TextGrid contained annotations of the target /t/, but when attempting to obtain images at the midpoint of these annotations, the images would be truncated. Therefore, based on the TextGrid file containing the time information of the utterance intervals, frames were extracted one second after the end time of the annotated interval in the video. This was done to obtain images after utterance to capture the target /t/.

All B/M-mode images were cropped to 800px × 300px, extracting only the M-mode section from the original 1024 × 768 images. The analysis focused on the central 200-pixel vertical section of the ultrasound images, where the tongue tip movement patterns during /t/ production are most clearly visible. This represents the range excluding 50 pixels from both the top and bottom of the original 300-pixel vertical dimension (Figure 3.6). The region was selected based on the direction and speed of tongue tip movement, which are the defining characteristics of /t/ variations. This approach ensured that the analysis concentrated on the most experimentally relevant areas, filtering out peripheral elements that could cause noise or redundancy.

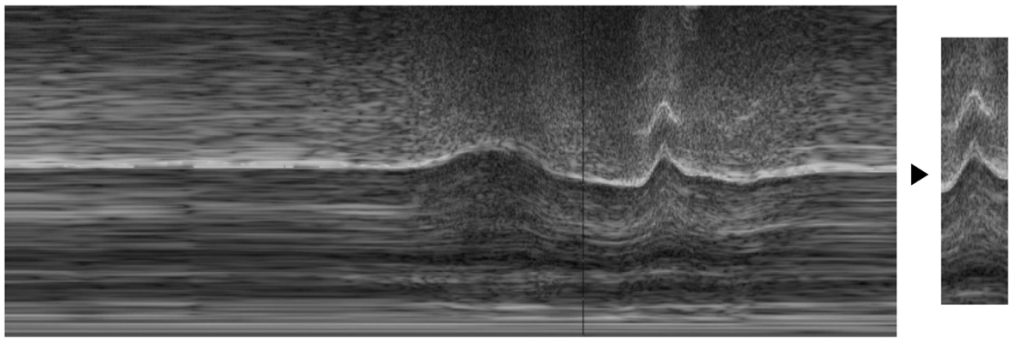


Figure 3.6: Image Preprocessing (Left: Before Preprocessing, Right: After Preprocessing)

doi:10.6342/NTU202500546

The images, cropped to the same size, were grouped by identical postures within each category, and their arithmetic mean images were calculated for use in similarity analysis. Figure 3.7 shows examples of arithmetic mean images for /t/ articulation within the same word.

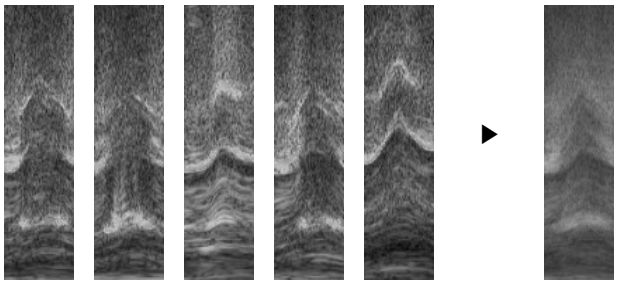


Figure 3.7: Five Tokens of /t/ in [sip-pat-a] from Participant M01 (left) and the Corresponding Averaged Image (right)

The analysis of preprocessed ultrasound images employs CoAtNet-2 (Dai et al., 2021), a neural network architecture that combines convolutional and attention-based mechanisms to compare postures across categories using cosine similarity measurements. CoAtNet-2 was initialized with pre-trained models from the TIMM library (Wightman, 2019). CoAtNet addresses the limitations of traditional CNNs by integrating Transformer technology, originally developed for natural language processing, with convolutional layers. Its hybrid architecture mitigates these challenges by blending local feature detection through convolutional layers with global pattern recognition through Transformer layers, making it effective for datasets of varying sizes. This study leverages CoAtNet's capabilities to extract detailed and comprehensive feature vectors from ultrasound images, enabling the analysis of intricate articulatory patterns even with limited data.

doi:10.6342/NTU202500546

Since CoAtNet-2 was initially designed for images with an input size of 224px × 224px, in this study, the cropped images were resized to the input dimensions while maintaining the core configuration for feature vector extraction.

Similarities between different /t/ realizations were quantified using cosine similarity measurements. This metric calculates the cosine of the angle between two vectors, producing values between 0 and 1, where 1 indicates identical patterns, and 0 means completely different patterns. Cosine similarity was used to analyse high-dimensional vectors, focusing on directional patterns over absolute magnitudes. This metric enables a quantitative approach to analyzing pattern similarities in ultrasound imaging data, supporting the exploratory nature of this study.

Before applying these analytical methods to the collected data, a preliminary analysis was conducted to verify whether CoAtNet-2's feature extraction and cosine similarity measurement could be effectively used for ultrasound images (Figure 3.8). The methodology was tested by analyzing the similarity of cropped /t/ articulation ultrasound images in the utterance *'ui-kʰut e piau-t͡siŋ'* from W04. The images were cropped along the x-axis using the highest point of /t/ as the reference point. First, the center x-coordinate (center_x) of the /t/ peak was identified. Using this center_x as a reference, the images were cropped at nine different positions, moving from -12 pixels to +12 pixels in 3-pixel increments. At each position, the image was cropped with a width of 50 pixels, taking 25 pixels on either side of the center point. For instance, at the -12 pixel offset, the crop started at (center_x - 12) - 25 pixels and ended at (center_x - 12) + 25 pixels. This process generated nine new images from each input image, each 50 pixels wide but centered at different positions along the x-axis. This method creates an effect similar to viewing the image through a window that gradually moves from left to right.

doi:10.6342/NTU202500546

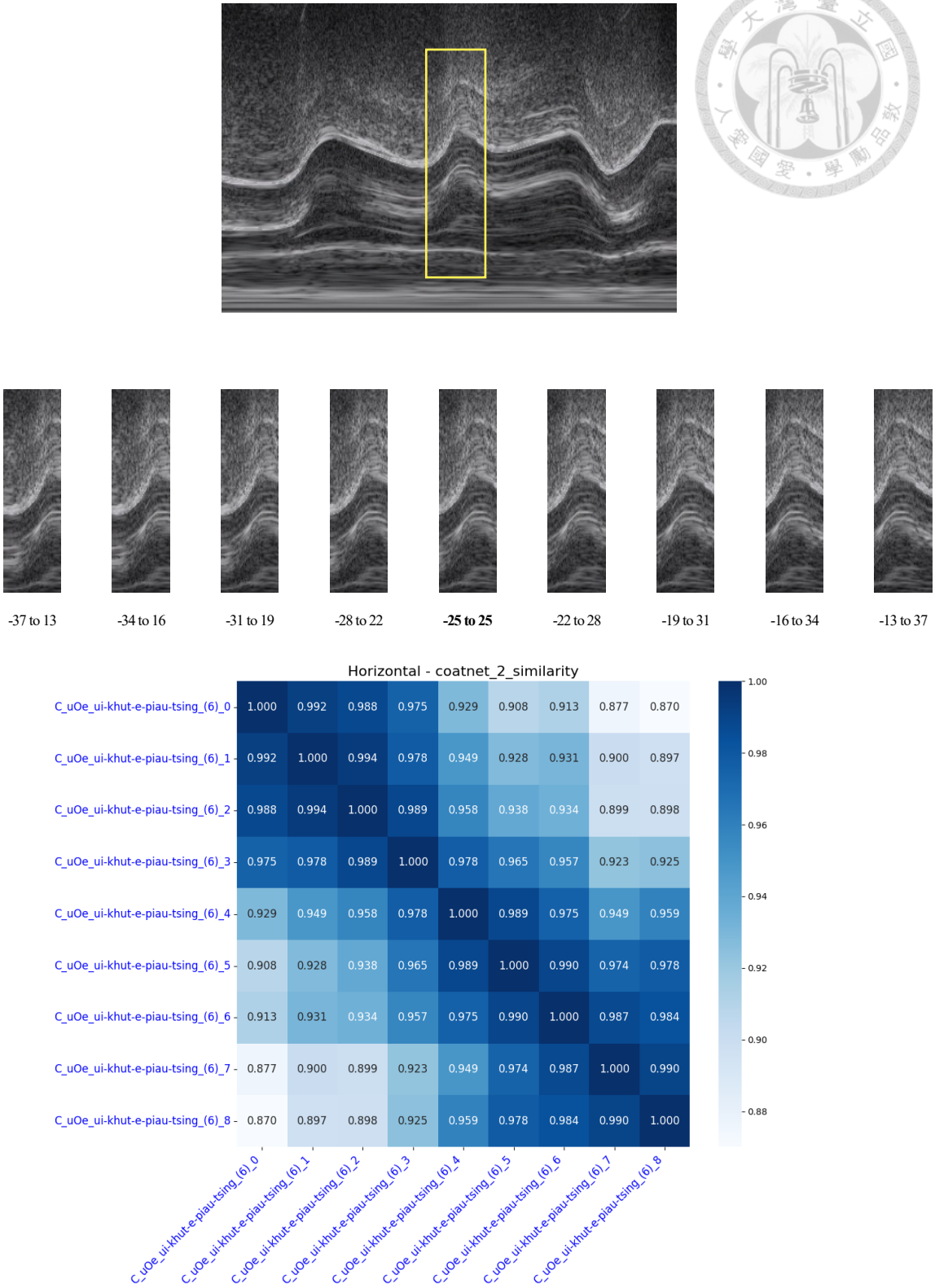


Figure 3.8: Validation of Ultrasound Image Analysis by Integrating CoAtNet and Cosine Similarity (W04 - /t/ in ui-k^{h}ut e piɑu-t͡siŋ).

doi:10.6342/NTU202500546

If the expectations were correct, images from adjacent positions should show more substantial similarities (indicated by darker colors). In comparison, images from more distant positions should show weaker similarities (indicated by lighter colors). The preliminary test results largely aligned with these initial expectations. Based on these promising results, it was determined that the combination of CoAtNet-2 feature extraction and cosine similarity measurement enabled the objective quantification of pattern similarities across categories despite some limitations in ultrasound image analysis, such as grayscale image processing or resizing issues. This led to the decision to apply this model to the collected data.

It should be noted that since this validation method examines the same utterance by shifting pixels from left to right over time, the range of similarities presented here does not correspond to the range of similarities shown in subsequent research results.

The results were visualized as heatmaps to demonstrate word similarity patterns across different morphological contexts. This methodology complemented visual inspection methods, enhancing the overall robustness of this analysis.

3.3.3. Expected Results

Based on the experimental procedures and analyses outlined above, the following outcomes are anticipated:

A. Acoustic Results

1. A clear acoustic distinction is expected between Categories A and B. For example, in terms of duration, Category A is expected to be longer, while Category B is expected to be shorter than A. As for the RMS amplitude, Category A is expected to be lower, while Category B is expected to be higher than A.

doi:10.6342/NTU202500546

2. Category C is expected to demonstrate acoustic characteristics more closely aligned with Category B, as it was hypothesized that intervocalic /t/ sharing morphological similarities would exhibit the same alternation patterns, regardless of whether they occur within words or across word boundaries.

B. Similarity Analysis Results

1. Cases exhibiting flap formation in Categories B and C are expected to show higher similarity values than other category pairs. For example, when compared to the expected [t] posture in Category A, the flaps [ɾ] in Categories B and C should demonstrate higher similarity scores with each other than with the [t] in Category A.

doi:10.6342/NTU202500546

Chapter 4 Results

4.1. Acoustic Results

4.1.1. Duration Analysis

Analysis of duration distributions through box plot visualization of the z-score data (Figure 4.1) indicates that condition A is characterized by significantly extended durations relative to conditions B and C. Moreover, condition A exhibits markedly higher duration variability, as evidenced by its broader distributional spread, whereas conditions B and C demonstrate more constrained duration distributions with heightened central tendency. This distinct pattern suggests fundamental differences across conditions.

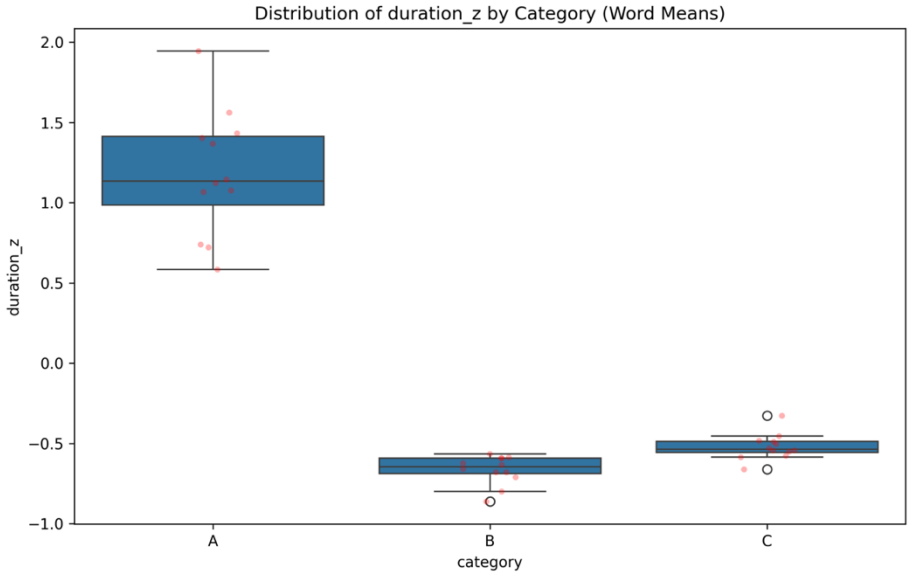


Figure 4.1: Box Plot of Duration Z-score Data of /t/ by Category.

A one-way ANOVA was conducted to examine the differences in standardized scores (z-scores) among groups categorized by duration. The z-scores were derived by standardizing the data for each participant, and further statistical analyses were performed on the combined standardized scores of all participants. The results showed a statistically significant effect of category on duration, $F(2, 33)=228.21$, $p < .0001$.

doi:10.6342/NTU202500546

Analysis through Tukey HSD post-hoc testing revealed that the mean difference between groups A and B was -1.8454, with a 95% confidence interval of [-2.0812, -1.6096], and a p-value less than 0.0001 ($p < .0001$), indicating a statistically significant difference. The mean difference between groups A and C was -1.7015, with a 95% confidence interval of [-1.9373, -1.4657] and a p-value less than 0.0001 ($p < .0001$), showing a significant difference. The mean difference between groups B and C was 0.1439, with a 95% confidence interval of [-0.0919, 0.3797] and a p-value of 0.305, indicating no statistically significant difference. The analysis revealed statistically significant differences in mean duration values between groups A and B, and groups A and C, but not between groups B and C.

4.1.2. RMS Amplitude Analysis

The box plot (Figure 4.2) illustrating RMS amplitude z-scores indicates that Categories B and C display significantly higher values compared to Category A. Moreover, unlike Duration measurements, RMS amplitude shows greater variability in Categories B and C. In contrast, Category A exhibits a more restricted RMS amplitude distribution characterized by a stronger central tendency. Importantly, Categories B and C, sharing similar morphological traits, reveal substantially broader overlapping distributions in comparison to Category A

doi:10.6342/NTU202500546

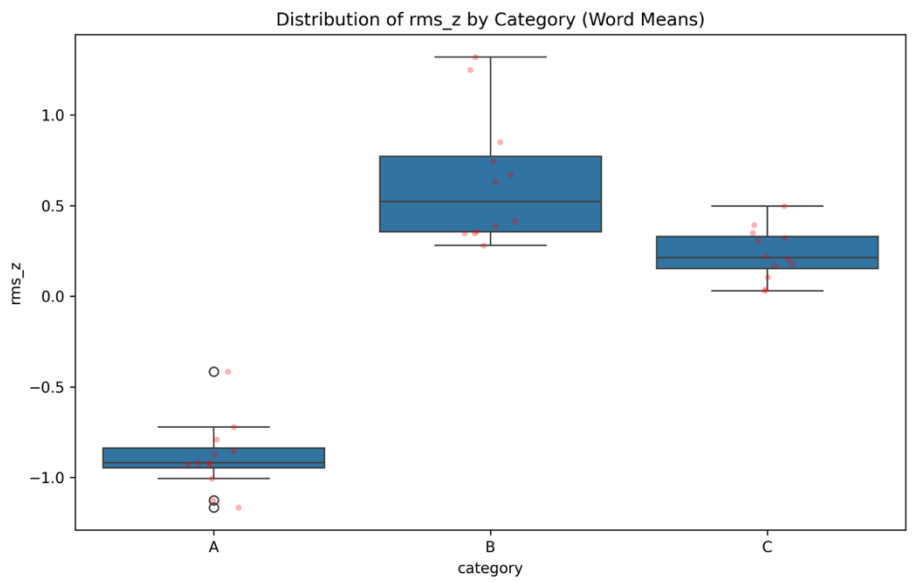


Figure 4.2: Box Plot for RMS Amplitude Z-score Data of /t/ by Category.

A one-way ANOVA was conducted to analyse the group differences in standardized RMS amplitude (z-scores) based on the categorical variable category. For z-scores, data were first standardized for each participant individually, and then statistical analyses were performed using the standardized scores from all participants. The analysis indicated a statistically significant effect of category on RMS amplitude, $F(2, 33)=121.70$ $p<0.0001$.

Subsequent Tukey HSD post-hoc analysis produced the following results: The mean difference between groups A and B was 1.5205, with a 95% confidence interval of [1.2726, 1.7685] and a p-value less than 0.0001 ($p < .0001$), showing a statistically significant difference. The mean difference between groups A and C was 1.1215, with a 95% confidence interval of [0.8735, 1.3695] and a p-value less than 0.0001 ($p < .0001$), indicating a significant difference. The comparison between groups B and C revealed a mean difference of -0.399, with a 95% confidence interval of [-0.647, -0.151] and a p-value of $p = 0.0011$, also showing a significant difference. The Z-score analysis confirmed

doi:10.6342/NTU202500546

that statistically significant differences in RMS values exist among all three groups of the categorical variable category.

4.2. Articulatory Results

4.2.1. Distribution of utterances

When analyzing all participants combined (Figure 4.3), for intervocalic /t/, category A showed [t] at 54.4% and [ʔ] at 36.8%. In category B, [ɾ] demonstrated a dominant proportion at 95.7%, while in category C, although not as high as B, [ɾ] maintained a notably high percentage at 83.7%.

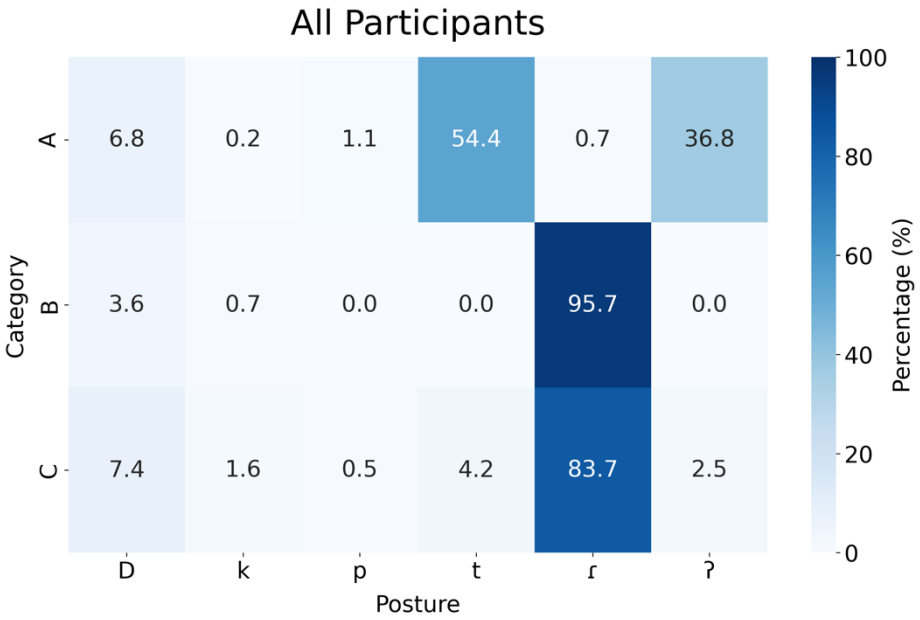


Figure 4.3: Overall Distribution of Intervocalic /t/ Postures (The X-axis Labels are in IPA, and the capital D represents deletion).

Figure 4.4 shows the distribution of utterances across all participants through a heatmap visualization. Note that the number of posture categories on the x-axis varies between participants, as postures not exhibited by an individual are excluded from their

doi:10.6342/NTU202500546

respective heatmap. Darker colors represent higher proportions, while lighter colors indicate lower proportions. The figure reveals several common patterns among participants.

In category A, most participants exhibited a dominant posture exceeding 70%; the most prevalent postures in category A were [t] and glottal stop [ʔ]. In the case of W01, [t] accounted for 45.8%. A notable observation was that for W04, both [t] and [ʔ] maintained high proportions ([t] at 58.6% and [ʔ] at 41.4%).

For category B, all participants except W01 produced flaps [ɾ] in over 90% of cases. W01 produced flaps [ɾ] in 81.9% of instances, with the remaining cases showing deletion [∅].

In Category C, all participants except M02 and W01 demonstrated flap [ɾ] distributions exceeding 70%, with M02 and W01 showing flap productions of 69.4% and 66.7%, respectively. While Category C's articulation patterns more closely resemble Category B than Category A's, it is distinguished by its greater diversity in postures than Category B.

doi:10.6342/NTU202500546

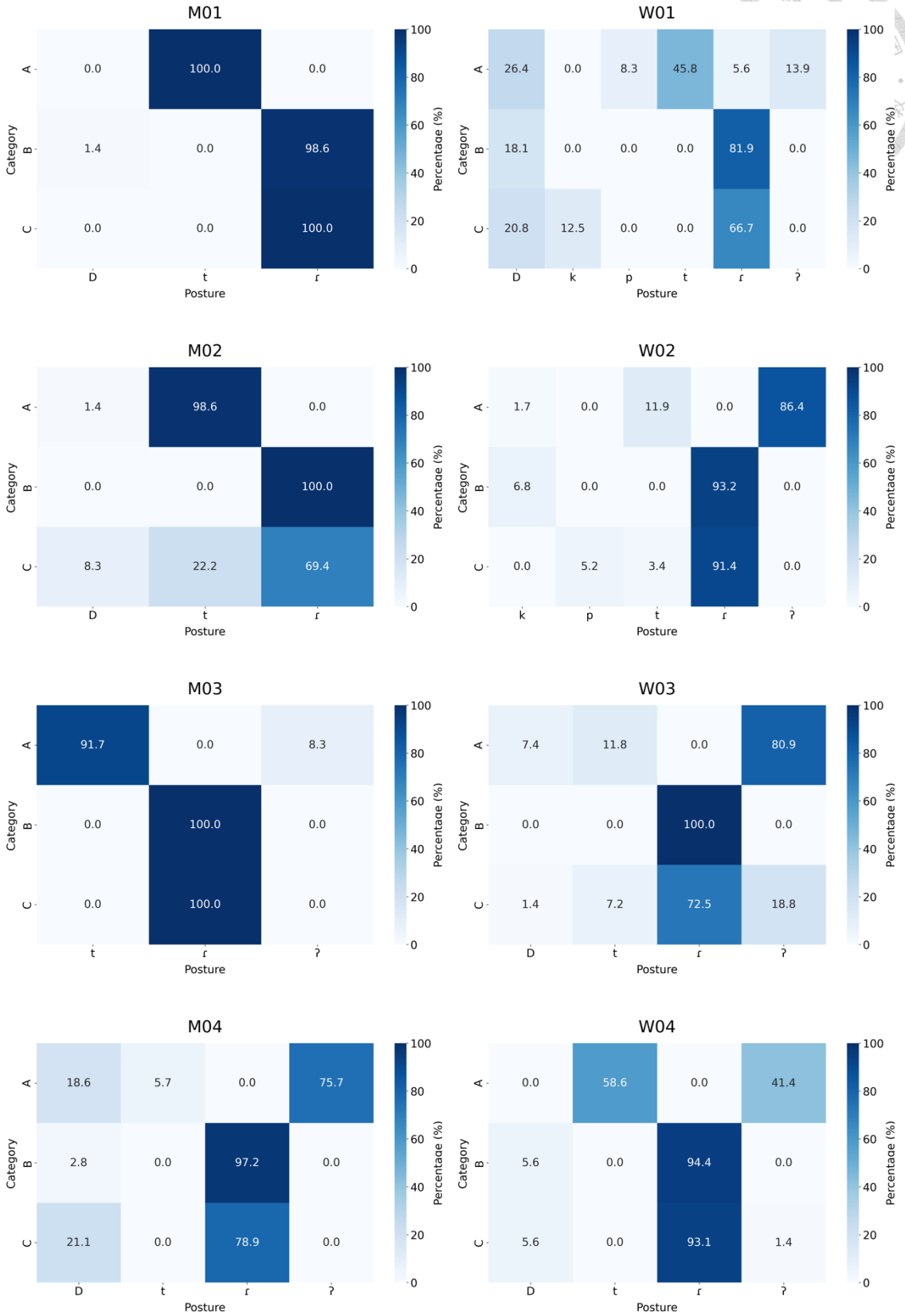


Figure 4.4: Distribution of Intervocalic /t/ Postures Across Categories (The X-axis Labels are in IPA, and the capital D represents deletion).

doi:10.6342/NTU202500546

4.2.2. Similarity Analysis

Figures 4.5 and 4.6 illustrate each participant's average images for /t/ utterances by posture. These visualizations are derived from feature vectors extracted using CoAtNet-2, and analysis was conducted using cosine similarity measurements.

As previously mentioned, category A's predominant postures were [t] and glottal stop [ʔ]. For posture [t], most average images display a smooth curved pattern that rises and falls. In contrast, the glottal stop [ʔ] posture shows an opposite directional pattern, with protrusions trending downward or forming a straight line at the bottom. This pattern resembles the [k] posture observed in W02's data.

The flap [ɾ] primarily occurs in categories B and C, displaying a directional pattern similar to posture [t], characterized by an initial upward movement followed by a downward trajectory. However, unlike [t], the flap features a sharper peak instead of a curve. This is supported by the findings of the acoustic analysis: the flap is articulated with the tongue briefly making contact with the alveolar ridge, and this articulatory action is captured in the M-mode images. The observed articulatory pattern aligns with previous descriptions of flaps, which involve quick tongue tip contact and release from the alveolar ridge (Zue & Laferriere, 1979).

Additionally, participant W02 demonstrated a [p] posture, which displayed a pattern similar to [t] with an extended straight-line formation. However, the peak amplitude for [p] consistently appeared lower than that observed for [t].

Analysis conducted using the CoAtNet-2 for visual feature extraction, coupled with cosine similarity metrics, revealed that the most salient pattern in the similarity heatmap was the high correspondence between flap movements occurring across categorical boundaries. This pattern demonstrated consistency across the entire participant.

doi:10.6342/NTU202500546

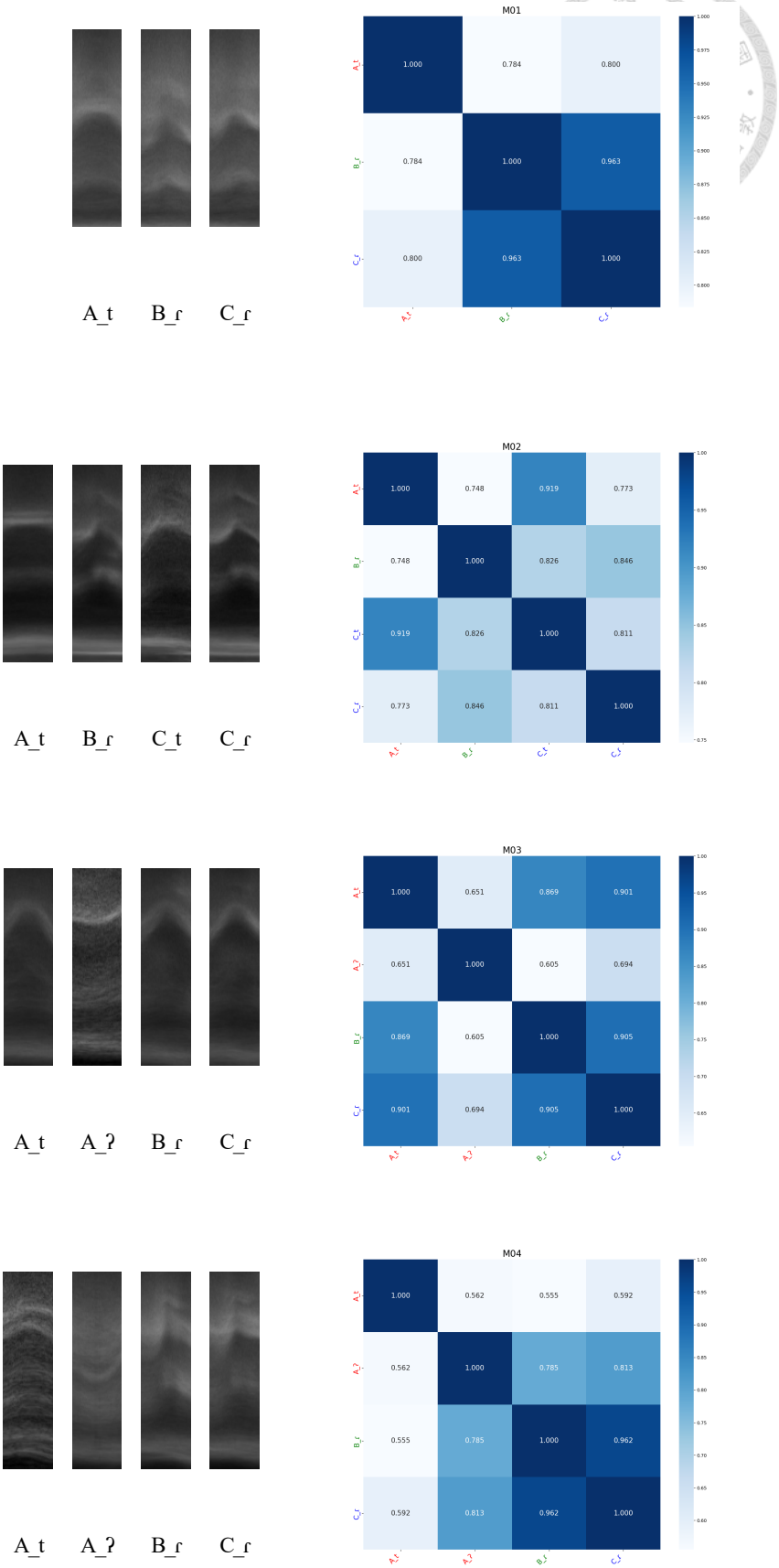


Figure 4.5: Averaged Images of /t/ and their Similarity Heatmap (M01-M04).

doi:10.6342/NTU202500546

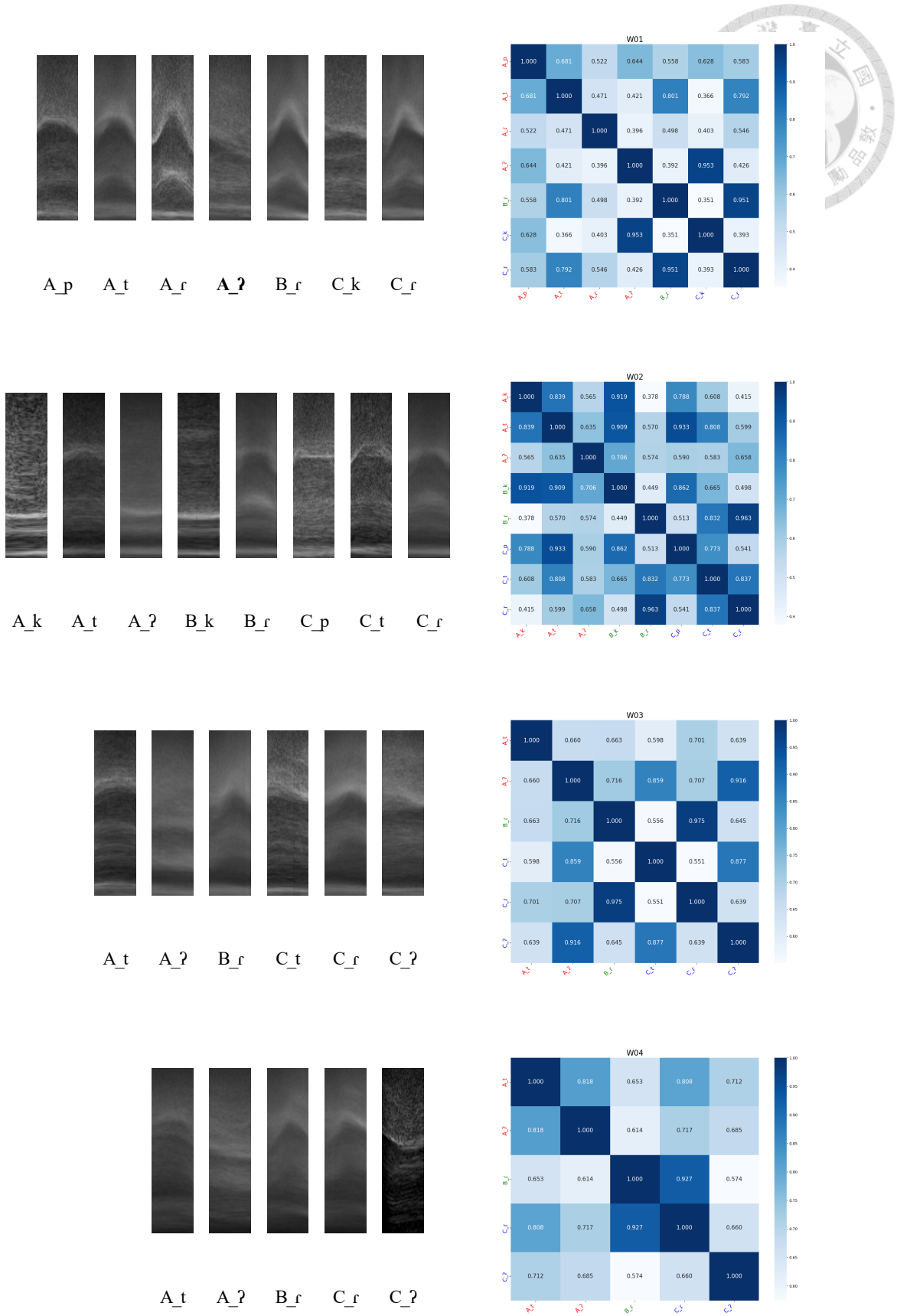


Figure 4.6: Averaged Images of /t/ and their Similarity Heatmap (W01-W04).

doi:10.6342/NTU202500546

Chapter 5 Discussion

5.1. Posture Distribution

This study hypothesized that cases with the same morphological structure would show similar patterns in posture distribution regardless of whether they occurred within words or at word boundaries. Accordingly, the data was divided into three categories. Category A consists of words composed of combinations of lexical morphemes, where intervocalic /t/ is positioned at the boundary between lexical morphemes. Category B includes words composed of lexical and grammatical morphemes combinations, where intervocalic /t/ is placed at the boundary between the lexical and grammatical morphemes. Finally, Category C consists of phrases with two words connected by a grammatical morpheme, with intervocalic /t/ positioned at the boundary between the first word and the grammatical morpheme. Therefore, Categories B and C were predicted to show similar variant distributions, as they share identical morphological contexts. At the same time, Category A was expected to display a different variant distribution pattern from these two categories.

The primary postures in Category A were [t] and [ʔ]. M01, M02, and M03 showed [t] ratios exceeding 90%, while W01 and W04 had lower but still predominant [t] ratios. As mentioned earlier, W04 displayed characteristics different from other participants (see Figure 4.4), with [t] and [ʔ] accounting for 58.6% and 41.4%, respectively, showing a unique pattern where both postures maintained high percentages even though [t] exceeded half. For M04, W02, and W03, [ʔ] emerged as the primary posture in Category A.

The predominant posture in Category B was [ɾ], frequently observed among all participants (see Figure 4.4). As previously mentioned, Category B includes words formed by combining lexical and grammatical morphemes. In Category C, [ɾ] was the predominant posture, a feature commonly observed across all participants (see Figure

doi:10.6342/NTU202500546

4.4). The occurrence rate of [ɾ] ranged from a minimum of 66.7% (W01) to a maximum of 100% (M01, M03). This result aligns with the study's expectations because I hypothesized that similar morphological contexts would induce similar intervocalic consonant alternations.

The variation in the ratio between Category B and Category C is a noteworthy phenomenon, and previous studies provide insights into this difference. Patterson and Connine (2001) found that in English, simple words exhibit a higher flapping rate than derivatives, suggesting that the structural stability of words influences their phonetic realization. Although the word-formation system of TSM does not entirely align with that of English, the current findings resonate with Patterson and Connine's (2001) observations. Notably, even when the structural composition involves a combination of lexical and grammatical morphemes, intervocalic /t/ within a word exhibits a higher flapping rate than intervocalic /t/ at word boundaries. This suggests that intervocalic /t/ within a word is realized as a "stabilized" form, leading to a higher fixed flapping rate compared to its realization at word boundaries.

When analyzing posture diversity, Category C shows notable differences from Category B. While Category B included only [ɾ] as a posture besides deletion instances, Category C exhibited up to three distinct postures along with deletion. Cho (2004) points out that flapping in American English is a multifaceted phenomenon shaped by numerous influences, including the properties of adjacent sounds, stress patterns, their organization within syllables and words, phonetic aspects, semantics, and speech pace.

In Category C, lexical and grammatical morphemes merge "between words." While it shares morphological patterns with Category B, the context between words adds complexities, such as differences in speech rate or prosodic emphasis. These factors can create distinct patterns compared to Category B, confined to a more limited intra-word

doi:10.6342/NTU202500546

context. This may explain the greater variety of posture variations seen in Category C compared to Category B.

Moreover, while not the primary phoneme examined in this study, the variant [ʔ] frequently occurs in the intervocalic /t/ of TSM and is also a common variant of the intervocalic /t/ in English. This occurrence has been documented in American English (Seyfarth & Garellek, 2020) and British English (Chong & Garellek, 2023). Roach (1973) noted the phenomenon of glottalization affecting English /p/, /t/, and /k/, finding that [ʔ] appears more frequently among younger speakers (under 40 years old) while being less common among older speakers. This generational variation reflects a growing trend of glottalization in English pronunciation, which signifies phonological evolution. Hence, detecting the [ʔ] variant in the intervocalic /t/ of TSM can provide significant insights into the phonetic similarities between English and TSM. This connection offers vital evidence that supports the universality of phonetic variation patterns and the similarities shared between the two languages.

5.2. Acoustic Aspect

5.2.1. Main Findings

From an acoustic perspective, Category A was expected to maintain the canonical acoustic characteristics of [t] (longer duration and low RMS amplitude), Category B was predicted to undergo flapping due to different morphological characteristics from A, thus expected to show short duration and high RMS amplitude, and Category C was hypothesized to show acoustic characteristics similar to B due to sharing the same morphological features. The acoustic aspects most heavily focused on in this study were duration and RMS amplitude, with the hypothesis that when intervocalic /t/ undergoes

doi:10.6342/NTU202500546

flapping, its duration would decrease. In contrast, its RMS amplitude would increase compared to [t].

Analyzing the duration reveals that Category A has the most diverse distribution. The extensive duration range in Category A arises from targets represented by five distinct phonemes—four stops ([t], [p], [k], [ʔ]) and [ɾ]—each having unique duration profiles. In Category B, the narrow and brief duration is linked to a uniform realization as [ɾ] by all participants (apart from instances of deletion). Meanwhile, even when excluding deletion cases, Category C shows a slightly broader and longer duration distribution attributed to specific realizations as stops.

The RMS amplitude showed different patterns from duration: Category A displayed the narrowest spread, while Category B showed the widest distribution. This can be attributed to Category A being predominantly realized as stops (except for 0.7% of [ɾ]), resulting in uniformly low RMS values. In contrast, [ɾ] tokens were sometimes so briefly positioned between vowels that measurement was "impossible," suggesting significant influence from surrounding vowels. While some tokens exhibited postures similar to Category A for Category C, the majority (83.7%) were realized as [ɾ], resulting in RMS amplitude patterns more identical to Category B.

The statistical results revealed an interesting pattern: in the box plots, Categories B and C showed more comparable results than Category A. However, Category C exhibited a slightly longer duration than Category B, while its RMS amplitude was somewhat lower than that of Category B. Nevertheless, ANOVA and Tukey statistical analyses confirmed that in duration, there was no statistically significant difference between B and C, while in RMS amplitude, all three categories—A, B, and C—were statistically different.

doi:10.6342/NTU202500546

The statistical difference between Category B and C in RMS amplitude somewhat deviates from the expectation that the acoustic characteristics of B and C would be similar. This can be attributed to the observations in Figure 4.3: although the majority of postures in Category C were realized as [ɾ], a small portion either shared postures with Category A (4.2% for [t] and 2.5% for [ʔ]) or were realized as phonemes with lower RMS amplitude compared to [ɾ] (1.6% for [k] and 0.5% for [p]). While category B consists entirely of flaps (except for deletion), Category C exhibits various stop postures. Another statistical analysis was conducted by extracting only the pure [ɾ] data from categories B and C, excluding the postures realized as stops.

When analyzing the duration difference between groups Category B's [ɾ] and Category C's [ɾ], the t-test revealed a t-statistic of -3.66 and a p-value of 0.0014 ($p < .01$). This indicates a statistically significant difference between the two groups. The negative t-statistic implies that the mean duration of group Category B's [ɾ] (-0.199) is lower than that of group Category C's [ɾ] (0.235). In other words, the duration of posture Category B's [ɾ] was shorter than that of posture Category C's [ɾ].

If this is the case, the assumption that adding postures realized as stops resulted in statistically different outcomes may not be the only account. When statistical analysis was performed simply by categories for duration, there were no statistical differences. However, when Category B's [ɾ] and Category C's [ɾ] were analysed separately, statistical differences were found between the two. This raises questions about other factors influencing these disparities in the statistical analyses. Before interpreting these results, let us analyse how RMS amplitude varies among the groups.

When analyzing the RMS amplitude difference between groups Category B's [ɾ] and Category C's [ɾ], the t-test revealed a t-statistic of 3.19 and a p-value of 0.0042 ($p < .01$). This indicates a statistically significant difference between the two groups. The

doi:10.6342/NTU202500546

positive t-statistic implies that the mean RMS amplitude of group Category B's [ɾ] (0.184) is higher than that of group Category C's [ɾ] (-0.217). In other words, the RMS amplitude of posture Category B's [ɾ] was higher than that of posture Category C's [ɾ].

In summary, the examination of the duration and RMS amplitude of Category B's [ɾ] and Category C's [ɾ] revealed consistent differences. Furthermore, box plots (Figures 5.1 and 5.2) generated from only Category B's [ɾ] and Category C's [ɾ] also displayed noticeable differences.

There is a potential explanation for why statistical testing indicated significant differences when analyzing only Category B's [ɾ] and Category C's [ɾ]. As shown in the previous paper, the flapping that occurs across word boundaries is affected by factors such as speech rate, frequency, and pauses, compared to flapping that occurs within words (Hwang & Kang, 2018). Therefore, even with the same morphological structure, flapping within a word versus across word boundaries may exhibit statistically significant acoustic differences (though both are more similar to each other than to the canonical stop [t]). Moreover, the possibility cannot be ruled out that statistical differences emerged because the following vowels of intervocalic /t/ in Categories B and C were fixed as /a/ and /e/, respectively. This discovery offers a fascinating perspective. While morphologically alike and displaying the same posture for intervocalic /t/, the acoustic characteristics might vary because of positional differences between word-internal and word-boundary contexts.

Despite their similar structural characteristics, these previously identified features offer indirect evidence for this study's finding that intervocalic /t/ variants exhibit statistically distinguishable attributes in both word-internal and word-boundary contexts. The brevity of [ɾ] presents a distinct advantage for acoustic analysis.

Nevertheless, the box plots illustrate that the differences in acoustic features between Category B's [ɾ] and Category C's [ɾ] do not fundamentally challenge this study's

doi:10.6342/NTU202500546

hypothesis. Exploring these subtle discrepancies between word-internal and word-boundary positions may present a valuable direction for future research.

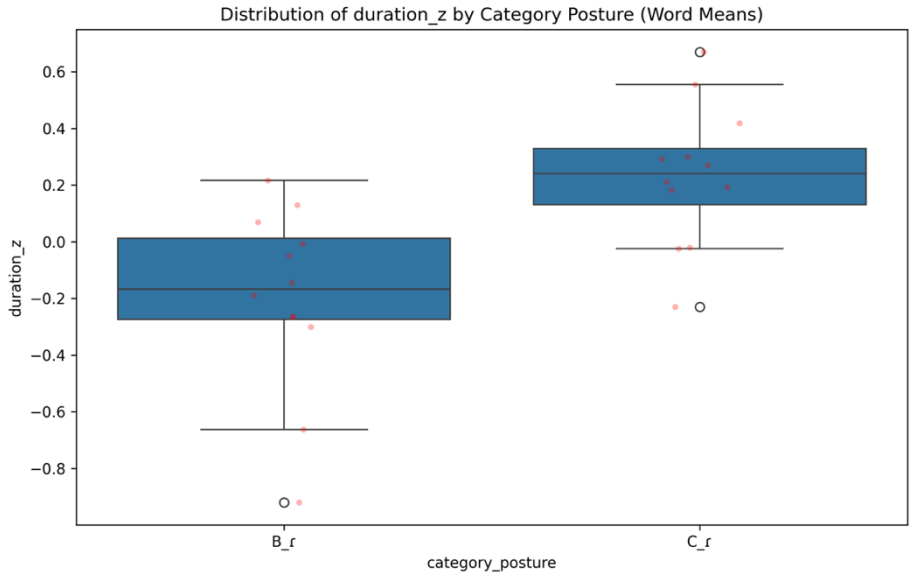


Figure 5.1: Box Plot of Duration (Z-score) for Categories B and C.

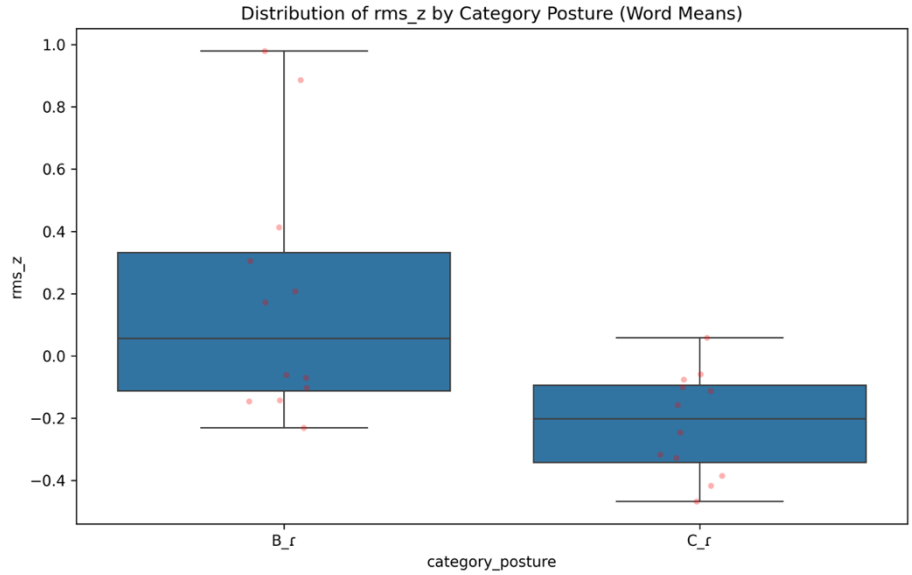


Figure 5.2: Box Plot of RMS Amplitude (Z-score) for Categories B and C.

doi:10.6342/NTU202500546

5.2.2. Implications

The initial hypothesis of this study was that intervocalic /t/ with identical morphological structure would show similar consonant alternation regardless of its position within or across word boundaries. This suggests that categories with the same morphological structure (B and C) should exhibit more similar acoustic characteristics compared to the other category (A).

According to previous research (Zue & Laferriere, 1979; Price, 1981), [ɾ] exhibits a shorter duration and greater intensity (RMS amplitude) (Lavoie, 2001; Parker, 2002; Warner et al., 2009; Herd et al., 2010) compared to [t]. The acoustic analysis of this study revealed that Categories B and C, which share the same morphological structure, demonstrated shorter durations and higher RMS amplitudes compared to Category A. There was no statistical difference in duration between Category B and C (MD = 0.1439, 95% CI [-0.0919, 0.3797], p = 0.305), both categories showed significant differences from Category A (A vs B: MD = -1.8454, 95% CI [-2.0812, -1.6096], $p < 0.001$; A vs C: MD = -1.7015, 95% CI [-1.9373, -1.4657], $p < 0.001$). In RMS amplitude, there were significant differences between all categories, with the difference between Category B and C (MD = -0.399, 95% CI [-0.647, -0.151], $p = 0.0011$) being relatively smaller than the differences between A and B (MD = 1.5205, 95% CI [1.2726, 1.7685], $p < 0.001$) or A and C (MD = 1.1215, 95% CI [0.8735, 1.3695], $p < 0.001$). Furthermore, the boxplot indicated that Categories B and C exhibited similar patterns, while category A displayed a distinct pattern.

The acoustic characteristics of flaps identified in previous studies, such as shorter duration and higher intensity than [t], support the hypothesis that morphological distinctiveness induces intervocalic /t/ flapping. This alternation occurs similarly both

doi:10.6342/NTU202500546

within and across word boundaries. Moreover, these similarities and differences in acoustic characteristics suggest further articulatory analysis.

5.3. Articulatory Aspect

5.3.1. Main Findings

Representative mean images across participant categories (Figure 5.3) reveal distinct articulatory characteristics. Within Category A, the representative posture [t] is characterized by tongue contact with the alveolar ridge followed by a release, resulting in prolonged alveolar contact compared to [ɾ]. This articulatory feature manifests in the ultrasound imaging data as relatively higher tongue positioning with smooth plateau curvature. Notably, some participants (M02) exhibited extended periods of alveolar contact, producing a linear configuration at the superior margin of the cropped temporal window.

Another representative posture in Category A, [ʔ], is characterized by a full closure of the vocal folds (Ladefoged & Maddieson, 1996). Unlike the aforementioned [t], where the tongue moves upward toward the alveolar ridge, [ʔ] does not exhibit an upward tongue posture. Given the absence of vertical movement, the tongue contour manifested either as a gentle downward-oriented curve (M04) or a downward linear configuration (W02, W03) rather than exhibiting sharp contours.

In contrast, Category B's representative posture [ɾ] involves a brief tongue tap against the alveolar ridge followed by downward movement. The key distinction lies in [ɾ]'s requirement for rapid vertical movement completion within a short timeframe, resulting in a sharp peak formation rather than the gradual slope observed in [t]. Regardless, it is noted that both Category B's [ɾ] and Category A's [t] show similar upward tongue movement patterns since both postures require alveolar contact. These two points

doi:10.6342/NTU202500546

(the same directionality and whether it is sharp or not) fittingly reflect the articulatory similarities and differences between [t] and [ɾ].

Visual examination of the representative posture [ɾ] in Category C reveals patterns that generally align with those observed in Category B. The trajectory height is comparable to Category A's [t] and Category B's [ɾ], demonstrating similar vertical positioning in the production. Regarding curvature characteristics, the contour exhibits sharp peak formations similar to Category B's [ɾ] rather than the gradual slopes observed in Category A's [t]. While some cases show less distinct differences upon initial visual inspection (M03), examining internal shading patterns reveals subtle variations in peak sharpness. This shows a slight difference in the subsequent similarity comparison.

doi:10.6342/NTU202500546

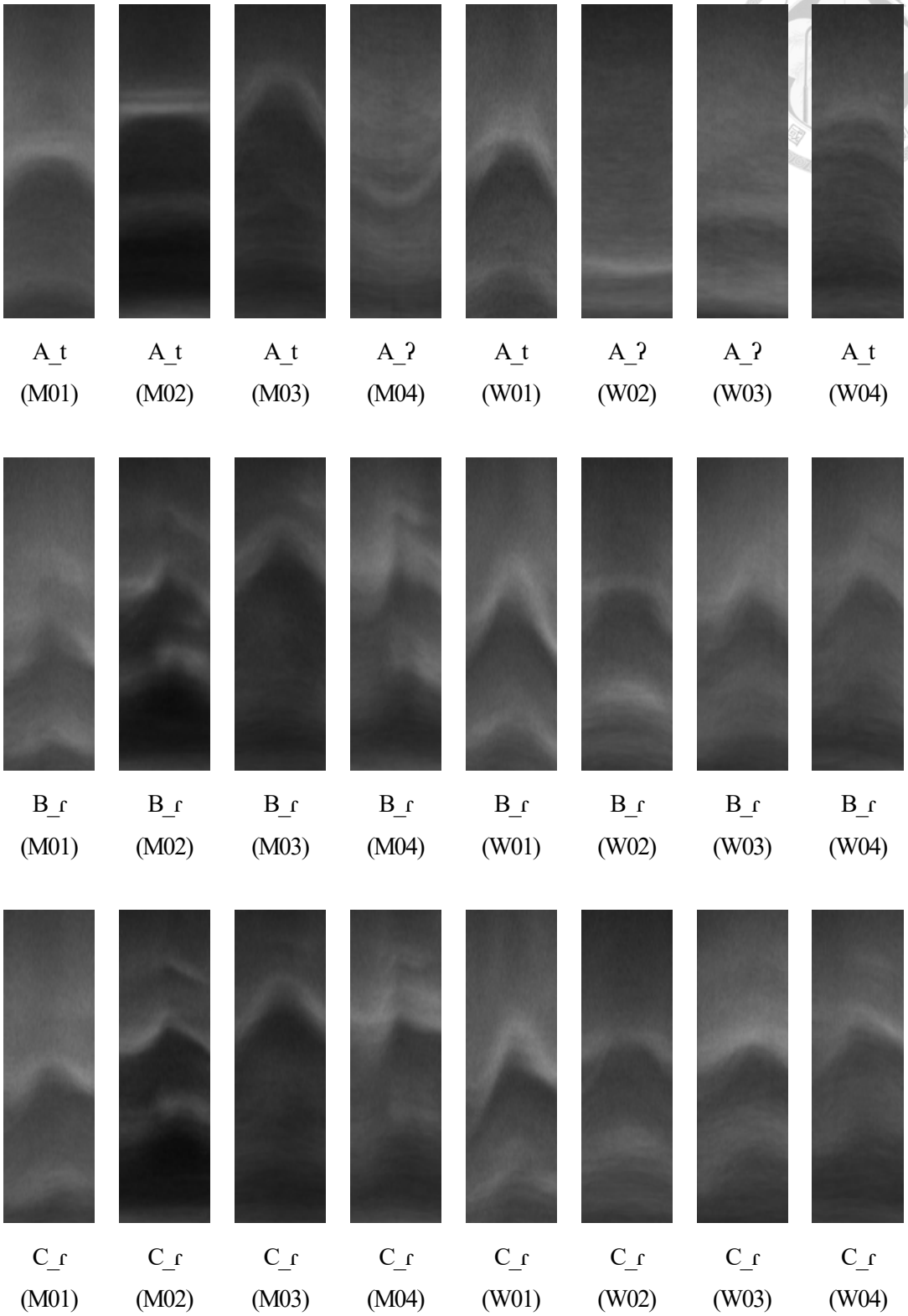


Figure 5.3: Category-wise Average Images of Primary Postures.

doi:10.6342/NTU202500546

The similarity scores between primary postures across categories for all participants (Table 5.1) reveal notable patterns that align with previously presented evidence. The similarity scores between [t] and [ɾ] cluster in a higher range (minimum 0.784 - maximum 0.901), while similarity scores between [ʔ] and [ɾ] distribute across a lower range (minimum 0.574 - maximum 0.813).

One interesting point here is W04's data. W04's Category A productions were represented as 58.6% [t] and 41.4% [ʔ], demonstrating both types of postures. "Fortunately," having such a participant meant that only in her case could it generate representative mean images for both postures. Her results aligned with the overall pattern, with [t]-[ɾ] similarity scores ranging from 0.808 to 0.818, while [ʔ]-[ɾ] similarity scores occupied a lower range (0.614 to 0.685). These findings demonstrate the reliability of this model's image similarity analysis methodology. Additionally, for participant M03, whose differences were less pronounced upon visual inspection, the similarity scores showed minimal variation, although the B-C combination exhibited the highest similarity score (0.905).

Table 5.1: Overall Similarity between Primary Postures between Categories.

	[t]					[ʔ]			
	M01	M02	M03	W01	W04	M04	W02	W03	W04
A_B	0.784	0.748	0.869	0.801	0.818	0.785	0.574	0.716	0.614
A_C	0.800	0.773	0.901	0.792	0.808	0.813	0.658	0.707	0.685
B_C	0.963	0.846	0.905	0.951	0.927	0.962	0.963	0.975	0.927

5.3.2. Implications

Previous studies often compared tongue postures using B-Mode ultrasound imaging. When capturing the tongue with B-mode ultrasound, it is possible to observe

doi:10.6342/NTU202500546

the sagittal plane, which is a vertical cross-section of the body from front to back, allowing observation of the anterior-posterior movement, superior-inferior movement, and shape changes (curvature) of the tongue. However, B-mode has technical limitations in detecting tongue changes over time. While these limitations can be overcome by observing tongue images at various time points of the target posture and visualizing them in temporal sequence, this method may have limited effectiveness for extremely brief targets like flaps (Stone, 2005). This is because frame limitations may cause critical tongue movements to be missed when touching the alveolar ridge during flap articulation.

Recognizing these limitations, this study utilized M-Mode for posture analysis. While M-Mode can only view a single fixed scan line, making it difficult to grasp the overall movement or shape of the tongue, it faithfully shows changes over time along the scan line set during experimental design. This characteristic can help study [t] and [ɾ], articulated at similar positions.

This study set the scan line based on the highest point where the tongue tip touches the alveolar ridge. Through this, it was confirmed that [t] and [ʔ], which dominated Category A, and [ɾ], which dominated Categories B and C, had different articulatory characteristics. While [t] and [ɾ] showed the same directionality, [ʔ] showed different directionality, and [t] and [ɾ], which showed the same directionality, exhibited different curvature patterns (more rounded or flat shapes for [t], more pointed shapes for [ɾ]) in their durational distinctions observed in the acoustic aspect.

Consequently, this study used M-mode ultrasound imaging to identify how articulatory posture changes in intervocalic /t/ (especially [t] and [ɾ]) differ. From an articulatory perspective, this supports the hypothesis that "the same alternation occurs in the same structure."

doi:10.6342/NTU202500546

5.4. Comparison with Previous Studies on TSM

5.4.1. Focus of Previous Studies

As noted, earlier research on the intervocalic /t/ in TSM commonly analysed it alongside other stops (Chiang, 1992; Wang & Liu, 2010; Fon & Khoo 2025). These studies focus on its structural characteristics, suggesting that it undergoes voicing and gemination and ultimately articulates as [l]. However, there has been a significant lack of comprehensive understanding regarding its contextual variations and a considerable absence of acoustic or articulatory examination of this phenomenon.

5.4.2. Duration Differences

According to earlier studies on English and Spanish (Lavoie, 2001), the average duration of intervocalic [t] in Spanish lies between 97 ms (non-pre-stress) and 106 ms (pre-stress), while intervocalic [l] ranges from 57 ms (non-pre-stress) to 69 ms (pre-stress), and intervocalic [ɾ] lasts between 24 ms (non-pre-stress) and 29 ms (pre-stress). These findings should be compared with the boxplot in Figure 5.4. The statistical summaries for Categories A, B, and C are as follows: Category A has a mean of 0.134 s, a standard deviation of 0.0523 s, and a variance of 0.0027 s; Category B has a mean of 0.036 s, a standard deviation of 0.0152 s, and a variance of 0.0002 s; Category C reports a mean of 0.044 s, a standard deviation of 0.0204 s, and a variance of 0.0004 s. Although a direct comparison with the acoustic features of Spanish phonemes is not feasible, it is noteworthy that the mean duration of Category B, consisting solely of [ɾ], is closer to the duration of [ɾ] in Spanish than that of [l].

doi:10.6342/NTU202500546

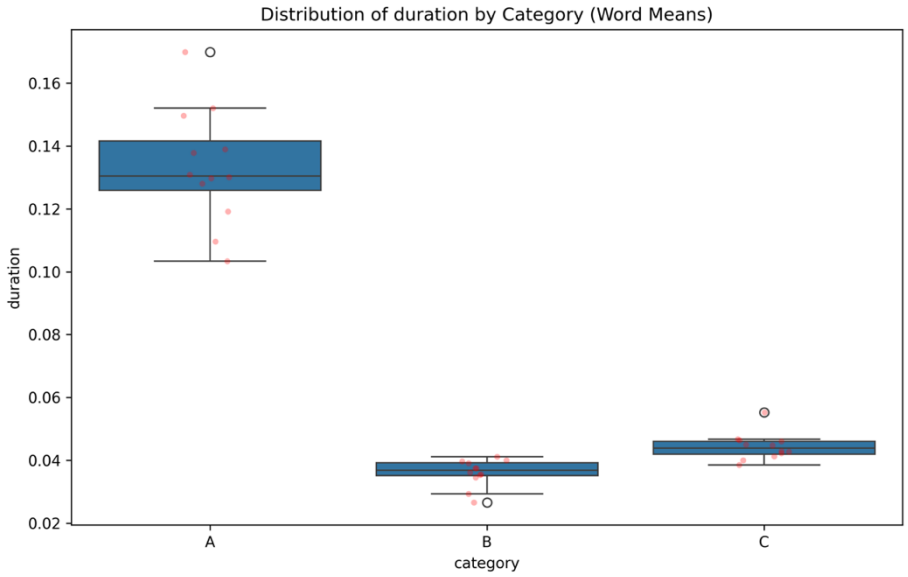
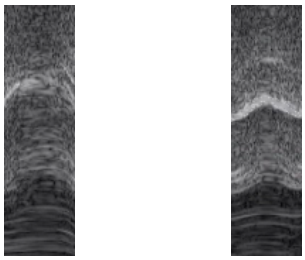


Figure 5.4: Raw Duration (in seconds) by Category.

5.4.3. Movement Pattern

To further examine the articulatory aspect, let's review the experimental phrase *'lau-let e ke-a'* (鬧熱的街仔) used in this study (Figure 5.5). The ultrasound images reveal that the [l] in "lɑu" displays a pattern similar to the [t], characterized by a rounded curve pointing upward. This indicates a prolonged tongue-to-alveolar contact time comparable to [t]. In contrast, the target sound of this study, [ɾ], exhibits a sharp peak shape, as previously described. This reflects the rapid tap of the tongue against the alveolar ridge when producing an intervocalic /t/ variant. These distinct patterns observed in the M-mode images provide additional evidence for determining whether this should be classified as [ɾ] or [l]. This suggests that while previous studies have discussed the realization of intervocalic /t/ as [l], there are also cases where it manifests as [ɾ].

doi:10.6342/NTU202500546



[l] [ɾ]

Figure 5.5: Differences Between [l] and [ɾ] in Ultrasound Imaging (Selected Images of 'lɑu-let e ke-a (鬧熱的街仔)' from W04)

The findings of this study from both acoustic and articulatory perspectives, along with the observation that intervocalic alternation occurs less frequently at word boundaries than within words, present an opportunity for further research into whether the intervocalic /t/ of TSM, situated between lexical and grammatical morphemes, reflects gemination (a form of strengthening) or flapping (a form of weakening). This is also a crucial reason why this paper defines the change in intervocalic /t/ as 'alternation'.

However, this study did not conduct a thorough acoustic or articulatory analysis to compare the variant defined as [ɾ] with the standard [l] pronunciation in TSM. Whether this variant is closer to [ɾ], closer to [l], or exhibits gradient characteristics remains an issue that requires further investigation. Future research should address this question by designing more systematic experiments and analyzing the results comprehensively.

5.5. Limitations and Future Directions

5.5.1. Definition of Duration

In this study, the duration of /t/ was measured from the end of the preceding vowel to the beginning of the following vowel (see Figure 3.4), following Herd et al. (2010). While Herd et al. (2010) was designed to compare the characteristics of /t/ and /d/ when

doi:10.6342/NTU202500546

realized as [ɾ], the primary objective of the present study was to measure duration to compare the acoustic properties of [t] and [ɾ]. A limitation of adopting the definition from the previous study is that certain acoustic features (particularly duration) may be overestimated, as [t] involves a closure period followed by a release before transitioning to the following vowel, whereas [ɾ] tends to blend more naturally with surrounding vowels.

In this regard, both acoustic and ultrasound data were collected, with the latter recorded in B/M-mode. B-mode imaging offers the advantage of visually identifying the moment when the tongue makes and breaks contact with the alveolar ridge. Therefore, in future research, if consistent criteria (e.g., from the end of the preceding vowel to the moment of release) were applied to both [t] and [ɾ] using B-mode imaging, more accurate results could be obtained in evaluating the acoustic characteristics of intervocalic /t/ variants in TSM.

5.5.2. Experimental Material Composition

5.5.2.1. Consideration of Vowels Before and After Intervocalic /t/

A primary limitation of this study concerns the composition of the experimental material. The main objective was to compare intervocalic /t/ similarities across different morphological contexts. To achieve this, three categories were established, and experimental materials were constructed using existing words and feasible combinations within each category. Four preceding vowels were also standardized across categories to enhance uniformity.

However, certain limitations arose when selecting actual words under these criteria. For all categories, the vowels preceding intervocalic /t/ were controlled to include three instances each of /a/, /e/, /i/, and /u/ (see Table 3.2). Nonetheless, it was not feasible

doi:10.6342/NTU202500546

to control the vowels following intervocalic /t/ uniformly across all categories. In Categories B and C, the following vowels were fixed as /a/ and /e/.

When selecting the experimental materials, I focused on using existing words (Categories A and B) or reasonable word combinations (Category C). As mentioned earlier, while the vowels preceding intervocalic /t/ were controlled to /a/, /e/, /i/, and /u/ with three items per vowel, Categories B and C yielded three or more suitable experimental items. However, Category A posed challenges in obtaining sufficient materials. For instance, when trying to control both preceding and following vowels as /a/, only one real word could be classified under Category A: *'at-am'* (遏泔; the act of preventing material from escaping when pouring soup). However, the extremely low frequency of this lexical item made it unsuitable as experimental material. Moreover, when fixing the preceding vowel as /a/ and the following vowel as /e/, no suitable words were found.

Therefore, the selection process for experimental materials in this study represents a necessary compromise to elicit natural pronunciation by using existing words. One potential approach to overcome these limitations in future research would be to conduct more in-depth studies with a narrower range of experimental materials.

Furthermore, in follow-up studies, modifying the surrounding vowels of intervocalic /t/ in Category A to two types (/a/ and /e/) and comparing them with Categories B and C, respectively, can reduce the noise that may arise from the surrounding vowels and yield statistically reliable results.

5.5.2.2. Word Frequency Impact

Some words had a notably low frequency of use (e.g., *kuã-let-a*; 寒熱仔), while others, though familiar in written form, were challenging for natural oral production due

doi:10.6342/NTU202500546

to limited use in spoken discourse (e.g., *t͡set-iok*; 節約). As prior research (Patterson & Connine, 2001; Kilbourn-Ceron & Goldrick, 2021) has shown, phenomena like flapping are significantly influenced by word frequency. Therefore, ensuring maximal consistency in word frequency for experimental materials in future studies would significantly enhance the reliability of the results.

5.5.2.3. Polyphones

The issue of experimental materials led to another problem caused by Chinese characters with multiple pronunciations. As mentioned earlier, in *'t͡set-iok'* (節約), the character '節' should be pronounced with the vowel [e], while in *'kuan-t͡sat iam'* (關節炎), the same character '節' should be pronounced with the vowel [a]. The fact that identical Chinese characters appearing in the same experimental materials had different pronunciations could potentially influence each other's pronunciation patterns. Indeed, one participant (W03) pronounced the vowel of '節' in '*kuan-t͡sat iam'* as [e]. In a situation where determining whether the vowel differences were due to mistakes or stemmed from the participants' natural pronunciation became difficult, it was unfortunate that acoustic characteristics could not be further subdivided by category based on preceding vowels, nor could ultrasound images be analysed by simply grouping the exact words. This limitation needs to be overcome in follow-up studies.

5.5.2.4. Phonetic Context Interference

Furthermore, one of Category C's experimental items, *'sẽ-lit e i-gi'* (生日的意義), presented a methodological challenge due to /l/ before the preceding vowel of the target, which obscured the clear vertical movement range of the tongue. However, while this

doi:10.6342/NTU202500546

limitation slightly affected the ultrasound image analysis, it had minimal impact on the acoustic analysis.

5.5.3. Reality of Intervocalic /t/ Variants

The experiment recruited participants based on basic conversation and self-assessment, which did not fully standardize individual language proficiency. Consequently, it was impossible to verify whether the small number of variant realizations observed in this study represent actual variations.

As shown in Figure 4.3, while the primary variants of intervocalic /t/ showed high consistency in categories B and C, category A exhibited a bifurcation in its main variants. Additionally, categories A and C displayed various variants beyond the primary variant(s), including variations (A_k, A_ɾ, A_p, B_k) that occurred in less than 1% of total tokens.

The non-uniform distribution of variations among participants has resulted in a wide distribution of acoustic characteristics, which made extracting average images during ultrasound image analysis complex. The low-frequency postures observed in specific individuals raise questions about whether they should be considered errors or legitimate variations.

Therefore, future research would benefit from:

1. Stricter control over participant selection

2. Establishment of rigorous criteria for excluding tokens from analysis

These measures would enable a more thorough analysis of ‘genuine’ variations.

doi:10.6342/NTU202500546

5.5.4. Image Similarity Analysis

5.5.4.1. Significant Similarity among Target Postures in Certain Participants

In this study, an experimental method that combines CoAtNet-2 and cosine similarity was employed to propose a quantitative approach for analyzing the similarity of ultrasound images. When this method was applied, the ranges of similarity varied among individual participants, suggesting that these differences may be due to variations in individual anatomical structures or articulatory processes strategies.

Upon examining the results (see Figures 4.5 and 4.6), small differences in similarity between [t] and [ɾ] were observed in some participants. For example, participant M03 exhibited a similarity of 0.901 between A_t and C_ɾ, and 0.905 between B_ɾ and C_ɾ, with a difference of only 0.004 recorded.

The minor difference in similarity of 0.004 between [t] and [ɾ] may raise questions from readers. To address these concerns, initial validation was carried out. As illustrated in Figure 3.8, even small numerical differences in similarity scores can indicate significant distinctions. When analyzing similarities based on extracted feature vectors, it's important to recognize that the differences between feature vectors obtained from M-mode images are generally smaller than those derived from conventional image similarity comparisons.

Furthermore, when comparing the high similarity between [t] and [ɾ] with previously reported acoustic analysis, it can be interpreted that for some participants, while the realization of intervocalic /t/ is similar, the overlap between /t/ and the vowels increases, which leads to increased voicing and alternation of intervocalic /t/.

For future research, the following approaches are recommended: From an acoustic perspective, focusing on analyzing the duration changes of preceding and following vowels; and from an articulatory perspective, either extending the time window of M-

doi:10.6342/NTU202500546

mode images (although this study utilized the maximum width supported by the ultrasound device) or analyzing postures from various angles by adjusting the M-mode scanning line position. Additionally, further refinements could include modifications to experimental materials (ensuring complete matching of preceding and following vowels) and implementing more detailed classifications (subcategorizing each category by phonemes). These modifications could lead to more universally acceptable results.

5.5.4.2. Difficulty in Integrated Analysis of All Experiment Participants

Making direct comparisons using original images from multiple participants is difficult. Generally, when using ultrasound equipment, while there is the advantage of seeing the tongue movements of experimental participants clearly, there is a slight disadvantage in that standardization is difficult due to differences in individuals' physical structures affecting their tongue's anterior-posterior and superior-inferior range of motion. In this study, with M-mode, when the detection line was designated, and the angle was well adjusted, it was confirmed that one could visually distinguish whether the tongue tip made contact with the alveolar ridge slowly to make a sound or quickly touched and passed by. However, even so, since the distance from the ultrasound probe to the tongue may vary among participants (Figure 5.6), if the images obtained through the experiment are cropped within the same range, they cannot be used as direct objects of comparison.

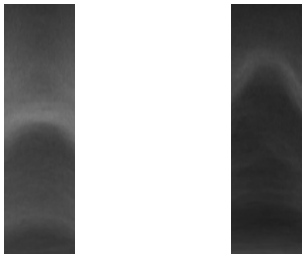


Figure 5.6: Tongue Height Differences for [t] in Category A.

doi:10.6342/NTU202500546

Of course, there could be a method to directly compare each individual's final calculated similarity values. However, as mentioned earlier, individual differences and various variables during image acquisition (such as contrast differences) may result in different similarity ranges when capturing ultrasound images. Regarding this issue, some improvement could be made in future research by appropriately adjusting factors such as gain according to each individual's situation during M-mode image acquisition while simultaneously standardizing the vertical reference points of the tongue when cropping images.

Furthermore, even within the same time window, depending on individual speech characteristics, the articulation of the corresponding phoneme may or may not have been completed. For example, looking at the articulation of [t] (Figure 5.7), within the same time window, M01 shows a gentle curve where the tongue has already made contact with the alveolar ridge and released. In contrast, M02 shows a straight line as the tongue is still in contact with the alveolar ridge. While this does not affect the comparison between representative postures for each Category, when developing future research, it is believed that more objective data analysis could be conducted by cropping images based on specific points of preceding and following vowels.

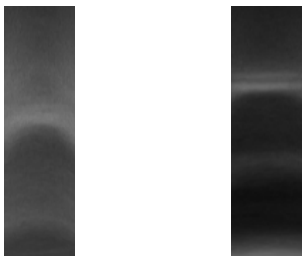


Figure 5.7: Differences in [t] Images from Category A.

doi:10.6342/NTU202500546

5.5.4.3. Blurring Caused by the Average Image

There is a limitation regarding the comparison of averaged images. In this study, to experimentally apply a new method of extracting feature vectors using a pre-trained model from M-mode images and comparing cosine similarity, average images for each Category_posture were used in the feature vector extraction stage. Figure 5.8 shows three Categories from W02. These three images show considerable similarity when visually inspected. However, in the similarity analysis, the similarity between A_k and A_ʔ was 0.565, A_k and B_k was 0.919, and between A_ʔ and B_k was 0.706. To analyse the reason for this, examining the distribution of these Category_postures revealed that A_k represented 1.7% (1 image), A_ʔ 86.4% (51 images), and B_k 6.8% (4 images) of the distribution.

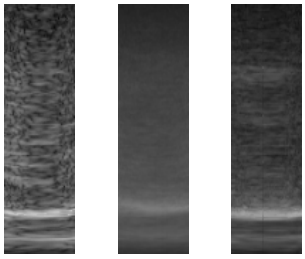


Figure 5.8: Differences between [k] and [ʔ] Images (W02).

This phenomenon seems linked to the average curvature observed when analyzing multiple averaged images, which resulted in simultaneous image blurring. This blurring effect affects the range of similarity measurements. During the analysis, a pattern emerges where blurred images show high similarity with other blurred images, and transparent images display high similarity with other clear images. In contrast, the similarity between blurred and clear images remains consistently low. To lessen this effect, regulating the number of images within the same categories or posture groups seems necessary.

doi:10.6342/NTU202500546

Of course, this study aimed to compare the similarity between the main postures of Categories A, B, and C – specifically [t] or [ʔ] and [ɾ] – and sufficient images were acquired to create average images for this purpose (although there were differences in quantity as seen in Figure 4.3). Therefore, there was no significant issue in analyzing the similarity between the main Category_postures of each Category. However, improvement in the analysis method seems necessary for more accurate analysis. The most intuitive approach would be to compare the similarity of all individual images per Category_posture and create a heatmap by averaging these similarities. However, conducting individual image analysis using the model employed in this study was found to be excessively time-consuming. Therefore, standardizing the number of images for all Category_postures during the experimental phase appears to be the most feasible method in follow-up research.

5.5.5. Additional Implications for Future Research

5.5.5.1. Comparison with Other Stops in TSM

As mentioned earlier, TSM intervocalic /t/ has been studied alongside other stop consonants such as /p/ and /k/, and has been considered to undergo gemination and voicing in specific environments (Chiang 1992; Fon & Khoo 2025). However, in light of the current research findings, TSM intervocalic /t/ may undergo weakening rather than strengthening in certain environments. Based on suggestions for future research, if more definitive evidence is found through acoustic and articulatory analyses that TSM intervocalic /t/ undergoes weakening rather than strengthening in specific environments, a comprehensive reanalysis may be necessary regarding how other TSM stop consonants are realized in intervocalic environments.

doi:10.6342/NTU202500546

5.5.5.2. Existence of Other Effects

Research has demonstrated that English flapping is influenced by word frequency as well as stress (Patterson & Connine, 2001; Kilbourn-Ceron & Goldrick, 2021). While this study aimed to exclude frequency effects when selecting experimental stimuli (except for those that had to be selected for vowel environment control, as previously mentioned), further research is necessary to examine the impact of frequency more thoroughly. Additionally, this represents a further task to be explored following the current study, which identified commonalities between intervocalic /t/ within words and at word boundaries in environments that share the same form.

5.5.5.3. Existence of Articulatory Variability

Derrick and Gick (2011) demonstrated that four variants of American English flap posture can be realized differently based on individual characteristics, indicating that flapping is not influenced solely by phonological context. Furthermore, Derrick and Schultz (2013) revealed that these four variants of American English flap posture can exhibit different acoustic characteristics, such as F1 or F2.

However, this study did not investigate whether different flap postures exist between category B and category C, or whether individuals employ different strategies for realizing postures. While it cannot be definitively stated whether posture diversity might be found in TSM flapping, given that its conditions differ from those of American English flapping, it would be meaningful to examine whether there are various variants of flap posture in TSM intervocalic /t/, as posture diversity can be investigated based on the methodology of previous research (Derrick et al., 2015).

doi:10.6342/NTU202500546

5.5.5.4. The influence of gender on the diversity of alternations

A noteworthy observation was that male participants exhibited less diversity in posture when pronouncing intervocalic /t/ compared to female participants. This finding is intriguing, particularly when contrasted with Herd et al. (2010), who reported that female speakers showed a higher flapping frequency than male speakers in acoustic and perceptual studies. As noted in 5.5.3, no objective standard is currently to determine whether the observed variants are genuine or incidental. While this study alone cannot verify the existence of these variants or whether the trends arise from characteristics of female speakers related to language change or individual linguistic abilities, Labov (2001) emphasized that children reconstruct language during acquisition, contributing to language change, with female children displaying this tendency more prominently. Therefore, it is essential to recognize that female participants in this study showed greater diversity in postures than their male counterparts, which warrants further investigation in future research.

doi:10.6342/NTU202500546

Chapter 6 Conclusion

TSM's intervocalic /t/ exhibits various phonetic variations when content and grammatical morphemes combine. This phonological process occurs across all generations and regions and is not limited to local dialects. While this phenomenon mainly occurs within words, this study aims to provide substantial and verifiable evidence of flapping phenomena across word boundaries, extending beyond single words.

To achieve this, both acoustic and articulatory methods were employed to examine how intervocalic /t/ is realized in TSM. This study explored the postures generated by each category by establishing morphologically similar and different contexts and broadening the analysis beyond just words to include word boundaries. When lexical and grammatical morphemes combine, flapping occurs more consistently within words. Although flapping is prevalent across word boundaries, there are instances where it does not happen, and the original phoneme is instead realized.

To summarize, duration and RMS amplitude were examined in terms of acoustics. Morphologically similar situations exhibited more comparable duration and RMS amplitude characteristics than morphologically different situations.

Flaps occurring in the same structural context may show statistical differences in acoustic characteristics depending on whether they are positioned within a word or at word boundaries. Conversely, this characteristic is difficult to detect in ultrasound image analysis, and it is anticipated that differences in acoustic characteristics, including the duration or RMS amplitude of word-internal flaps versus word-boundary flaps, warrant further in-depth research.

From an articulatory perspective, the comparison of postures, which is challenging to quantify, was analysed through similarity assessment using ultrasound M-

doi:10.6342/NTU202500546

Mode images, confirming that morphologically similar situations showed higher similarity values between postures than morphologically different situations.

The findings from previous studies, which suggest that TSM's intervocalic /t/ has characteristics of [l], show a slight divergence from the results of this study. However, this study indirectly revealed that the variant exhibits acoustic and articulatory features more akin to [ɾ]. More conclusive evidence is anticipated from future research that directly compares TSM's typical [l] with the intervocalic /t/ realized as [ɾ].

This study can be considered experimental, as it explores how recent developments in artificial intelligence can be applied to linguistics. Neural network models can be used alongside acoustic analysis methods and ultrasound-based articulation studies to extract features and analyse articulatory patterns.

Of course, there are clear limitations since the model and analysis methods adopted this time were not explicitly developed for analyzing ultrasound images. Future research is expected to be more reliable, as standardized image processing methods and modifications to the model to make more subtle distinctions suitable for black-and-white ultrasound images will be introduced.

doi:10.6342/NTU202500546

References

Boersma, P., & Weenink, D. (2024). *Praat: doing phonetics by computer* (Version 6.4.23) [Computer program]. Retrieved October 27, 2024, from http://www.praat.org/

Chiang, W.-Y. (1992). *The prosodic morphology and phonology of affixation in Taiwanese and other Chinese Languages* (Doctoral dissertation). University of Delaware.

Cho, H.-M. (2004). The phonology of flapping in American English. *The Jungang Journal of English Language & Literature*, *46*(1), 163-192.

Chong, A. J., & Garellek, M. (2023). Glottalisation of voiceless stops in Multicultural London English. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, 730, 2169-2173.

Dai, Z., Liu, H., Le, Q. V., & Tan, M. (2021). CoAtNet: Marrying Convolution and Attention for All Data Sizes. *arXiv preprint arXiv:2106.04803*. https://doi.org/10.48550/arXiv.2106.04803

De Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics, 26*, 283–310. https://doi.org/10.1006/jpho.1998.0077

Derrick, D., & Gick, B. (2011). Individual variation in English flaps and taps: A case of categorical phonetics. *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, *56*(3), 307–319. https://doi.org/10.1353/cjl.2011.0024

Derrick, D., & Schultz, B. (2013). Acoustic correlates of flaps in North American English. *Proceedings of Meetings on Acoustics*, *19*, 060260. https://doi.org/10.1121/1.4798779

doi:10.6342/NTU202500546

Derrick, D., Stavness, I., & Gick, B. (2015). Three speech sounds, one motor action: Evidence for speech-motor disparity from English flap production. *The Journal of the Acoustical Society of America, 137*(3), 1493–1502. http://dx.doi.org/10.1121/1.4906831

ELAN (Version 6.8) [Computer software]. (2024). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fon, J., & Khoo, H. L. (2025). *The phonetics of Taiwanese*. Cambridge University Press. https://doi.org/10.1017/9781009566834

Fukaya, T., & Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association, 35*(1), 45–58. DOI: 10.1017/S0025100305001891

Herd, W., Jongman, A., & Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics, 38*(4), 504–516. https://doi.org/10.1016/j.wocn.2010.06.003

Huffman, M. K. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics, 33*(3), 335–362.

Hwang, B., & Kang, S. (2018). A study of flaps in American English based on the Buckeye Corpus. *Phonetics and Speech Sciences, 10*(3), 9–18. https://doi.org/10.13064/KSSS.2018.10.3.009

Kahn, D. (1976). *Syllable-based generalizations in English phonology* (Doctoral dissertation). Massachusetts Institute of Technology.

Kilbourn-Ceron, O., & Goldrick, M. (2021). Variable pronunciations reveal dynamic intra-speaker variation in speech planning. *Psychonomic Bulletin & Review, 28*, 1365–1380. https://doi.org/10.3758/s13423-021-01886-0

doi:10.6342/NTU202500546

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, *10*(3), 421–441. http://www.jstor.org/stable/4178120

Kirchner, R. M. (1998). *An effort-based approach to consonant lenition* (Doctoral dissertation). University of California, Los Angeles.

Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2). Wiley-Blackwell.

Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages. Blackwell.

Lavoie, L. M. (1996). Consonant strength: Results of a data base development project. *Working Papers of the Cornell Phonetics Laboratory, 11*, 269–316.

Lavoie, L. M. (2001). *Consonant strength: Phonological patterns and phonetic manifestations (1st ed.)*. Routledge. https://doi.org/10.4324/9780203826423

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249–336. http://www.jstor.org/stable/4177987

Parker, S. G. (2002). *Quantifying the sonority hierarchy* (Doctoral dissertation). University of Massachusetts Amherst.

Patterson, D. & Connine, C. (2001). Variant Frequency in Flap Production. *Phonetica*, *58*(4), 254-275. https://doi.org/10.1159/000046178

Price, P. J. (1981). *A cross-linguistic study of flaps in Japanese and in American English* (Doctoral dissertation). University of Pennsylvania.

Roach, P. J. (1973). Glottalization of English /p/, /t/, /k/ and /tʃ/: A re-examination. *Journal of the International Phonetic Association, 3*(1), 10–21. https://doi.org/10.1017/S0025100300000633

Riehl, A. K. (2003). American English flapping: Evidence against paradigm uniformity

doi:10.6342/NTU202500546

with phonetic features. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS-15)* (pp. 2753–2756). Barcelona, Spain.

Seyfarth, S., & Garellek, M. (2020). Physical and phonological causes of coda /t/ glottalization in the mainstream American English of central Ohio. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(1), 1–33. https://doi.org/10.5334/labphon.213

Shport, I. A., Johnson, G., & Herd, W. (2018). Flapping before a stressed vowel-Whatever! *Proceedings of Meetings on Acoustics, 31*, 060004. https://doi.org/10.1121/2.0000823

Steriade, D. (2000). Paradigm uniformity and the phonetics/phonology boundary. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 313–334). Cambridge University Press.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, *19*(6–7), 455–501. https://doi.org/10.1080/02699200500113558

Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *The Journal of the Acoustical Society of America, 125*(5), 3317–3327. https://doi.org/10.1121/1.3097773

Wang, H. S., & Liu, H.-C. J. (2010). The morphologization of liaison consonants in Taiwan Min and Taiwan Hakka. *Language and Linguistics, 11*(1), 1–20.

Wightman, R. (2019). *PyTorch Image Models* [Computer software]. GitHub. https://github.com/rwightman/pytorch-image-models. https://doi.org/10.5281/zenodo.4414861

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America, 66*(4), 1039–1050. https://doi.org/10.1121/1.383323

doi:10.6342/NTU202500546

Appendix A – Participant Information

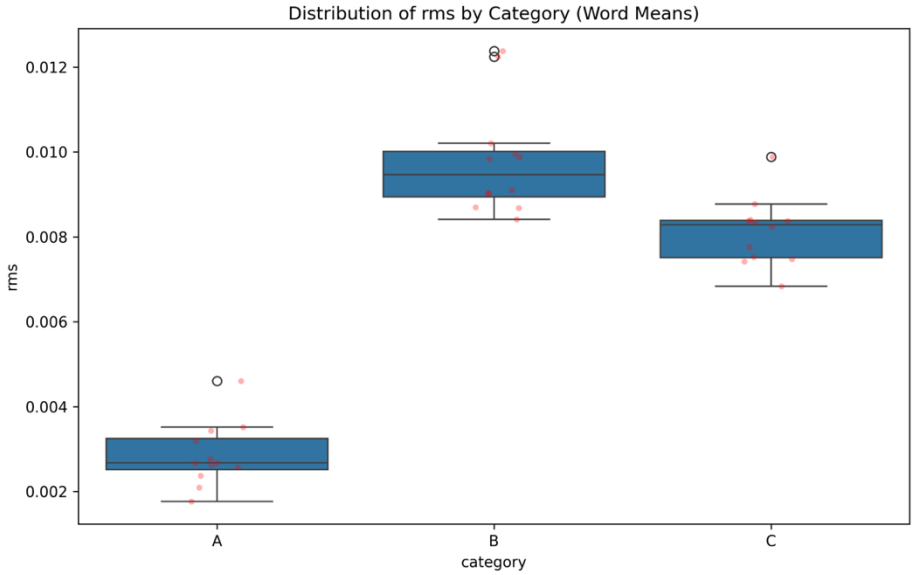
A.1. Participant Information

Code	Sex	Age	Hometown
M01	Male	20	Kaohsiung
M02	Male	21	Keelung
M03	Male	18	Taichung
M04	Male	21	New Taipei
W01	Female	20	Tainan
W02	Female	30	Taipei
W03	Female	32	New Taipei
W04	Female	40	Taipei

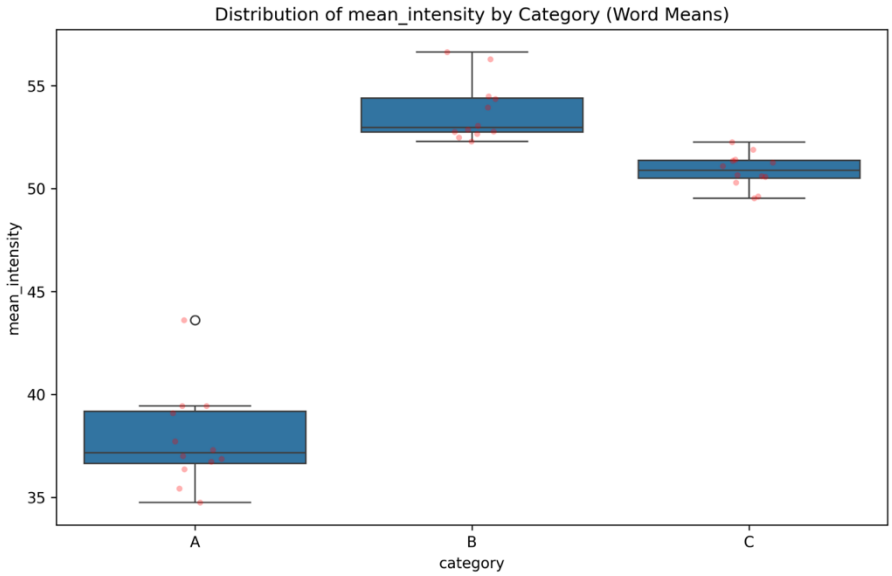
doi:10.6342/NTU202500546

Appendix B – Additional Figures and Tables

B.1. Raw RMS amplitude by Category



B.2. Raw Intensity (in dB) by Category



doi:10.6342/NTU202500546

B.3. Statistical Analysis of Raw Duration (ANOVA and Tukey)

ANOVA

Feature	Group_By	F-Statistic	df1	df2	P-Value
duration	category	278.36	2	33	$p < 0.0001$

Tukey

Group1	Group2	Mean_Difference	Lower_CI	Upper_CI	P-Value	Reject_Null
A	B	-0.0972	-0.1084	-0.086	$p < 0.0001$	TRUE
A	C	-0.0889	-0.1002	-0.0777	$p < 0.0001$	TRUE
B	C	0.0083	-0.0029	0.0195	$p = 0.1808$	FALSE

B.4. Statistical Analysis of Raw RMS Amplitude (ANOVA and Tukey)

ANOVA

Feature	Group_By	F-Statistic	df1	df2	P-Value
rms	category	162.38	2	33	$p < 0.0001$

Tukey

Group1	Group2	Mean_Difference	Lower_CI	Upper_CI	P-Value	Reject_Null
A	B	0.0069	0.0059	0.0079	$p < 0.0001$	TRUE
A	C	0.0053	0.0043	0.0062	$p < 0.0001$	TRUE
B	C	-0.0017	-0.0027	-0.0007	$p = 0.0006$	TRUE

B.5. Statistical Analysis of Raw Intensity (ANOVA and Tukey)

ANOVA

Feature	Group_By	F-Statistic	df1	df2	P-Value
mean_intensity	category	309.42	2	33	$p < 0.0001$

Tukey

Group1	Group2	Mean_Difference	Lower_CI	Upper_CI	P-Value	Reject_Null
A	B	15.884	14.2126	17.5554	$p < 0.0001$	TRUE
A	C	13.0519	11.3805	14.7233	$p < 0.0001$	TRUE
B	C	-2.8322	-4.5036	-1.1608	$p = 0.0006$	TRUE

doi:10.6342/NTU202500546