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 Arts

National Taiwan University

Master's Thesis

年齡對臺灣華語元音聲調產製中的舌骨喉頭接近動作之影響

Age-related Effects on Hyolaryngeal Approximation during Taiwan Mandarin Vowel-Tone Production

姚勇瑞

Yung-Jui Yao

指導教授：邱振豪 博士

Advisor: Chenhao Chiu, PhD

共同指導教授：蕭名彥 醫師

Co-Advisor: Ming-Yen Hsiao, MD, PhD

中華民國 114 年 7 月

July 2025

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

Acknowledgements

很高興能夠順利完成研究並從語言學研究所畢業。首先，想感謝家人與兩位指導老師。感謝家人在就學期間的支持與鼓勵使我能專注於學業，以及兩位指導老師在實驗與論文撰寫上的專業指導與啟發，使得研究能夠更深入、嚴謹與完整。同時，也感謝語言所的師長與同學，在研究歷程中所給予的協助與建議。最後，特別感謝實驗室成員莊詠甯同學的大力協助，使實驗執行過程更加順利。

感謝你們，使這項研究得以完成。

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

摘要

在語音學中，舌骨與甲狀軟骨的交互作用相對較少被討論，然而在醫學領域中，兩者因在吞嚥與吞嚥障礙中扮演關鍵角色而備受重視。本研究探討語音產製過程中舌骨喉頭接近動作的時間變化，並檢視此現象是否會隨年齡和母音與聲調環境而有所不同。本研究招募了無吞嚥障礙、流利使用臺灣華語的年輕成人與年長者作為受試者。研究中運用超音波影像與廣義加成模型來分析語音產出中的舌骨喉頭接近動作，並以表面肌電圖來對肌肉活化模式進行定性分析。超音波結果顯示語音產製中存在舌骨喉頭接近動作，且其表現出年齡差異與母音與聲調環境的影響；年長者表現出較弱的舌骨喉頭接近程度。這些差異可能受到老化與不同世代間的聲調變異所影響。肌電圖的結果顯現年輕成人受試者的肌肉活化模式具有聲調與母音特異性，而年長者的結果則暗示其舌部姿態的啟動時間可能早於年輕成人受試者。本研究結果為了解語音產製中的舌骨喉頭動態提供了初步的認識，並可能有助於未來與吞嚥過程中舌骨喉頭動態的比較研究。

關鍵字: 舌骨喉頭接近動作、母音產製、聲調產製、臺灣華語、年齡變異

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

Abstract

The interplay between the hyoid bone and thyroid cartilage has been relatively underexplored in phonetics despite receiving significant attention in medicine due to their critical roles in deglutition and dysphagia. This study investigated hyolaryngeal approximation changes over time during speech production and examined whether it varies across ages and vowel-tone contexts. Young adult and elderly non-dysphagic fluent Taiwan Mandarin speakers were recruited. Ultrasound imaging with generalized additive mixed modeling was used to characterize hyolaryngeal approximation during speech while surface electromyography (EMG) was applied to qualitatively analyze muscle activation patterns. Ultrasound results revealed hyolaryngeal approximation during speech, along with age-related differences modulated by vowel-tone contexts; elderly speakers exhibited more reduced approximation. These differences are likely influenced by aging and cross-generational tonal variations. The EMG results showed tone- and vowel-specific patterns for the young adult speakers while the results for the elderly speakers suggested earlier activation for lingual gestures compared to the young adult speakers. These findings represent a preliminary step to understanding hyolaryngeal dynamics during speech production and may help inform future comparisons with hyolaryngeal dynamics in swallowing.

doi:10.6342/NTU202503180

Keywords: hyolaryngeal approximation, vowel production, tone production, Taiwan Mandarin, age variation

doi:10.6342/NTU202503180

Contents

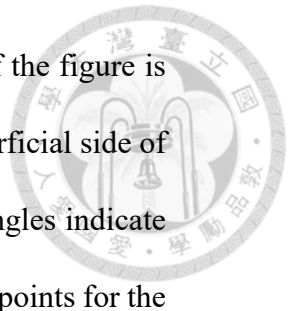
doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

List of Figures

doi:10.6342/NTU202503180



doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

List of Tables

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

Chapter 1 Introduction

Speech production involves coordinated movements of various articulators to accomplish specific speech targets. Among these articulators, the roles of the hyoid bone and larynx have been extensively explored (Ewan & Krones, 1974; Honda, 1983; Menon & Shearer, 1971; Moisik et al., 2014; Poh, 2024). The movements of the hyoid bone and larynx are linked to both lingual articulation and pitch control. For instance, the hyoid is typically displaced in a supero-anterior direction during the production of higher pitch, and high vowels are associated with a more anterior hyoid position (Honda, 1983). Meanwhile, vertical movements of the larynx are commonly assessed by tracking the position of the thyroid cartilage. Laryngeal height generally follows the pitch contour, with notable exceptions (Honda, 1995; Honda et al., 1999; Moisik et al., 2014). Beyond pitch, laryngeal height is also influenced by segmental properties; for example, rounded vowels are generally associated with a lower laryngeal position (Ewan & Krones, 1974). Nevertheless, the dynamic interaction between these two structures remains underexplored in the realms of speech.

In medicine, the approximation of the hyoid bone and larynx, referred to as hyolaryngeal approximation, has received attention for its critical physiological role in airway protection during the pharyngeal phase of swallowing, when the bolus is transported to the esophagus

doi:10.6342/NTU202503180

through the pharynx, the shared pathway of the alimentary and respiratory tracts (Ambrosi & Lee, 2021; Leonard et al., 2000). Crucially, a reduced extent of hyolaryngeal approximation during swallowing has been observed in individuals with dysphagia, and in healthy older adults specifically during effortful swallows (Bahia & Lowell, 2023; Y.-L. Huang et al., 2009; Matsuo & Matsuyama, 2021). Aging may lead to reductions in muscle strength and speed, which can modulate the temporal and spatial features of swallowing behavior (Bahia & Lowell, 2023; Groher, 2016; Kang et al., 2010; Kim & McCullough, 2008; Logemann et al., 2000, 2002; Robbins, 1996), potentially altering the extent and coordination of hyolaryngeal approximation.

Hyolaryngeal approximation is not only vital in deglutition but may also play a role in speech production. Previous research has indicated a shared motor and potential neurological basis underlying both functions. Hiiemae and Palmer (2003) examined anatomical evidence and lingual postures shared across speech and feeding. Their finding suggested a "kinetic chain" linking the jaw, hyoid and tongue serving as a basis for both speech and feeding. The shared musculature between speech and deglutition may also support the notion of a linked low-level motor control between two tasks despite the fact that a growing number of scientific publications may point to a shared neuromuscular network at the higher level (Castillo-Allendes et al., 2025; McFarland, 2006).

A number of studies have examined hyolaryngeal approximation in various phonation tasks (Carlson, 2021; Miller et al., 2012; Miloro et al., 2014). However, these studies primarily focused on single time points during the non-linguistic production tasks and have reported varied evidence regarding the presence of hyolaryngeal approximation. To date, its temporal dynamics and behavior in natural linguistic contexts have received limited attention. More-

doi:10.6342/NTU202503180

over, age-related variation in hyolaryngeal approximation remains largely unexplored in linguistic tasks.

This study aims to investigate whether the extent of hyolaryngeal approximation and associated muscle activation patterns vary across age groups and linguistic contexts. Based on the hypothesis that there is a shared neuromuscular basis, it is expected that hyolaryngeal approximation occurs during speech production and the extent of hyolaryngeal approximation would be reduced in older speakers compared to the younger speakers, especially in more demanding contexts.

To investigate these questions, fluent speakers of Taiwan Mandarin, a tonal language, were recruited. The use of lexical tones in tonal languages provides better control over the pitch contour during production. The production targets in this study were monosyllables consisting of [a], [i], and [u] produced in the four lexical tones in Taiwan Mandarin. Hyolaryngeal approximation was recorded with a hand-held linear probe and analyzed with generalized additive mixed models. Surface electromyography was used to record muscle activation patterns and analyzed qualitatively via visual inspection.

This study contributes to the growing body of literature on hyolaryngeal gestures during speech production. Moreover, it examines age-related variations of these gestures, drawing on insights from the deglutition literature, where age-related changes in hyolaryngeal function are more extensively documented. In addition, this study represents a preliminary step towards characterizing hyolaryngeal dynamics, with the view to enabling future comparisons with deglutition and supporting the potential use of speech gestures as an adjunct tool in screening or clinical contexts.

The remainder of this thesis is organized as follows. Chapter 2 reviews previous research

doi:10.6342/NTU202503180

on the roles of the hyoid and larynx in speech production and deglutition, as well as relevant background on tonal patterns in Taiwan Mandarin. Chapter 3 outlines the methodology, including participant recruitment, data collection procedures, and analytical approaches. Chapter 4 presents the results of the ultrasound and surface electromyography analyses. Chapter 5 discusses the results in relation with previous studies. Finally, Chapter 6 concludes the thesis with a summary, limitations, and suggestions for future research.

doi:10.6342/NTU202503180

Chapter 2 Literature Review

2.1 An Overview of the Anatomy of the Larynx and Hyoid

The literature provides a concise overview of the anatomy of the larynx and hyoid, as described in Gick et al. (2013), Honda (1995), and Moore et al. (2018). These structures and their associated musculature play crucial roles in speech production and deglutition, and are central to the current investigation.

The human larynx is composed primarily of cartilage, with the thyroid cartilage occupying substantial portions of its structure. Below the thyroid cartilage is the cricoid cartilage. Two arytenoid cartilages sit on the posterior plate of the cricoid. The vocal folds contain the vocalis muscle and vocal ligaments and are connected to the thyroid notch anteriorly and the vocal processes of the arytenoid cartilages posteriorly.

The muscles of the larynx are typically grouped into intrinsic and extrinsic laryngeal muscles. The intrinsic muscles are those that connect between laryngeal cartilages while the extrinsic ones are those provide support from the outside.

With regard to the functions of the intrinsic muscles, the vocalis muscle alters the tension of the vocal folds, while the cricothyroid muscle plays an important role in increasing pitch by rotating the thyroid cartilage, tightening the vocal folds. The lateral and posterior

doi:10.6342/NTU202503180

cricoarytenoid muscles adduct and abduct the vocal folds, respectively (Figure 2.1).

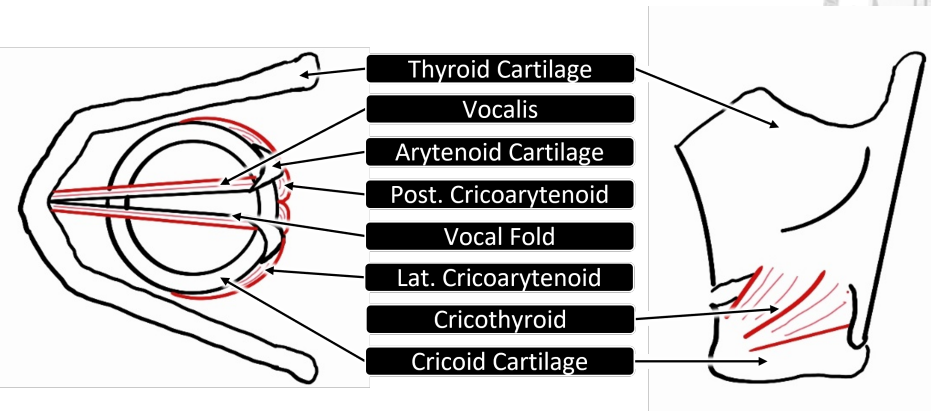


Figure 2.1: A simplified hand-drawn diagram showing intrinsic laryngeal muscles, based on Gick et al. (2013) and Moore et al. (2018). The diagram one the left shows the top view of the larynx and the diagram on the right shows the lateral view. The left sides of both diagrams are the anterior side of the human body.

The extrinsic laryngeal muscles include the infrahyoid and suprahyoid muscles. Specifically, the infrahyoid muscles are the sternohyoid, omohyoid, sternothyroid, and thyrohyoid muscles, whereas the suprahyoid muscles consist of the mylohyoid, geniohyoid, stylohyoid, and digastric muscles (Figure 2.2).

In general, the suprahyoid muscles may elevate the hyoid and, indirectly, the larynx. In contrast, the infrahyoid muscles depress the larynx, with thyrohyoid being a notable exception, which elevates the larynx. Other extrinsic muscles include the stylopharyngeus which also elevates the larynx.

Closely linked to the larynx both anatomically and functionally, the hyoid bone is not connected to any other bones, but is suspended approximately at the level of the third cervical vertebra and serves as a crucial anchor for many extrinsic laryngeal muscles, except for the sternothyroid muscle. This sets its role in coordinating laryngeal movements. In addition, the hyoid provides structural support to the tongue and facilitates its movements. The mylohyoid

doi:10.6342/NTU202503180

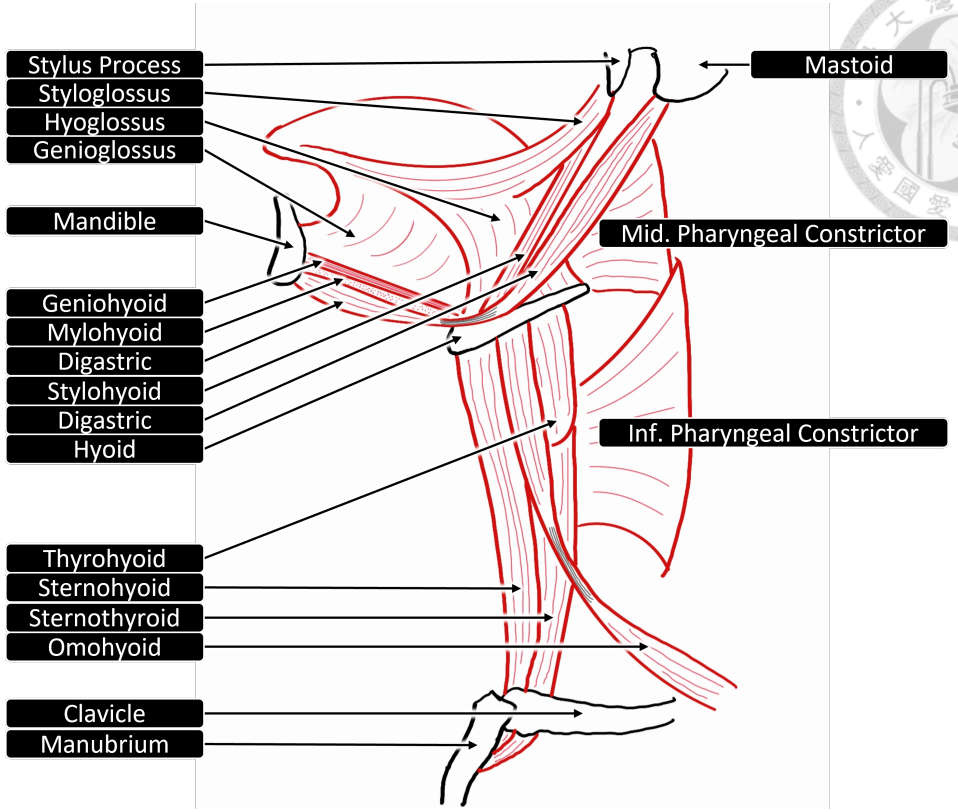


Figure 2.2: A simplified hand-drawn side-view diagram showing muscles attached to the larynx and hyoid, based on Gick et al. (2013) and Moore et al. (2018). The left side of the diagram is the anterior side of the human body.

muscle, a suprahyoid muscle, stiffens the floor of the mouth and elevates the hyoid and tongue. The genioglossus and hyoglossus muscles are also attached to the hyoid bone. The posterior fibers of the genioglossus muscle may pull the tongue root and hyoid bone forward, squeezing the tongue superiorly. A summary of the musculature related to larynx and hyoid movements can be found in Table 2.1.

The movements of larynx and hyoid have been found to have correlates in various language production targets as well as important functions in deglutition. The following sections provide a review on laryngeal and hyoid movements in speech production and deglutition.

doi:10.6342/NTU202503180

Function	Agonist Muscles
Larynx elevation	suprahyoid muscles, thyrohyoid
Larynx depression	sternohyoid, sternothyroid, omohyoid
Larynx anterior tilt	cricothyroid
Vocal cord shortening	vocalis
Vocal cord adduction	lateral cricoarytenoid
Vocal cord abduction	posterior cricoarytenoid
Hyoid elevation	suprahyoid muscles
Hyoid advancement	geniohyoid, genioglossus
Hyoid depression	infrahyoid muscles
Muscle Groups:	
Suprahyoid muscles	mylohyoid, geniohyoid, stylohyoid, digastric
Infrahyoid muscles	sternohyoid, sternothyroid, thyrohyoid, omohyoid

Table 2.1: Summary of Musculature Affecting the Larynx and Hyoid, based on Gick et al. (2013) and Moore et al. (2018).

2.2 Laryngeal and Hyoid Movements in Speech Production

2.2.1 Laryngeal Movements in Speech Production

We shall begin by outlining laryngeal movements involved in speech production. Laryngeal movements have been relatively well-documented in the articulatory literature. Several studies have revealed laryngeal height and state contributes to pitch control, register, and vowel qualities.

Moisik et al. (2014) tracked laryngeal height changes during Mandarin tone production with ultrasound and laryngoscopy to record the state of the larynx. Their results revealed that the height of the thyroid and pitch contours, in general, showed a similar trend in Tones 2 and 4. However, there could be gradual rising of thyroid when producing Tone 1, which was interpreted as a strategy to compensate for the gradual drop in subglottal pressure during production. Tone 3, on the other hand, could be realized with two distinct gestures, laryngeal lowering and laryngeal elevation with constriction.

doi:10.6342/NTU202503180

It has been widely observed that laryngeal height substantially correlates with F0 (Honda, 1995). When the larynx is lowered, the posterior plate of the cricoid cartilage remains parallel to the tangent to the anterior outline of the lordotic cervical vertebrae. This leads to a rotation on the cricothyroid joint in favor of shortening the vocal cords, resulting in a lower F0 (Honda et al., 1999). The larynx depressors, the sternohyoid and, less consistently, the sternothyroid muscle, have also been implicated in lower F0 production (Honda, 1995), as shown in previous research on Mandarin (Boysson-Bardies et al., 1986), Thai (Erickson, 1993), and English (Atkinson, 1978). However, some evidence suggests that the sternothyroid may also contribute to F0 elevation (Niimi et al., 1988). Beyond pitch modulation, the sternohyoid also appears to participate in jaw opening and tongue retraction (Honda, 1995).

Honda et al. (1999) also explored how vertical movement of the larynx would contribute to high F0 production. However, the effects appeared to be relatively modest. Despite the well-known phenomenon where larynx height is positively associated with F0, there does not seem to be a convincing explanation (Honda, 1995). Aside from vertical laryngeal movements, cricopharyngeus, an intrinsic laryngeal muscle may contract to rotate the cricothyroid joint, lengthening and tensing the vocal cords, leading to an increased F0 (Gick et al., 2013; Honda, 1995; Moore et al., 2018). As for the extrinsic laryngeal muscles associated with pitch elevation, it has been observed that the thyrohyoid muscle activation shares a similar contour to that of F0 (Sawashima et al., 1973).

Medialization of the thyroid cartilage was also observed on ultrasound imaging in Singaporean Mandarin (Poh, 2024). The phenomenon was reported to be partially related to lower F0 and possibly related to laryngeal constriction. Horizontal movement and rotation of the larynx were suggested as other factors contributing to medialization on ultrasound image.

doi:10.6342/NTU202503180

However, certain patterns of medialization observed in Singaporean Mandarin tone production remains to be understood.

The other noteworthy finding by Moisik et al. (2014) was that low pitch, aside from laryngeal lowering, could be achieved with a constricted, raised larynx. It was argued that this laryngeal configuration would allow contact of the ventricular and vocal folds, which would increase the vibrating mass and bring about changes in vibration mode, resulting in a lower F0 and a creaky voice, a feature present in low pitch regions in Mandarin tones, including Taiwan Mandarin (Kuang, 2017). On the contrary, in breathy phonation, the larynx moves downwards, separating the vocal cords more apart, resulting in less contact and possibly breathy phonation (Gick et al., 2013; Moisik et al., 2014). These phonatary types could be integrated into the tone system as observed in languages such as Bai (Tibeto-Burman) (Edmondson & Esling, 2006). Nevertheless, current literature does not suggest that register is a necessary target of Taiwan Mandarin tones.

Aside from pitch, laryngeal position is also dependent on segment properties. In general, rounded vowels are associated with lower laryngeal position. Evidence suggests that the lowering of larynx lengthens the vocal tract reducing the overall formant frequency in cooperation with lip rounding (Dusan, 2007; Ewan & Krones, 1974; Hoole & Kroos, 1998; Riordan, 1977). Consonants also affect laryngeal height. Others being equal, aspirated and dental stops, in general, were associated with a higher laryngeal position compared to non-aspirated and bilabial stops, respectively (Ewan & Krones, 1974).

doi:10.6342/NTU202503180

2.2.2 Hyoid Movements in Speech Production

The hyoid is located at the base of the tongue and has connections to numerous muscles. In contrast to descriptions of the thyroid during speech production, literature specifically describing hyoid movements has been relatively sparse.

Notably, Honda (1983) used electromyography (EMG) and an optical tracking system to record muscle activity and changes in hyoid position for different vowels produced with different pitch contours over two-mora non-words, consisted of combinations of /a/ and /i/ flanked either with /m/ or not. The subject was one Japanese speaker and the pitch contours were either flat, raising and falling. Regarding horizontal hyoid movements, the results showed that hyoid movements generally followed pitch contours, with a higher pitch associated with a more anterior position. Meanwhile, the hyoid body had an anterior position during /i/ production and a posterior position with /a/ production. It was argued that the hyoid was anteriorly displaced by the contraction of the posterior fibers of the genioglossus, advancing the tongue root.

With regard to vertical movement, the hyoid can be elevated with a rising pitch contour. However, it did not consistently depress with a falling contour. /a/ was associated with a more superior hyoid position in comparison with /i/. It was suggested that when pronouncing /i/, the tongue increased in height but the base area of the tongue reduced, leading to an increased downward force per unit area. Vertical hyoid movement was larger for vowel quality changes than for pitch changes.

Menon and Shearer (1971) included vowels [i], [u], [ɑ] in the analysis. Participants were instructed to produce syllable sequences consisting of three sets of CV syllables; each repeated

doi:10.6342/NTU202503180

7 times. The consonant was [p] and the vowel was one of the three target vowels. CV sets were pronounced in three orders. The results showed that hyoid position varied across vowels, [i], [u], [ɑ], ordered from anterior to posterior and inferior to superior, respectively. Moreover, the findings indicated that the hyoid position for a given vowel could be influenced by the neighboring vowels, suggesting coarticulatory effects in syllable sequences.

2.2.3 The Interaction of Hyoid and Laryngeal Postures in Speech Production

Much of the prior linguistic research on hyoid-laryngeal interaction in speech has centered around intrinsic F0, a prevalent phenomenon where a higher fundamental frequency is associated with a high vowel compared to low vowels (Whalen & Levitt, 1995). A common explanation to this phenomenon is that the slight changes in pitch arise from the pulling force from the tongue. This is referred to as the "tongue-pull" hypothesis. When the tongue is raised along with the hyoid, which in turn pulls the larynx upwards or rotates the thyroid cartilage and tightens the vocal cords, leading to a higher pitch. (Honda, 1983; Ladefoged, 1968).

Aside from the intrinsic F0, which focuses on presumably universal inherent pitch alterations, research suggested that hyoid/lingual gestures may also interact with the larynx in lexical tone productions. Shaw et al. (2016) showed that for tones starting with a low pitch (Tones 2 and 3), the tongue body was lowered for vowel /a/ but elevated for vowel /i/, in comparison to tones starting with a high pitch (Tones 1 and 4). It was suggested that the two low tone-producing strategies reported in Moisik et al. (2014), namely laryngeal lowering and elevation with constriction (discussed in Section 2.2.1), could have contributed to the observed difference in Shaw et al. (2016) and that the preferred laryngeal gestures may vary across

doi:10.6342/NTU202503180

different vowel contexts.

Focusing onto hyolaryngeal approximation, this aspect of speech production has been less explored in the linguistic context. Notably, Carlson (2021) examined whether the thyrohyoid space[1] could be reliably measured using ultrasound in vocally healthy individuals, with the view of further studies on individuals with muscle tension dysphonia. Participants from two age groups were recruited, the mean ages being 27.5 and 54.25 respectively. Participants sustained /i/ at habitual pitch and loudness, falsetto, low pitch, high pitch, loud voice, and soft voice, each repeated three times. Sonography was done in the supine position. Ultrasound images were captured approximately at the three-second mark of sustained phonation. The thyrohyoid space was determined as the distance between the acoustic shadows of the two structures. Absolute thyrohyoid space and percent reduction of thyrohyoid space[2] were compared between age groups and genders. Absolute thyrohyoid space was generally greater in the younger group except for resting and low pitch productions. For the percentage of thyrohyoid space reduction, only the female group in low pitch /i/ showed significantly less reduction compared to the male participants. On the other hand, for the percentage change of thyrohyoid space, all comparisons were insignificant except for the low /i/ condition where females showed less reduction of thyrohyoid space.

Two other MRI studies have examined hyolaryngeal approximation, measured between the hyoid bone and the larynx, during effortful pitch glide (Miloro et al., 2014) and humming (Miller et al., 2012). Miloro et al. (2014) examined hyolaryngeal approximation during effortful pitch glide and compared it with a repeated swallow task. The effortful pitch glide was

[1]The measuring method was different from the present investigation. Thus the term used in the Carlson (2021) was used.

[2]Referred to as "relative laryngeal reduction" in the article.

doi:10.6342/NTU202503180

performed by asking participants to produce /i/ starting from a modal voice, gradually elevating the pitch, and ending with a forceful /i/ at their highest pitch. The maneuver elicited an increased approximation (0.27 ± 0.28 cm) while swallowing did not (-0.17 ± 0.5 cm). Miller et al. (2012) measured hyolaryngeal distance during resting, low pitch target humming and high pitch target humming and found no significant differences. A reduction in hyolaryngeal distance appeared to occur progressively from the resting state to low-pitch and then high-pitch targets (Miller et al. (2012), Table 3).

These prior studies, while informative, were limited to sustained phonation or non-linguistic tasks and typically lacked temporal data. Therefore, it may not capture the full range of articulatory dynamics involved in linguistically driven gestures. A summary of the laryngeal and hyoid movements invovlved in speech production is presented in Table 2.2.

	Patterns
Larynx	Laryngeal height follows F0 Laryngeal height is higher for [i], [a] compared to [u]
	Alternatives: Tone 1: elevated to compensate for dropping subglottal pressure Tone 3: alternatively raised and constricted to achieve a low F0
Hyoid	More supero-anterior for higher pitch More anterior for high vowels
Hyoid-Larynx	Mainly focused on intrinsic F0 Inconclusive results with phonation tasks

Table 2.2: A summary of laryngeal and hyoid movements in speech production, based on Carlson (2021), Ewan and Krones (1974), Honda (1983, 1995), Menon and Shearer (1971), Miller et al. (2012), Miloro et al. (2014), and Moisik et al. (2014)

doi:10.6342/NTU202503180

2.3 Hyoid and Laryngeal Movements in Deglutition

Aside from speech production, the muscles connected to the hyoid and larynx also play an important role in deglutition and growing evidence suggests overlapping in neuromuscular components of the two functions. This section provides a concise introduction to the physiology and age-related changes of deglutition. The connection between speech production and deglutition is reviewed in the following section (Section 2.4)

2.3.1 An Overview of the Physiology of Deglutition

Traditionally, deglutition is classified into three stages: the oral, pharyngeal, and esophageal phases (Ambrosi & Lee, 2021; Palmer & Hiiemae, 1997). In the oral phase, food is masticated and formed into a bolus, which is then propelled into the pharynx. Numerous protective mechanisms are involved in the pharyngeal phase to reduce the risk of the bolus entering the nasopharynx and larynx. The esophageal phase denotes the passage of the bolus through the esophagus into the stomach.

Relevant to the present investigation, during pharyngeal swallow, the hyoid bone and the thyroid cartilage, together known as the hyolaryngeal complex, are displaced in a superoanterior direction by the action of the suprahyoid and thyrohyoid muscles. This positions the hyolaryngeal complex beneath the tongue root, which leads to posterior tilting of the epiglottis, covering the laryngeal inlet. The upper esophageal sphincter also opens to permit bolus entry into the esophagus. Other protective strategies include vocal fold adduction and velopharyngeal port closure.

If the airway is not sufficiently protected, the bolus may enter the larynx during degluti-

doi:10.6342/NTU202503180

tion. “Laryngeal penetration” occurs when the bolus enters the larynx above the vocal cords while “aspiration” arises when the bolus passes through the vocal cords. Penetration and microscopic amounts of aspiration may occur in healthy individuals, while visible aspiration on videofluoroscopy or endoscopy is pathological and linked to pneumonia or airway obstruction.

2.3.2 Hyolaryngeal Approximation in Deglutition

As mentioned above, the hyolaryngeal complex moves supero-anteriorly during the pharyngeal phase of deglutition (Ambrosi & Lee, 2021). Aside from the overall movement of the complex, the distance between the hyoid and larynx is also reduced during a normal swallow, as revealed by a videofluoroscopic study on swallow in normal adults. This maneuver was considered to be related to airway protection (Leonard et al., 2000). Decreased hyolaryngeal approximation thus may be associated with reduced airway protection during deglutition.

Kuhl et al. (2003) first applied hyolaryngeal approximation with ultrasound to measure laryngeal elevation and examined patients with different conditions. Thereafter, with a curvilinear probe placed on the throat midline, Y.-L. Huang et al. (2009) measured the hyolaryngeal distance during swallow in normal individuals, non-dysphagic stroke patients, and dysphagic stroke patients. They verified the intra- and inter-rater reliability of this method and reported a decrease in hyolaryngeal distance in dysphagic stroke patients compared to normal and non-dysphagic stroke patients. Alternatively, Matsuo and Matsuyama (2021) calculated the ratio of hyoid bone displacement to thyroid cartilage displacement, referring to as “motion ratio.” Increased motion ratio was observed for stroke patients with neurogenic oropharyngeal dysphagia. Notably, hyolaryngeal distance changes appeared to be insensitive to position ac-

doi:10.6342/NTU202503180

cording to a study that compares the hyolaryngeal distance before swallowing and the shortest hyolaryngeal distance during a swallow (Ahn et al., 2015). The evidence indicates that hyolaryngeal approximation measurements may have a role in evaluating patients for neurogenic oropharyngeal dysphagia.

2.3.3 Changes in Deglutition in Normal Aging

Changes in deglutition can occur as a part of normal aging and may be attributed to alterations in muscle strength, contraction speed, and sensory perception (Groher, 2016). As reviewed in Jones (2003), changes in deglutition with aging include a prolonged oropharyngeal phase, a lower hyoid position, xerostomia, and reduced sensory perception. Nevertheless, it remains a clinical challenge to differentiate aging effects on deglutition from those induced by pathological causes (Groher, 2016).

Regarding hyolaryngeal approximation, Bahia and Lowell (2023) recruited two groups of participants, younger (mean age = 21.95 years) and older (mean age = 70.10 years) individuals, to perform three swallowing tasks. The tasks included (1) normal swallowing, (2) effortful swallowing with tongue emphasis, and (3) effortful swallowing with pharyngeal squeezing. The first condition served as the control. The second and third conditions were elicited by instructions asking the participants to either squeeze hard with their tongues or squeeze hard with their throats. The effortful swallowing conditions showed greater levels of relative hyolaryngeal approximation[3] compared to normal swallowing. Notably, the younger group exhibited significantly greater approximation during effortful swallowing with pharyngeal squeezing in comparison to the older group. In the other two conditions, the younger

[3]Defined as the percentage change relative to the resting distance

doi:10.6342/NTU202503180

group consistently showed a greater level of approximation although the differences were not statistically significant. These findings suggest a potential age-related decline in hyolaryngeal approximation capacity. Beyond hyolaryngeal approximation, the effect of aging on hyoid and laryngeal movement is mixed with evidence showing increased and reduced movement (Kang et al., 2010; Kim & McCullough, 2008; Logemann et al., 2000, 2002).

2.4 The Connection between Speech and Deglutition

Speech production and deglutition appear to be different functions and studies directly comparing the two are rare. One such study is Hiiemae et al. (2002) in which lateral-projection videofluoroscopy was used to examine speech and swallowing in normal young adults. Specifically, they looked at the jaw, tongue, and hyoid.[4] Compared to feeding, during speech (1) jaw movement was smaller and on overage more supero-anterior, (2) hyoid movement was more anterior and limited in area, and (3) tongue movement was less variable but the centroids were no different from those of feeding. Later, Hiiemae and Palmer (2003) reviewed previous research on the jaw-hyoid-tongue complex in relation to feeding[5] and speech. Examining various studies, including Abd-El-Malek (1955), Hiiemae et al. (2002), and Stone and Lundberg (1996), Hiiemae and Palmer (2003) hypothesized a functional linkage of the jaw, hyoid and tongue that serve as the basis for feeding as well as speech and that tongue shapes in feeding served as the basis for those in speech, considering that many tongue gestures in speech could be found in feeding.

Aside from the tongue, the laryngeal movements and hyolaryngeal approximation in speech (Section 2.2) and feeding (Section 2.3) likely involve shared components. Hyolaryngeal

[4]The marks for the jaw was the lower canine. There were two markers, anterior and posterior, for the tongue.
[5]In this context, ”feeding” likely refers to the behavior associated with food intake.

doi:10.6342/NTU202503180

movement depends on the intrinsic and extrinsic laryngeal musculature. The two behaviors should at least share a common neural pathway at the lower motor neuron level.

Pertaining to central mechanisms, McFarland (2006) hypothesized a shared brain network as the basis for both speech and deglutition. More specifically, previous studies revealed that humans have two representations of laryngeal muscles in the motor region of the brain. These are referred to as the laryngeal motor cortices (LMCs), which include a dorsal and a ventral component (dLMC and vLMC respectively), located at the two ends of the orofacial somatotopic map of the precentral gyrus (Belyk et al., 2021). The laryngeal motor cortices are considered to be responsible for the movements of laryngeal musculature for both speaking and deglutition (Li-Jessen & Ridgway, 2020).

Due to the overlapping musculature and neural networks involved in speech and deglutition, it is expected that functional deficits or interventions in one domain may influence the other. Supporting this concept, Castillo-Allendes et al. (2025) reviewed studies on volitional voice tasks and intentional voluntary voice exercises/activities, and concluded that such tasks held potential as assessment or treatment tools for swallowing disorders.

However, it remains to be examined as to how hyolaryngeal approximation, an airway protective measure in deglutition, change over time during the production of linguistic targets, and how different age groups may realize these productions differently. The result may further contribute to the the growing body of literature supporting the link between speech and deglutition.

doi:10.6342/NTU202503180

2.5 Tones in Taiwan Mandarin

Taiwan Mandarin is the Mandarin variety spoken in Taiwan. It emerged after World War II, receiving influence not only from pre-existing local languages, including the widely spoken Taiwan Southern Min, but also other Mandarin varieties brought by Mandarin-speaking immigrants (Kubler, 1985; Kuo, 2005).

Similar to Beijing Mandarin, there are four lexical tones in Taiwan Mandarin. The four tones in Beijing Mandarin are described as high-flat (55), high-rising (35), low-dipping (214), and high-falling (51) (Chao, 1948). The numbers in the parentheses represent the pitch with 5 being highest and 1 being the lowest (Chao, 1930). On the other hand, with trisyllabic target words, Fon and Chiang (1999) revealed that the four tones in Taiwan Mandarin were 44, 323, 312, and 42, with a narrower and lower pitch range. Alternatively, Dan et al. (2006) showed the four Taiwan Mandarin tones to be 44, 23, 211/21, and 51, based on monosyllabic production targets.

Nevertheless, Taiwan Mandarin is not uniform across different ages. Sanders (2008) noted that Tone 2 and Tone 3 patterns varied across age groups. While the older speakers produced Tone 2 more commonly with a raising contour, the younger speakers showed an increased use of dipping contours. The older speakers mainly realized Tone 3 with a dipping contour, whereas younger speakers showed increased preference for a falling contour. On the other hand, Hsu and Tse (2009) showed that older speakers, regardless of ethnical background, showed a greater pitch range for Tone 4.

Aside from generational effects, central Taiwan Mandarin, sometimes referred to as “Taichung accent”, has been reported to have a generally lower and even narrower tonal range (Y.-H.

doi:10.6342/NTU202503180

Huang & Fon, 2011; Khoo, 2020). Taiwanese Southern Min proficiency was also found to be related to Taiwan Mandarin tone performances, along with gender effects. Higher Taiwanese Southern Min proficiency was correlated with a lower pitch range in males, and a narrower pitch range in females (Wu, Fon, et al., 2010).

Apart from pitch, tones in Taiwan Mandarin are associated with other features. One important feature is the use of creaky voice during tone production. Kuang (2017) reported different usages of creaky voice across Taiwan Mandarin tones by analyzing data from a corpus containing 6 Taiwan Mandarin speakers (3 males, 3 females). The probability of creaky voice was found to be highest for Tone 3, and in order, Tone 2, Tone 4, and Tone 1. This finding resonates with the findings in Moisik et al. (2014) in that low F0 may be realized by a raised, constricted laryngeal gestures that promotes the occurrence of creaky voice (see Section 2.2.1).

2.6 Research Questions

Previous research in the field of linguistics has revealed the role of larynx height, rotation, and state in controlling pitch (Gick et al., 2013; Honda, 1995; Honda et al., 1999; Moisik et al., 2014). Regarding the hyoid, being connected to numerous muscles, it serves as the support for the tongue and transfers forces from suprahyoid muscle to the larynx, playing an important role in pitch control and the phenomenon of intrinsic F0, by rotating, or possibly elevating, the larynx (Honda, 1983, 1995; Honda et al., 1999; Menon & Shearer, 1971). Previous studies on the interaction of the hyoid and larynx have explored its role in intrinsic f0 (Honda, 1983; Ladefoged, 1968), and an interaction between lingual postures and tone targets (Shaw et al., 2016). Importantly, previous works also indicated the existence of hyolaryngeal

doi:10.6342/NTU202503180

approximation in phonation, though with varied results (Carlson, 2021; Miller et al., 2012; Miloro et al., 2014). Nevertheless, these studies relied on non-linguistic targets and did not account for temporal changes over productions.

Meanwhile, in the field of medicine, the interaction between the hyoid and thyroid cartilage has important implications in airway protection during the pharyngeal phase of deglutition (Ambrosi & Lee, 2021; Leonard et al., 2000). The reduction of hyolaryngeal approximation has found to be associated with dysphagia (Y.-L. Huang et al., 2009; Matsuo & Matsuyama, 2021). A reduction in hyolaryngeal approximation was also shown in older individuals under effortful swallow (Bahia & Lowell, 2023).

A shared muscular and/or neurological basis for speech and deglutition has been proposed (Belyk et al., 2021; Castillo-Allendes et al., 2025; Hiiemae & Palmer, 2003; Hiiemae et al., 2002; Li-Jessen & Ridgway, 2020; McFarland, 2006). Consequently, aging of the cervical musculature is expected to impact both speech production and deglutition. This age-related effect is likely to be most pronounced when vowel and tone targets place greater demands on the cervical musculature.

By examining hyolaryngeal gestures in non-dysphagic Taiwan Mandarin speakers across different age groups, this study aims to answer the following questions:

1. How does hyolaryngeal approximation pattern over time and vary between different age groups under different linguistic contexts?
2. Does the pattern of muscular activation associated with hyolaryngeal movement vary between different age groups across linguistic contexts?

Together, the study aims to provide a preliminary understanding of hyolaryngeal dynamics

doi:10.6342/NTU202503180

in speech production and serve as a stepping stone for future comparisons with studies on deglutition.

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

Chapter 3 Methods

3.1 Participants

Two groups of participants, young adult and elderly, were recruited. The inclusion criteria required participants to be fluent in Taiwan Mandarin and familiar with Bopomofo, a phonetic transcription system for Taiwan Mandarin, based on self-report. The young adult group comprised individuals aged between 20 and 30 years whereas the elderly group included those aged 65 years and above.

The exclusion criteria required no history of neurodegenerative disorders, dysphagia, or oral or neck surgery deemed to have a significant impact on hyolaryngeal motor function, also based on self-report. Participants scoring 3 or higher on the Eating Assessment Tool (EAT-10) questionnaire (Belafsky et al., 2008; P.-S. Huang, 2023) were excluded. Details of the EAT-10 questionnaire is described in Section 3.2.

Of all recruited participants, three were excluded as they did not meet the requirements. The young adult group included 13 male participants and 10 female participants. The mean age was 24.35 years, and the standard deviation was 2.98 years, with a maximum of 30 years and minimum of 20 years. The elderly group consisted of 5 males and 10 females. The mean age was 70.07 years, and the standard deviation was 3.61 years, with a maximum of

doi:10.6342/NTU202503180

75 years and minimum of 65 years. A Chi-squared test of independence was conducted to examine the relationship between age group and gender. The association was not statistically significant ($\chi^2(1, N = 38) = 1.96,\ p = .162$) suggesting that gender distribution did not differ significantly between the two age groups.

	Young Adult (n = 23)	Elderly (n = 15)
Male	13	5
Age (mean)	24.35	70.07
Age (SD)	2.98	3.61
Max age	30	75
Min age	20	65

Table 3.1: Summary of included participants

3.2 Questionnaire

Participants were asked to complete two surveys before the experiment: the Taiwan Mandarin version of EAT-10 (Belafsky et al., 2008; P.-S. Huang, 2023) and a linguistic background questionnaire.

The EAT-10 survey was conducted to rule out participants with potentially undiagnosed dysphagia. EAT-10 is a self-report questionnaire in which respondents rate the severity of 10 symptoms on a scale of 0 to 4. The Taiwan Mandarin translation of EAT-10 was adopted from P.-S. Huang (2023). At a cut-off score of ≥ 3, the sensitivity and specificity for distinguishing between dysphagic and nondysphagic individuals were reported to be 86% and 96%, respectively (P.-S. Huang, 2023). Therefore, by excluding participants scoring 3 or higher, chances of including individuals with undiagnosed dysphagia ought to be low.

The language background questionnaire collected the following information: (1) the age of the participant, (2) the place where the participant grew up, defined as the location where

doi:10.6342/NTU202503180

they lived prior to the age of 18, (3) the languages spoken by the participant, including their first language, (4) the languages spoken by the participant's primary caregiver(s) during childhood, and (5) whether the participant has lived overseas for more than six months, and if so, the corresponding time period(s).

3.3 Stimulus Design

There were 12 stimuli, corresponding to three Taiwan Mandarin vowels, [a], [i], and [u], across the four Taiwan Mandarin lexical tones (see Section 2.5 for a description of the tones). These stimuli were shown to the participants in the form of Bopomofo (Table 3.2).

	Tone 1	Tone 2	Tone 3	Tone 4
[a]	ㄚ	ㄚˊ	ㄚˇ	ㄚˋ
[i]	ㄧ	ㄧˊ	ㄧˇ	ㄧˋ
[u]	ㄨ	ㄨˊ	ㄨˇ	ㄨˋ

Table 3.2: Visual stimuli in Bopomofo used in the ultrasound and electromyography tasks

3.4 Experiment Procedure, Apparatuses, and Set-Up

Participants were first asked to complete the two questionnaires before the experiments (see Section 3.2). Two experiments were conducted: the ultrasound and electromyography experiments. The order of the ultrasound and electromyography experiments was counterbalanced across participants. Nevertheless, the electromyography electrodes were always attached to the speakers' necks prior to both experiments, regardless of the order. For both experiments, the stimuli were presented to the participants with PsychoPy (Peirce et al., 2019).

Each block consisted of all twelve stimuli, with each stimulus presented once in random order. The block was repeated 5 times, resulting in 60 tokens for electromyography and ul-

doi:10.6342/NTU202503180

trasound experiments, respectively. Pacing was controlled by the experimenters to ensure no overlapping of gestures between two productions. Participants were refrained from production for 4 seconds if they noticed themselves coughing, swallowing, burping or having other lingual or laryngeal movements unrelated to articulation. The experimenters also paid additional attention to such incidents and verbally asked participants to delay their production. If the production target was inaccurate or the quality of the recorded data were poor, the experimenters would correct the participants or ask the participants to repeat.

Ultrasound was recorded with a portable ultrasound machine (CGM OPUS5100) using a linear probe (LA80N). The probe was placed on the right anterior neck, deviating within 30 degrees from the midsagittal line (Figure 3.1). It was confirmed that both the hyoid and thyroid were included in the visual field before the actual experiment began. To align the audios and ultrasound videos, the participants were asked to cough three times at the beginning of the recording. The movements of the coughs were later used as a cue to align the audio and video files.

The ultrasound imaging system recorded data at a frame rate of 46 frames per second (fps), and the video output was captured to the computer at 60 fps with a capture card (AverMedia) for subsequent analyses. The audio was recorded with a microphone (Audio-Techinica AT2035) in Praat (Boersma, 2001) sampled at 44.1 kHz.

For the electromyography experiment, surface electromyography was used to collect the activity of muscle groups in the neck. The signals were collected with electrodes, acquired with the data acquisition device PowerLab (ADInstruments) and logged by the LabChart software (ADInstruments). The signals were recorded at a sampling rate of 40,000 Hz. Three pairs of electrodes were attached to the left anterior aspect of the neck with 3M^{TM} Trans-

doi:10.6342/NTU202503180

Figure 3.1: The probe was placed within 30 degrees deviating from the midsagittal line. For demonstrative purposes, the surface electromyography nodes were not attached.

pore™ tape. The first pair was attached to the submental area, anterior to the hyoid, and aimed to record the activation of the submental musculature. The second electrode pair was placed on the superolateral side of the thyroid cartilage, above the lamina, while avoiding the curved transition between the neck and the jaw. If the space between the thyroid cartilage and the curved transition was too narrow for electrode placement, the electrodes were instead positioned partially or completely over the lamina. This pair was intended to record the muscle activity in the region between the hyoid bone and the thyroid cartilage, with particular attention to the thyrohyoid muscles. The third pair was attached to the neck at the midpoint between the thyroid prominence and the sternoclavicular joint. The nodes were placed to retrieve activation from the infrahyoid muscles, especially the activation from the sternohyoid and sternothyroid muscles. The ground was attached to the left dorsal wrist, serving as a baseline for muscle activity. The location of the three pairs of electrodes were shown in Figure 3.2.

doi:10.6342/NTU202503180

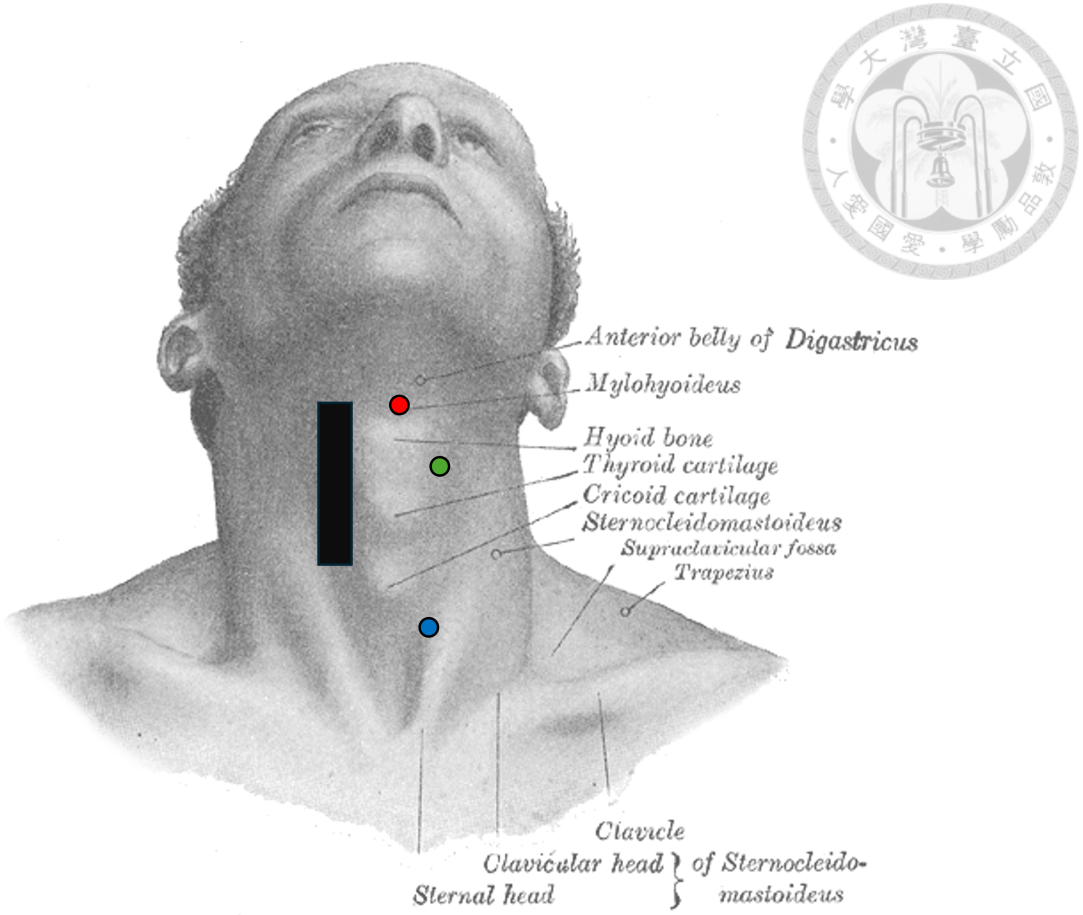


Figure 3.2: Locations of the surface electromyography electrodes. The first pair was attached to the speaker submental area (indicated by the red circle), the second pair at the superolateral side of the thyroid cartilage (i.e., the green circle), the third pair at the mid point between the thyroid prominence and the sternoclavicular joint (indicated by the blue circle). The black rectangle represents the placement of the ultrasound probe. Image adapted from Structure of Adam's Apple (Public domain), via Wikimedia Commons.

Source: https://commons.wikimedia.org/wiki/File:Structure_of_Adam%27s_apple.png

A microphone (Audio-Technica AT2035) was used to record the sound. The microphone was connected to USBPre 2 (Sound Devices), a high-resolution USB-audio interface. Two outputs were routed from USBPre 2. One was transferred directly into the computer and was recorded with Praat (Boersma, 2001). The other output was sent into the PowerLab machine and logged in the LabChart software as the fourth channel. The acoustic data were used in the annotation process to mark the onset time of each token.

doi:10.6342/NTU202503180

3.5 Ultrasound Data Processing

3.5.1 Alignment and Annotation

The ultrasound videos and audios were aligned in Kdenlive, a free and opensource video editing software. The audio and video were aligned with the first frame where the thyroid cartilage began to move in an opposite direction from the previous frames during a coughing movement. Video and audio files were then exported from the aligned product. The audio file was then processed in Praat (Boersma, 2001) to mark the onsets and offsets of voicing for each token as determined by the onsets and offsets of the voice bar on the spectrogram. 2.5-second-long video segments were extracted, each beginning 1 second before the voice onset and ending 1.5 seconds after the onset. The script also calculated the duration of each production, which was used for later data processing.

A user-defined python script using the *CSRT tracker* (Lukezic et al., 2017) provided by the OpenCV package (Bradski, 2000) was used to track structures in the video segments. The most caudal point of the hyoid (the red dot in Figure 3.3) and the most rostral point of the thyroid cartilage (the blue dot in Figure 3.3), as visible on the ultrasound frames, were manually selected on the first frame. Rectangular regions of interest (ROIs), each measuring 150 × 70 pixels, were then created for the two structures, with the selected points serving as the centers of their respective ROIs. The ROIs were then tracked by the CSRT tracker (Lukezic et al., 2017) (Figure 3.4).

After tracking, the results were manually checked and adjusted if needed. Frames with any target considered hard to annotate were removed from analysis. Tracked data were saved as Cartesian coordinates. The origin of the coordinate plane in each frame located at the

doi:10.6342/NTU202503180

top-left corner.

One participant was removed from the ultrasound entirely due to overall bad image quality. In addition, one token was accidentally skipped during the recording phase. Six tokens were removed due to unclear hyoid and/or thyroid landmarks on the first frame. With a total loss of 67 tokens, this summed up to a total loss of approximately 2.94% of all tokens. Each video was 2.5 seconds, and hence 150 frames given a fps of 60[1]. In total, over 300,000 frames were processed.

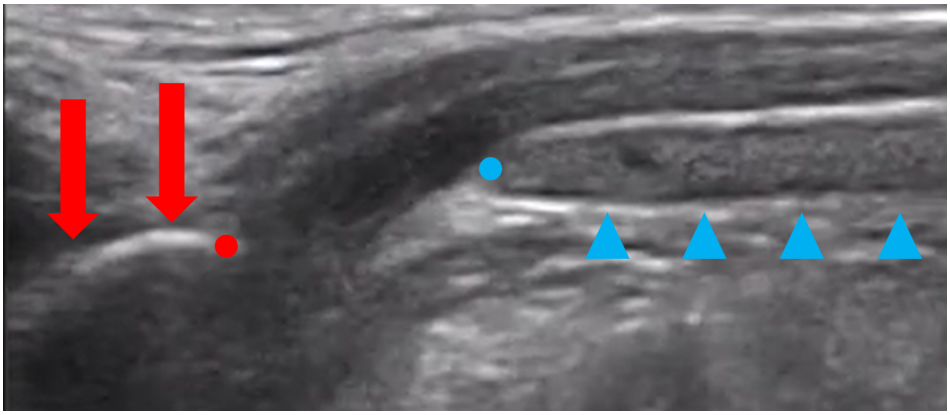


Figure 3.3: An example of the annotated points in ultrasound. The left side of the figure is the rostral side of the neck. The upper side of the figure is the superficial side of the neck. The red arrow points at the hyoid bone and the blue triangles indicate the thyroid cartilage. The red and blue dots represent the annotated points for the hyoid and the thyroid cartilage, respectively.

3.5.2 Ultrasound Data Preparation

The hyolaryngeal approximation was defined as the Euclidean distance between the annotated points for the hyoid bone and the thyroid cartilage (Figure 3.3). This Euclidean hyolaryngeal distance (hereafter *HT*) for each frame was calculated from the coordinates of the annotated hyoid and thyroid (Equation 3.1). Since each production and each person may be-

[1] Very occasionally, there are videos with only 149 frames.

doi:10.6342/NTU202503180

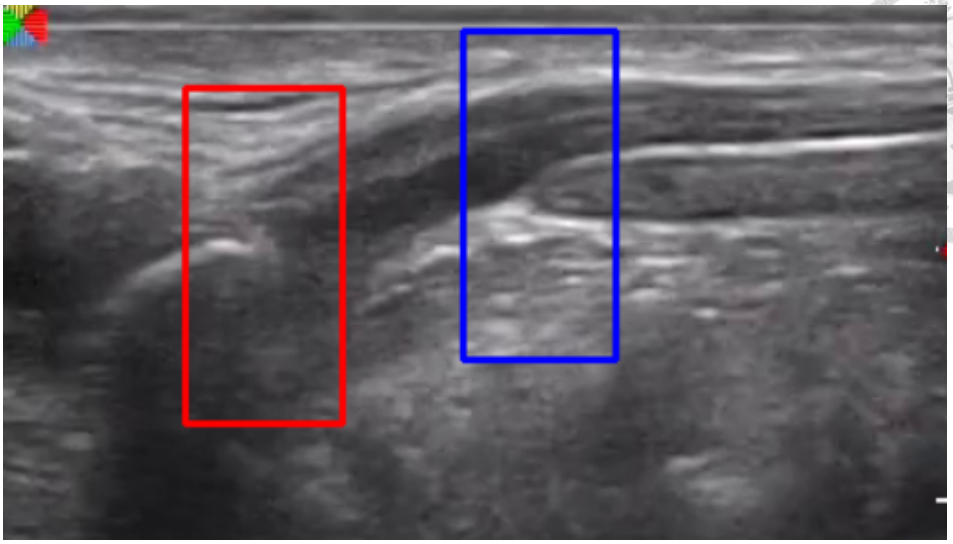


Figure 3.4: An example of the tracking process of the *CSRT tracker*. The red and blue boxes are the tracking box (ROI) of the hyoid bone and thyroid cartilage, respectively. The center of the boxes are the tracking targets.

gin with a different *HT*, the *HT* of the first frame was subtracted from the *HT* of each frame. The result is the absolute *HT* change, referred to as *aHT* (Equation 3.2). The *HT* from each frame was alternatively normalized to the *HT* of the first frame, assuming the first frame is associated with the resting *HT*; this normalized *HT* is referred to as *nHT* (Equation 3.3). In the equations, *i* refers to the frame number, *x* and *y* refer to the (x, y) coordinates of the annotated hyoid bone and thyroid cartilage.

$$HT_i = \sqrt{(Hyoid_{\text{x.axis},i} - Thyroid_{\text{x.axis},i})^2 + (Hyoid_{\text{y.axis},i} - Thyroid_{\text{y.axis},i})^2} \tag{3.1}$$

$$aHT_i = HT_i - HT_1 \tag{3.2}$$

$$nHT_i = \frac{HT_i}{HT_1} \tag{3.3}$$

In addition to distance, two temporal references were used. The absolute time of each token was calculated directly from the frame number. Since the frame rate of the videos was

doi:10.6342/NTU202503180

60 fps, dividing frame numbers with 60 yielded the corresponding absolute time in seconds. In each extracted video, the voice onset consistently occurred at the 1-second mark within each video (see Section 3.5.1). To move the voice onset to time zero, one was subtracted from the absolute time values. The resulting temporal reference system was the absolute time dataset, referred to as *abs_time*. An additional temporal reference was obtained by normalizing the voicing duration to 100% (i.e., *nor_time*). This was achieved by dividing each *abs_time* value by the corresponding token duration, which had been extracted in a previous step (see Section 3.5.1).

3.6 Surface Electromyography

3.6.1 Annotation and Exportation

Electromyography data were annotated using LabChart. Each token was marked based on the voice onset in the audio. The onset was verified through examination of the spectrogram, which was generated using a fast Fourier transform (FFT) in LabChart with a 256-point window. The annotated data were downsampled to 2,000 Hz and exported for ease of further processing. This sampling rate was more than twice the highest frequency of interest in subsequent analyses, thus satisfying the Nyquist criterion and minimizing the risk of aliasing.

One participant was excluded from the analysis due to unusable acoustic data in LabChart resulting from a setup error. Three additional participants were excluded from the analysis due to poor data quality. The remaining dataset thus consisted of 22 young adult participants (male = 12) and 12 elderly participants (male = 5).

doi:10.6342/NTU202503180

3.6.2 Electromyography Data Processing

The electromyography data processing was based on Clancy et al. (2023) and follows the order of token extraction, notch filter application, bandpass filter application, rectification, and moving window average.

The electromyography data for each token was extracted, between 1s before and 1.5s after the onset points annotated in LabChart. The data of each token were first examined for continuous data loss. The token was discarded in the case where any of the electromyography channels (i.e. except for channel 4, which contained audio signals) contained any continuous NaN segments (i.e., out of range data) over 5 ms. Thirty-eight tokens were removed due to prolonged missing data in addition to the four participants removed due to set up errors or overal poor signal quality. This resulted in a total loss of approximately 12.19% of all expected tokens.

For the retained token data, missing values (NaNs) were linearly interpolated, with any remaining leading or trailing NaNs filled using forward and backward filling, respectively. The electromyography data for each token was filtered first with a 60 Hz notch to remove power line interference, implemented with a second-order IIR notch filter designed with `iirnotch()`, and then applied with zero-phase distortion using `filtfilt()`. After the notch filter, a 4th-order Butterworth bandpass filter with the upper and lower limit at 450 Hz and 20 Hz designed with `butter()` was applied using `filtfilt()` to preserve phase integrity (Virtanen et al., 2020) The filtered signals were subsequently rectified and processed using a centered moving-window average filter with maximum overlap. The window size was 101 sample points.

doi:10.6342/NTU202503180

An example of the transformation of the electromyography data is shown in Figure 3.5. The first row in the figure showed the raw data of a channel after token extraction with visible low frequency components in the -0.5 to 1s window that are presumably motion artifacts. After applying the filters (notch and bandpass), the low frequency component was removed. The third row shows the result of rectification and the fourth row showed the signal curve after applying the moving window average filter.

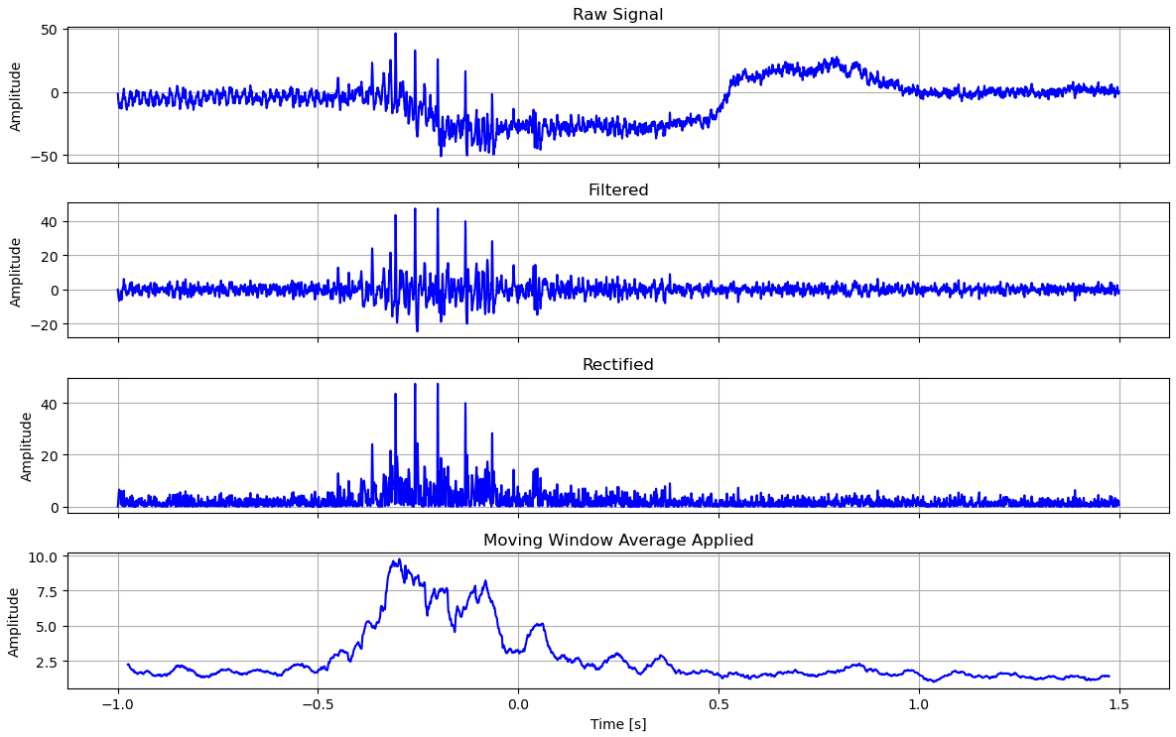


Figure 3.5: The change of electromyography signals with processing. The first row shows the raw data. The second row shows the same signal after applying notch and bandpass filters. The third row shows the signals after rectification. The last row shows the data after applying moving window average.

The data points from the initial and final 25 ms were removed as they contained no values (NaN) resulted from centered 101-sample window average application. EMG signals were then normalized by the maximum value observed in each channel across all tokens for each participant.

doi:10.6342/NTU202503180

3.7 Analysis

3.7.1 Analysis of Ultrasound Data

The aims of this study are to explore how hyolaryngeal approximation pattern temporally in speech production and whether the extent of approximation would differ between age groups under different vowel-tone combinations. To answer these questions, *aHT* and *nHT* were modeled for each vowel-tone combination using generalized additive mixed modeling (GAMM), implemented in the mgcv package (Wood, 2011).

GAMM allows for nonlinear relationships between predictors and the response. Hyolaryngeal approximation is expected to vary nonlinearly over time during speech production, and factors such as age group may also exert nonlinear effects. These characteristics make GAMMs an ideal choice for modeling hyolaryngeal approximation. Only tokens containing all 150 frames were included, resulting in 1,643 tokens, approximately 72.06% of the expected total.

The dependent variable in the generalized mixed additive model was hyolaryngeal approximation, modeled with either *aHT* or *nHT*. The model included age group (young adult and elderly groupings) as a parametric effect. A smooth term for time, modeled with either *abs_time* or *nor_time*, was included and was set to vary by age group, capturing group-specific temporal trajectories. Factor smooths were specified for each participant within each age group, allowing for non-linear, group-specific participant-level variability over time. Additionally, a random intercept for each production was included to account for variations across productions.

To account for autocorrelation between the residuals of successive time points in a time

doi:10.6342/NTU202503180

series, an AR1 error model was included. The model requires two inputs, *Rho*, the amount of the estimated autocorrelation, and *AR.start*, which tells the model where a new time series begin (van Rij et al., 2022; Wieling, 2018; Wood, 2011).

This model structure allows for flexible, non-linear changes over time while accounting for speaker-specific and production-specific variation in the dynamics of hyolaryngeal movement. The model described above is shown in Equation 3.4.

$$y \sim \text{age} + s(\text{Time}, \text{by} = \text{age}) + s(\text{Time}, \text{participant}, \text{by} = \text{age}, \text{bs} = \text{'fs'}, m = 1) + s(\text{id}, \text{bs} = \text{'re'}), \text{method} = \text{'fREML'}, \ \text{Rho}, \ \text{AR.start} \tag{3.4}$$

The significance of both parametric and smooth terms was evaluated, and to improve parsimony, a term was removed from the model if not significant for all vowel-tone combination. Between the two distance and temporal references (*aHT* and *nHT*; *abs_time* and *nor_time*), the optimal models were selected based on the Akaike Information Criterion (AIC; Akaike (1974)) and Bayesian information criterion (BIC; Schwarz (1978)), both of which evaluate the trade-off between model fit and complexity. Lower AIC or BIC values, when comparing models fitted to the same data, indicate better model performance.

The selected models were then plotted with `plot_smooth` and `plot_diff`, both provided by the `itsadug` package (van Rij et al., 2022), to visualize the fitted curves for both speaker groups and the difference curves between the two groups.

3.7.2 Analysis of Electromyography Data

To explore potential differences in muscle activation patterns between age groups, surface electromyography signals were qualitatively analyzed through visual inspection. Electromyo-

doi:10.6342/NTU202503180

graphy data from young adult and elderly speakers were separated for subsequent analyses. The electromyography data for the same vowel-tone combination were summed. The data were then baseline-corrected by subtracting the value at the first time point, such that each summed curve starts at zero. Subsequently, the data for each channel were normalized again to 100% to facilitate visual comparison across the three channels. Summed signals were examined for broad trends in activation timing and contour shape across conditions.

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

Chapter 4 Results

4.1 Ultrasound Results

For the general additive mixed models constructed, the parametric effect *age* was found to be nonsignificant in all models and was therefore removed. Moreover, all models analyzing *aHT* showed larger values for AIC and BIC compared ot those analyzing *nHT*. Thus, the *nHT* models were preferred and presented in this section. The resulting models analyzing *nHT* are expressed in Formula 4.1. The AIC and BIC results for all models fitted with Formula 4.1 are presented in Appendix A.1.

$$\begin{aligned} \text{nHT} \sim\ & s(\text{Time}, \text{by} = \text{age}) + s(\text{Time}, \text{participant}, \text{by} = \text{age}, \text{bs} = \text{'fs'}, m = 1) + \\ & s(\text{id}, \text{bs} = \text{'re'}),\ \text{method} = \text{'fREML'},\ \text{Rho},\ \text{AR.start} \end{aligned} \tag{4.1}$$

Regarding the temporal parameters, AIC measures generally favored *abs_time* models while BIC measures favored *nor_time* models. This section presents results from models fitted with both temporal references. Summaries of all *nHT* models are provided in Appendix A.2 for the *abs_time* models and Appendix A.3 for the *nor_time* models.

The ultrasound results for vowel [a] is shown in Figure 4.1. There are two subfigures in

doi:10.6342/NTU202503180

the figure, each containing eight plots. In Figure 4.1a, the top row shows fitted *nHT* trajectories, with *abs_time* as the temporal reference. The x-axis shows *abs_time* in seconds, with time 0s marking the voice onset; the y-axis indicates the distance reference *nHT*, where a value of 1 corresponds to the hyolaryngeal distance in the first frame, representing the resting distance. A lower value thus means hyolaryngeal approximation. The plots are also arranged by tones, Tone 1 to Tone 4, from left to right. Red and blue curves show the results of the young adult and elderly speakers, respectively, with translucent color shading indicating the 95% confidence intervals. The bottom row depicts the differences between the curves in the corresponding top-row plots. The translucent shading again denotes the 95% confidence intervals. Additionally, in the bottom-row plots, x-axis segments bounded by dotted red vertical lines and a red horizontal line along the x-axis mark time intervals where the two curves differ significantly.

From Figure 4.1a, it can be observed that both groups exhibited hyolaryngeal approximation during speech production, and that significant differences in the change of hyolaryngeal distance over time were found between the two age groups in Tones 2, 3, and 4. Significant differences were observed near or overlapping with time 0s for Tones 2, 3, and 4, in which young adult speakers exhibited greater hyolaryngeal approximation. In Tone 3, an additional period of significant difference appeared around 0.7 to 0.9s, during which elderly speakers showed greater approximation.

On the other hand, Figure 4.1b shows models fitted with *nor_time* as the temporal reference. The x-axis represents *nor_time* in percentage. Time 0% and 100% mark the onset and offset of voicing, respectively. The plots are otherwise arranged the same way as Figure 4.1a. Significantly more approximation was observed in [a] productions in Tones 2, 3

doi:10.6342/NTU202503180

and 4, overlapping with voicing. The differences in Tones 3 and 4 began prior to onset. The difference in Tone 4 was relatively short compared to those of Tones 2 and 3. For Tone 1, a similar pattern was observed for the difference curves from both groups, showing a tendency of greater hyolaryngeal approximation, though the difference was not statistically significant.

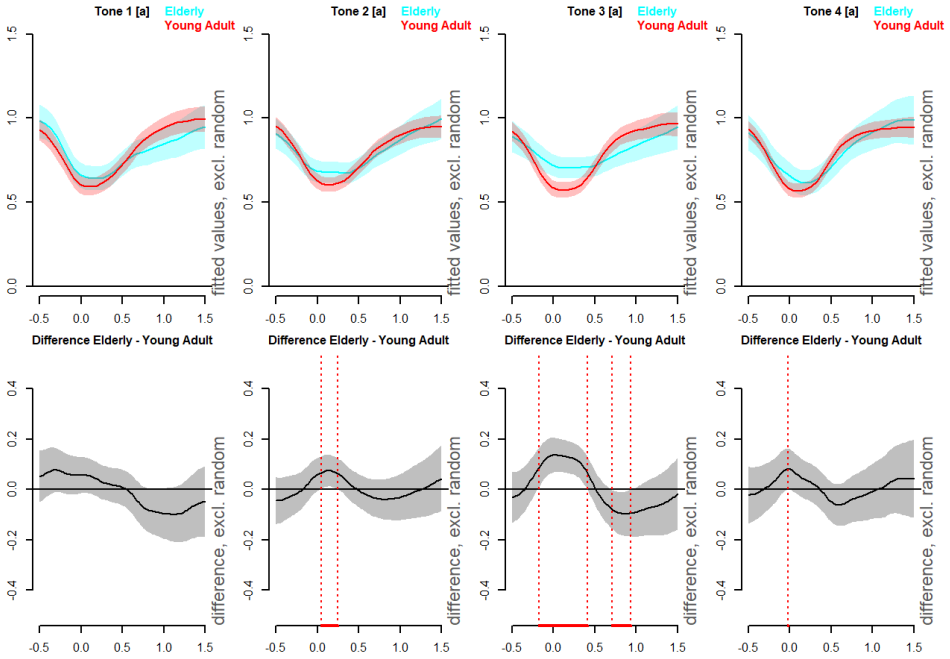
Figure 4.2a shows the results of the *abs_time* models for [i]. Significantly greater approximation in young adult speakers was observed in Tones 2, 3, and 4. The difference for Tone 2 spanned the longest time window, while Tone 3 was the shortest. Figure 4.2b displays the results of the *nor_time* models for [i]. Only Tone 2 and Tone 4 productions showed significantly greater hyolaryngeal approximation in the young adult speakers. The difference curves for Tone 1 and Tone 3 productions indicated a nonsignificant trend of greater approximation for the young adult speakers.

Figure 4.3a shows the results of the *abs_time* models for [u]. Significantly greater hyolaryngeal approximation in the elderly group was only found in Tone 2 over a very short time window. No significant difference was elsewhere observed. Significant differences were not found for the *nor_time* models (Figure 4.3b). Nevertheless, the difference cruves of Tone 1 and Tone 4 productions suggested a nonsignificant but greater hyolaryngeal approximation in the young adult speakers.

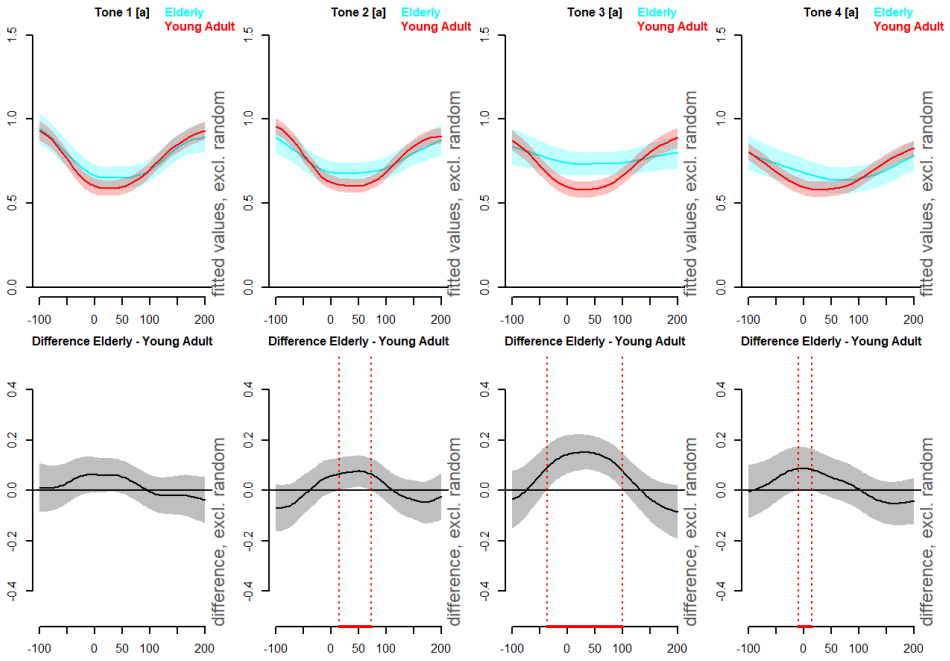
4.2 Surface Electromyography Results

The summed surface electromyography results for [a] produced by the young adult speakers are presented in Figure 4.4. From top to bottom, each plot corresponds to Tones 1 through 4, respectively. The x-axes indicate absolute time in seconds, with 0s marking voice onset. The y-axes represent relative amplitude, where 100% corresponds to the maximum summed

doi:10.6342/NTU202503180



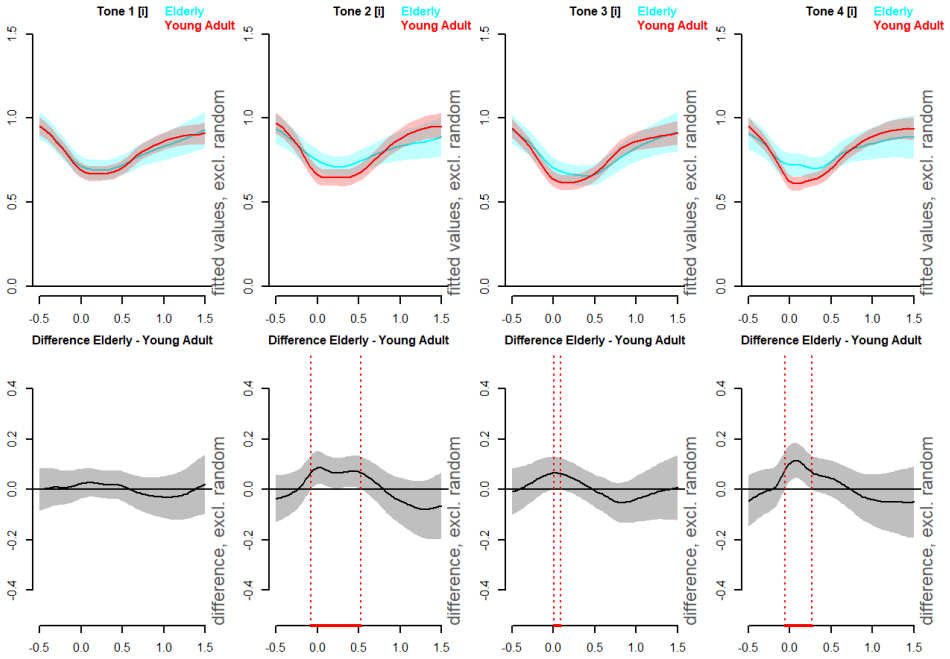
(a) *abs_time* models



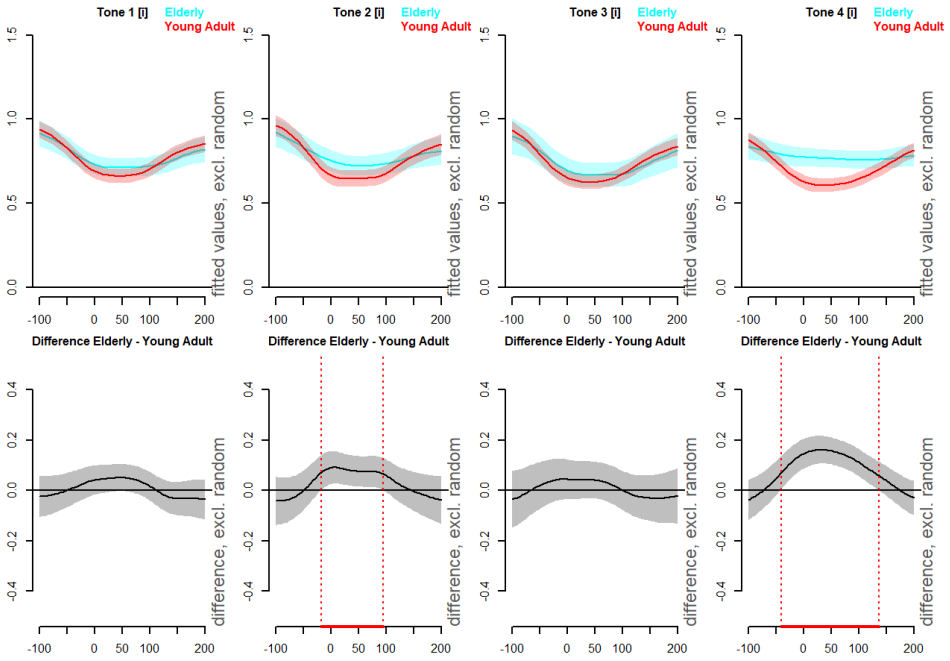
(b) *nor_time* models

Figure 4.1: Generalized additive mixed model results for [a]

doi:10.6342/NTU202503180



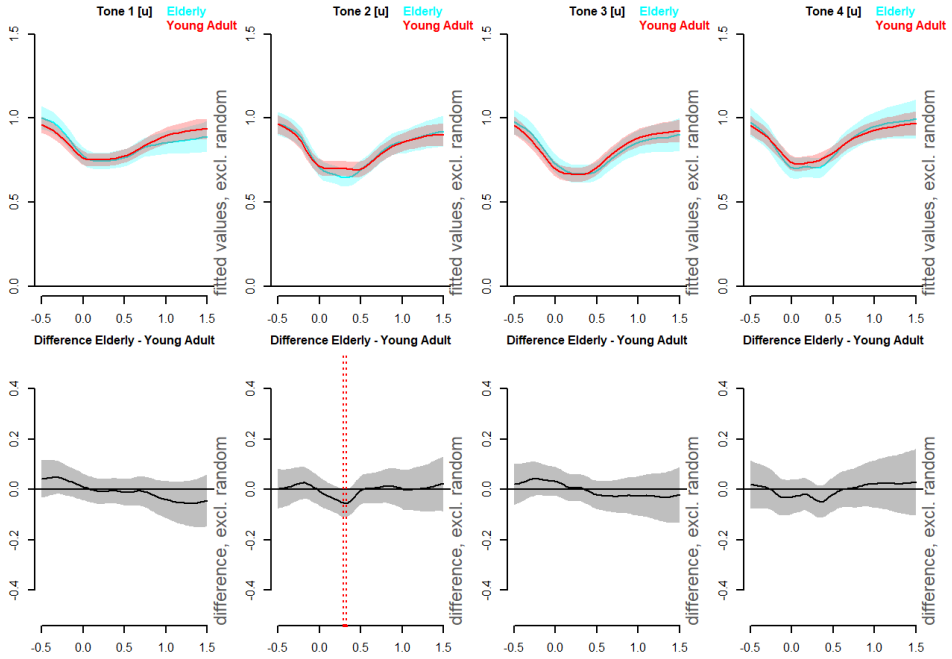
(a) *abs_time* models



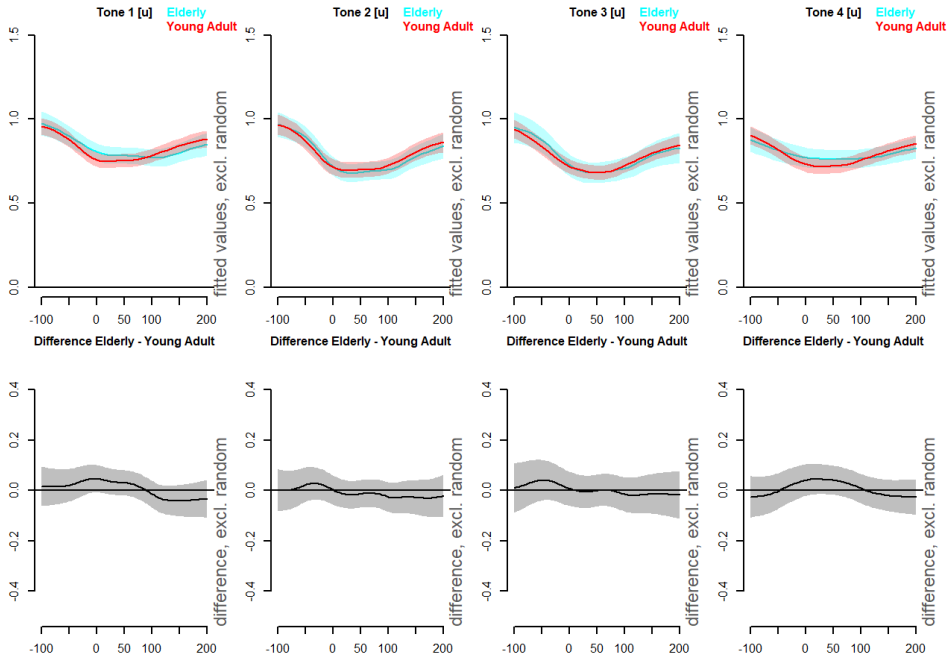
(b) *nor_time* models

Figure 4.2: Generalized additive mixed model results for [i]

doi:10.6342/NTU202503180



(a) *abs_time* models



(b) *nor_time* models

Figure 4.3: Generalized additive mixed model results for [u]

doi:10.6342/NTU202503180

amplitude observed for a given channel within the specific vowel-tone combination, and 0% corresponds to the amplitude at the earliest time point in the resepective summed electromyography series.

The red, green, and blue lines represent the electrode pairs at the submental area, the space between the hyoid bond and thyroid cartilage, and the midpoint between the thyroid prominence and sternoclavicular joint, respectively (Section 3.6). The channels are correspondingly labeled as submental, hyoid-thyroid, and lower neck in the figures.

For young adult productions of [a] (Figure 4.4), all three channels exhibited an early local peak in activity approximately 0.2 seconds prior to the voice onset (arrowheads in Figure 4.4). For Tone 1, the activity in all channels began to decline following this peak. In Tone 2, the submental and hyoid-thyroid channels demonstrated more prolonged activity relative to the lower neck channel after the voice onset. For Tone 3, the activity across all three channels persisted briefly after the voice onset before declining as denoted by the arrow in the figure. In Tone 4, a prominent peak, marked with an arrow, was observed in the all three channels shortly after the voice onset.

For young adult productions of [i] (Figure 4.5), channel activity in Tones 1 and 3 showed a gradual decline shortly after the voice onset. Tone 2 again exhibited prolonged activation in the submental and hyoid-thyroid channels after the voice onset. In Tone 4, a distinct post-onset peak, denoted with an arrow, was observed in the all three channels.

For young adult productions of [u] (Figure 4.6), the activity of Tones 1 and 2 exhibited extended activation in the submental and hyoid-thyroid channels compared to the lower neck channel after the voice onset. The activity of all three channels declined after voicing onset in Tone 3. Tone 4 featured a post-onset peak in all three channels as with [a] and [i], marked

doi:10.6342/NTU202503180

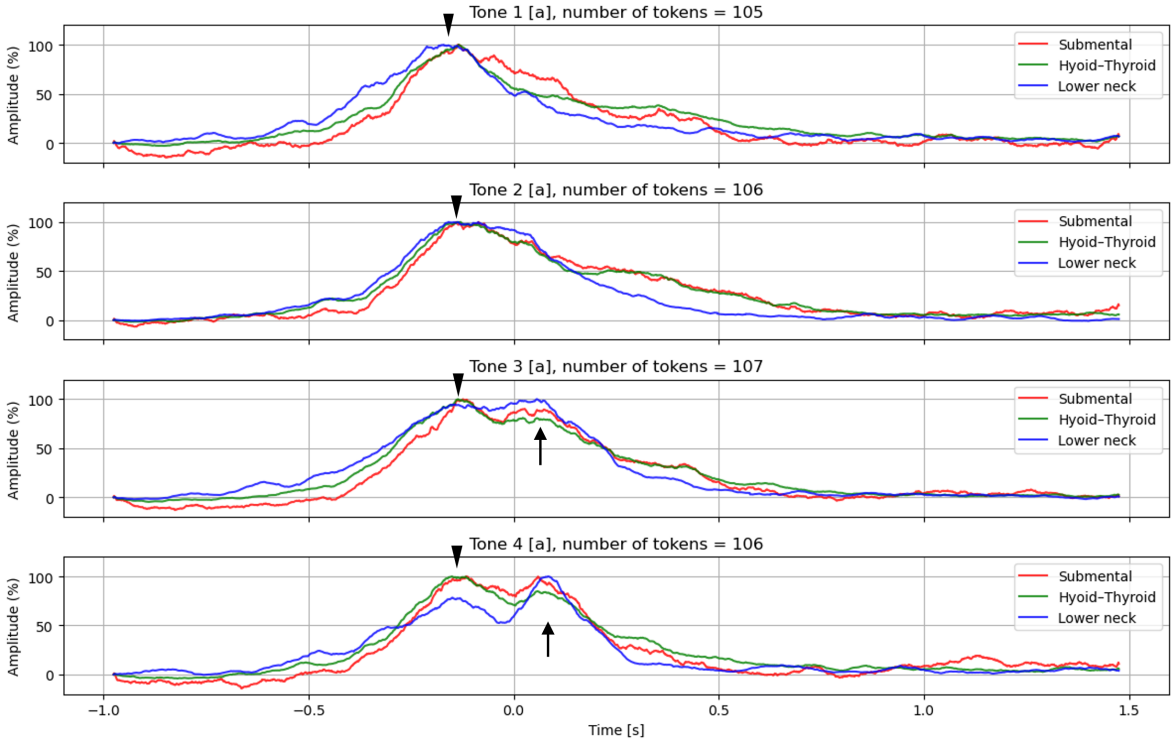


Figure 4.4: Electromyography results for [a] from the young adult group.

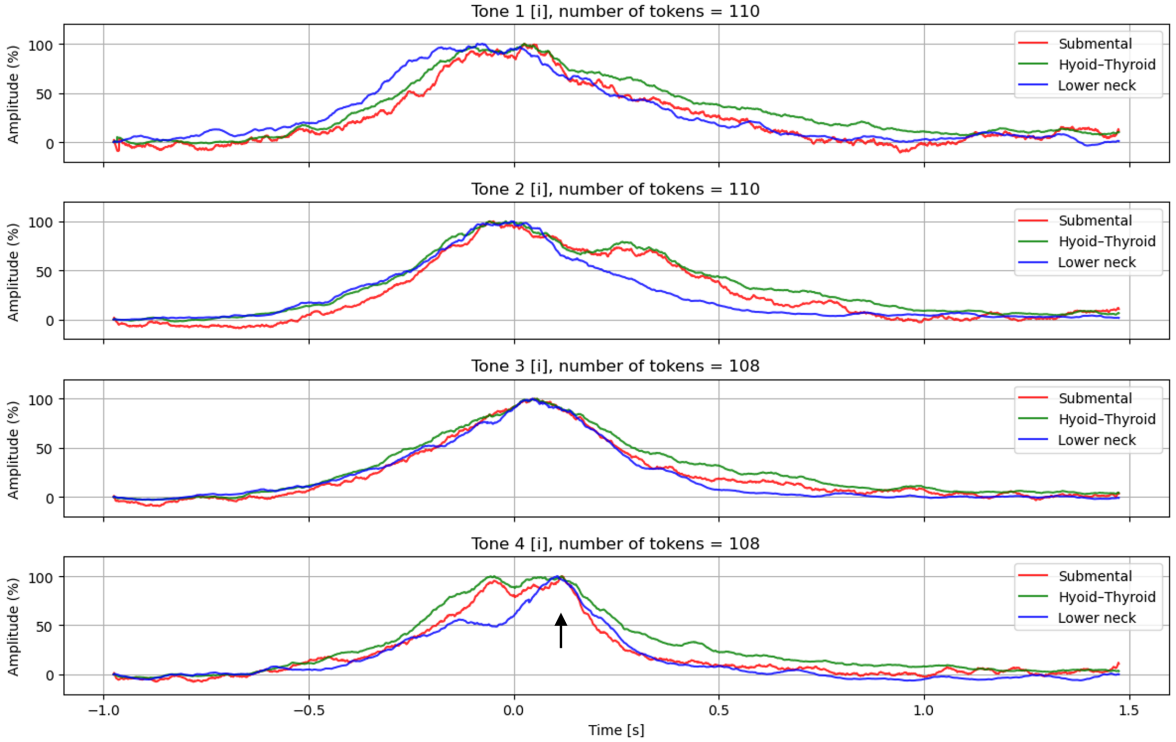


Figure 4.5: Electromyography results for [i] from the young adult group.

doi:10.6342/NTU202503180

with an arrow.

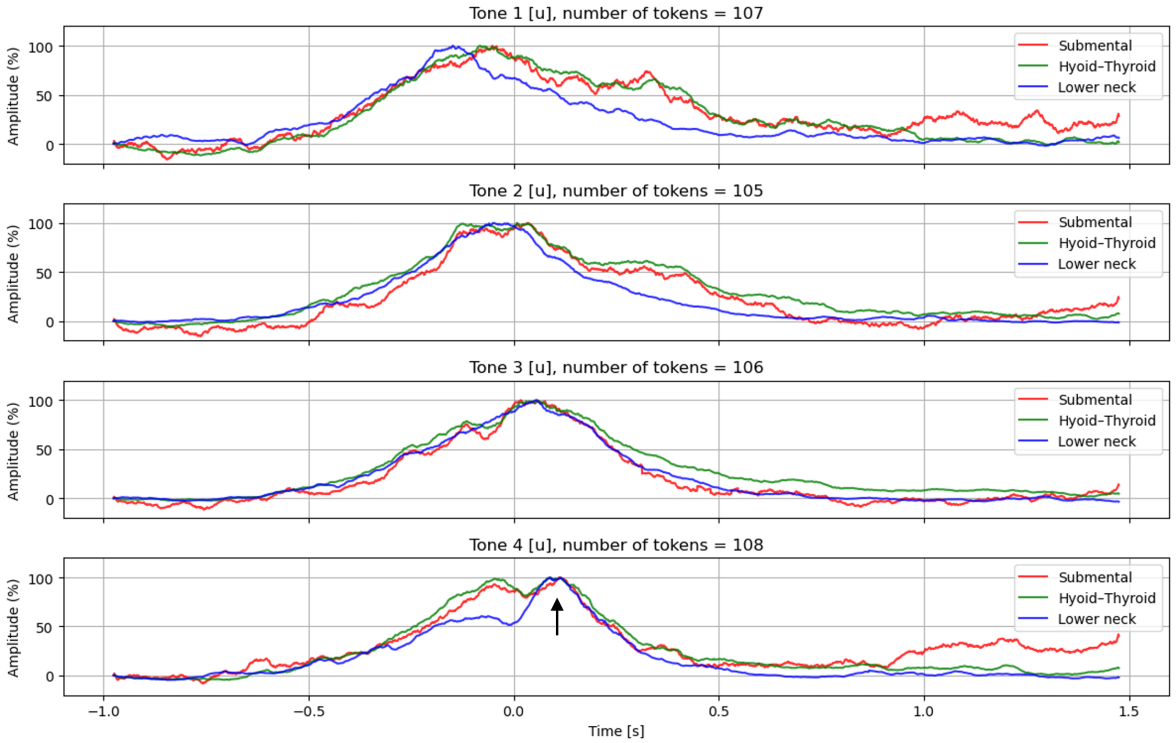


Figure 4.6: Electromyography results for [u] from the young adult group.

Turning to the electromyography results for the elderly group, a more varied pattern was observed. For [a] productions (Figure 4.7), the activities of all three channels in Tone 1 decline gradually beginning slightly before the voice onset. Following the voice onset, Tone 2 and Tone 3 activity in the submental and hyoid and hyoid-thyroid channels were relatively prolonged compared to the lower neck channel. Tone 4 again showed a post-onset peak, denoted with an arrow, in the lower neck channel though less robust compared to those observed in the young adult productions (*cf.* Figure 4.4).

Different from elderly [a], elderly [i] productions revealed earlier activation of the submental and/or hyoid-thyroid channels prior to the voice onset (arrowheads in Figure 4.8). Production patterns for Tones 1 and 2 appeared more variable. In Tone 1, activation in the

doi:10.6342/NTU202503180

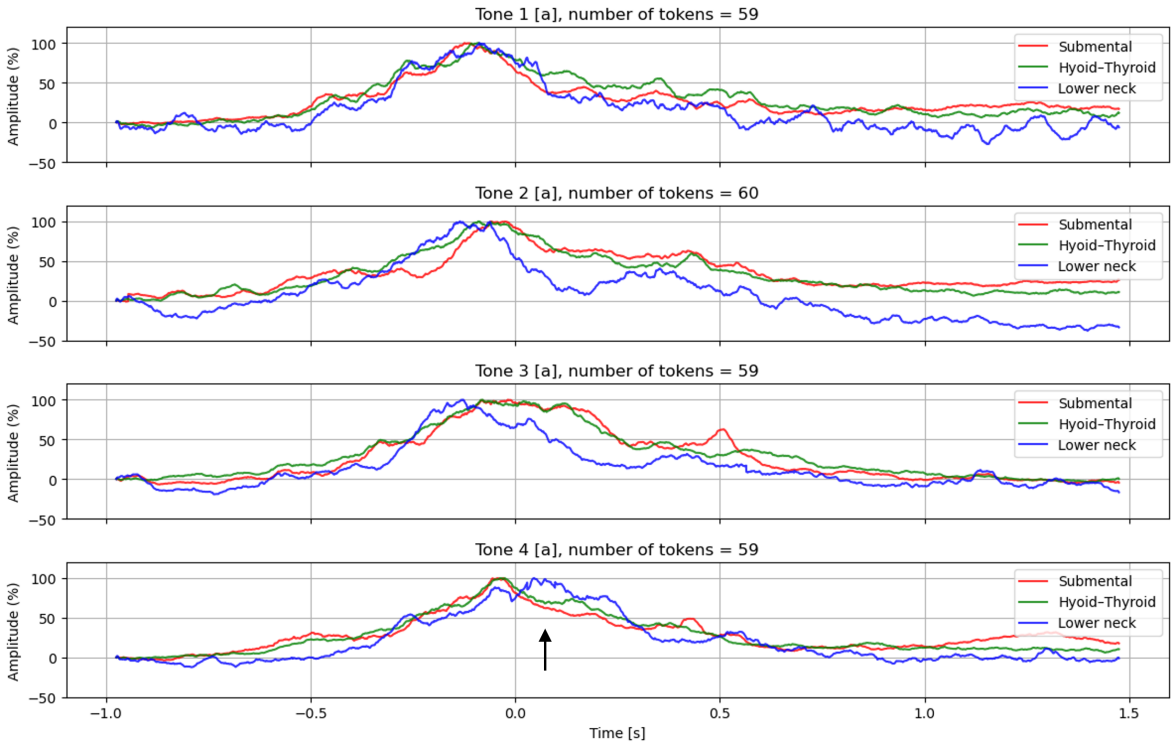


Figure 4.7: Electromyography results for [a] from the elderly group.

hyoid-thyroid channel persisted longer than the other two channels, as marked with an arrow. In Tone 2, all three channels appeared to follow a similar post-onset pattern. For Tone 3, as annotated with an arrow, the activity in all channels peaked shortly after the voice onset and then declined; an additional local maximum was observed around 0.4 to 0.5 seconds post-onset, labeled with a dotted arrow. Tone 4 exhibited a pattern in which all three channels followed a similar activation trend after voice onset.

Similar to [i] productions, elderly [u] productions again revealed earlier activation of the submental and/or hyoid-thyroid channels (arrowheads in Figure 4.9). Production patterns for Tones 1 and 2 were also varied. In Tone 1, the hyoid-thyroid and lower neck channels appeared to persist longer than the submental channel as annoated with an arrow. In Tone 2, activity across all channels followed a similar post-onset trend. Tone 3 production showed

doi:10.6342/NTU202503180

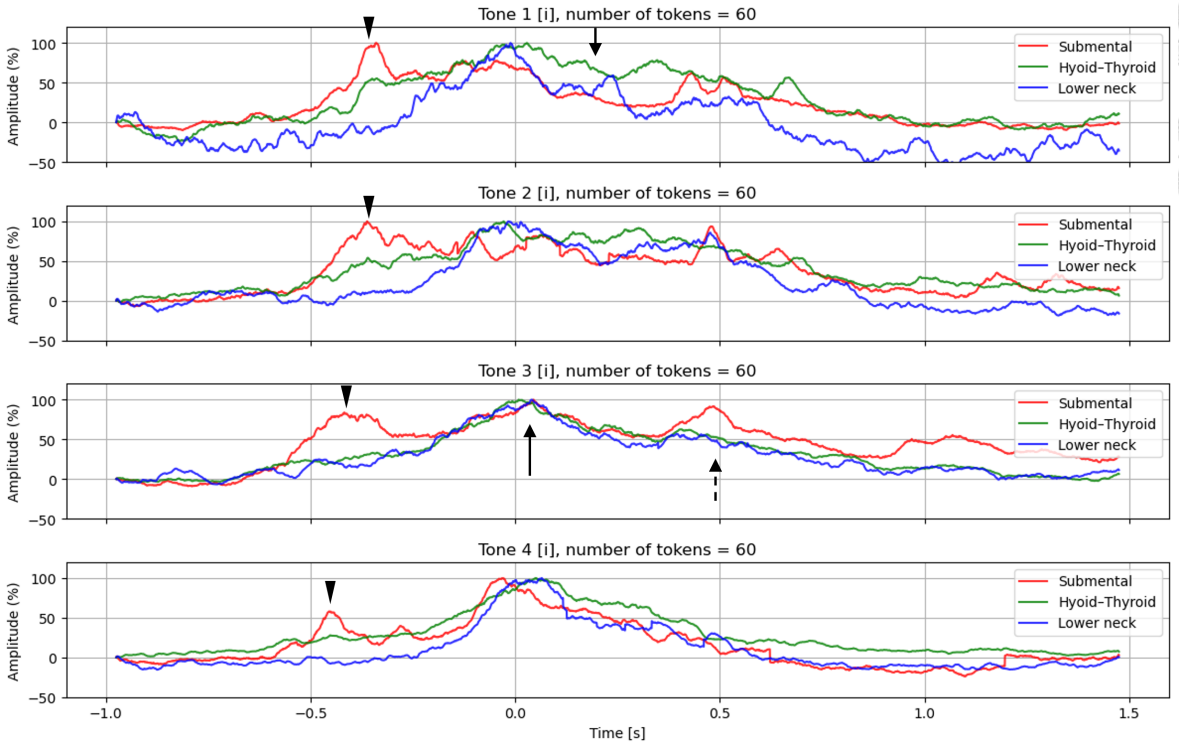


Figure 4.8: Electromyography results for [i] from the elderly group.

a gradual decline in lower neck channel activity after the voice onset. Local peaks, marked with an arrow, in the submental and hyoid-thyroid channels were observed approximately 0.2s after the voice onset. In Tone 4, activation declined gradually after the voice onset in the submental and lower neck channels while a local spike, annotated with an arrow, was observed in the hyoid-thyroid channel around 0.2s after voice onset.

doi:10.6342/NTU202503180

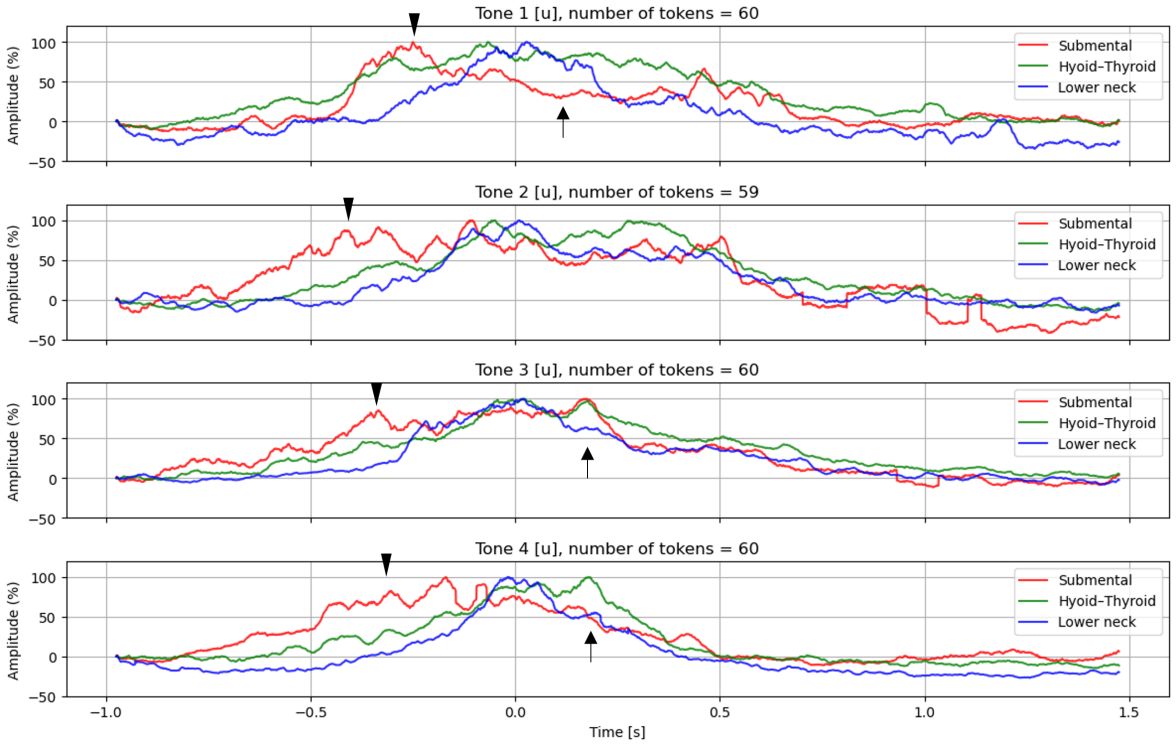


Figure 4.9: Electromyography results for [u] from the elderly group.

doi:10.6342/NTU202503180

Chapter 5 Discussion

This study asked two research questions. The first question was whether hyolaryngeal approximation patterned differently over time across age groups under different linguistic contexts. The first part of the discussion focuses on explaining the differences in hyolaryngeal approximation observed between age groups, considering its relationship to laryngeal height and potential effects of tonal variation and muscle aging. This corresponds directly to the ultrasound experiment examining hyolaryngeal approximation (Section 4.1).

The second question was whether activation patterns of the relevant musculature differed between the two age groups. This question is approached through an examination of the electromyography activity (Section 4.2), with a focus on interpreting how muscular activation patterns during vowel and tone production differ between age groups.

In the third section, the discussion compares the ultrasound findings with prior research on hyolaryngeal approximation in deglutition, highlighting future research directions and potential clinical relevance. Finally, the limitations of the current study are discussed in the last section of this chapter.

doi:10.6342/NTU202503180

5.1 Hyolaryngeal Approximation During Vowel and Tone Production

In this section, the results of *nHT* are explored in the context of the previous literature which largely focused on the hyoid and the larynx. To draw insight from previous studies, functional principal component analysis (fPCA) was conducted to understand the relation between laryngeal/hyoid movements and *nHT* (Sections 5.1.1 and 5.1.2). The advantages of measuring hyolaryngeal approximation over measuring laryngeal height was highlighted in Section 5.1.3. It is then argued in Section 5.1.4 that the differences in the two age groups may be explained by age-related changes in the laryngeal musculature and variation in tonal production.

5.1.1 Contribution of Hyoid and Laryngeal Movements to Hyolaryngeal Approximation during Speech Production

This section examines the respective contributions of hyoid and thyroid cartilage movements to hyolaryngeal approximation along the rostral-caudal axis during speech production. This axis corresponds to the x-axis in the ultrasound images (See Section 3.5). Movements along other axes are not analyzed, as they may not be adequately captured given the current positioning of the ultrasound probe.

The hyoid bone movement along the x-axis, referred to as *xH*, was calculated by subtracting its initial x-axis position from its position in each frame (Equation 5.1). An analogous calculation was performed for the thyroid cartilage to determine its x-axis displacement over time (Equation 5.2). *i* represents the frame number in the equations.

doi:10.6342/NTU202503180

$$\mathrm{xH}_i = \mathrm{Hyoid}_{\mathrm{x.axis},i} - \mathrm{Hyoid}_{\mathrm{x.axis},1} \tag{5.1}$$

$$\mathrm{xT}_i = \mathrm{Thyroid}_{\mathrm{x.axis},i} - \mathrm{Thyroid}_{\mathrm{x.axis},1} \tag{5.2}$$

The ultrasound data in this study comprised temporal sequences of articulatory movement during speech production. To assess whether the time course of *nHT* followed patterns similar to hyoid or laryngeal movement within this dataset, a post-hoc fPCA was conducted with the `fda` package (Ramsay, 2025) in R (R Core Team, 2024). Only productions with data points for all 150 frames were included in the analysis.

xH, *xT*, and *nHT* trajectories for each production were smoothed using 15 B-splines and converted into functional data objects to perform fPCA. Subsequently, canonical correlation analysis (CCA) was performed on the extracted scores of the first five fPCs to examine whether the patterns of variation in *xH* and *xT* movement were systematically correlated to those observed in *nHT*.

The fPCA results showed a similar pattern of change over time for *xH*, *xT*, and *nHT* (Figure 5.1). fPC1 contributed for 74%, 72% and 60% of variation observed in *xH*, *xT*, and *nHT*, respectively. The first five principal components account for over 90% of variation for all three parameters.

The five canonical correlations computed between the first five fPC scores of *xH* and *nHT* were 0.39, 0.26, 0.20, 0.19, and 0.09, indicating a negligible to moderate correlation between the two variables. A Wilks' Lambda test for the full set of canonical correlations (CC1 to CC5) suggested that the relationship was statistically significant ($p < .001$).

Meanwhile, the five canonical correlations between *xT* and *nHT* were 0.62, 0.48, 0.38,

doi:10.6342/NTU202503180

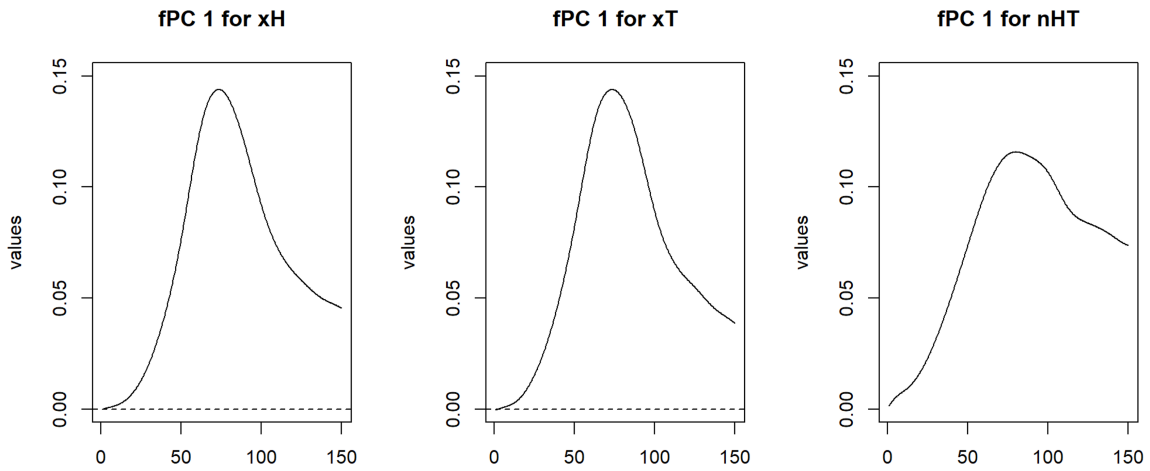


Figure 5.1: fPC1 for *xH*, *xT*, and *nHT*. The x-axis represents the frame number.

0.35, and 0.30, suggesting moderate to moderately strong correlation. The corresponding Wilks' Lambda test for the whole set of canonical correlations (CC1 to CC5) also indicated a statistically significant relationship ($p < .001$). These results suggest that thyroid cartilage movement exhibits a stronger association with hyolaryngeal approximation than hyoid movement does.

5.1.2 Relating *nHT* to Laryngeal Height during Speech Production

The dynamics of hyolaryngeal approximation during speech production have been rarely addressed in the literature. By contrast, previous studies have shown that changes in hyoid bone and laryngeal position are associated with pitch variation and vowel environments (Honda, 1983, 1995; Honda et al., 1999; Menon & Shearer, 1971).

To draw insight from previous studies, it may be informative to examine the current *nHT* findings with former studies on hyoid and laryngeal movements. While such positional changes of the two structures may occur along multiple anatomical axes, the data in the present

doi:10.6342/NTU202503180

study mainly captured the movements along the rostral-caudal and depth axes. Fortunately, both pitch-related and vowel-related variations have known correlates in the rostral-caudal dimension, particularly for laryngeal movements.

Section 5.1.1 showed that in the current dataset, thyroid cartilage movement along the rostral-caudal axis showed a stronger correlation with *nHT* than did hyoid movement in the same dimension. The calculation of *nHT* relied solely on the location of the hyoid bone and thyroid cartilage within the Cartesian plane defined by the ultrasound scan plane (Equation 3.3). Anatomically, thyroid movement occurs primarily along the rostral-caudal axis rather than the depth axis. The moderate to moderately strong canonical correlation between the fPC scores of *xT* and *nHT* may suggest *xT* is an important contributor to changes in *nHT*. The following sections therefore examine *nHT* patterns, drawing on findings from previous studies on laryngeal height in speech production.

5.1.3 Hyolaryngeal Distance as Compared to Laryngeal Height

To contextualize *nHT* in relation to previous ultrasound studies on larynx height during speech production, it is important to first highlight its methodological advantages. The advantages distinguish measuing hyolaryngeal distance from only measuring laryngeal height, providing an straightforward way to compare across speakers and making integration with other recording devices easier.

Previous studies mainly focused on larynx height during speech production (Moisik et al., 2014; Poh, 2024). Nevertheless, as discussed in Moisik et al. (2014), the optimal way to normalize larynx height data remains unclear. By contrast, the hyolaryngeal distance in the current study provides a straightforward approach to normalize data and facilitate inter-

doi:10.6342/NTU202503180

participant comparison. The hyolaryngeal distance is calculated from the positions of the hyoid bone and thyroid cartilage on each frame. The distance between the two targets during the resting state may serve as a baseline to normalize the production data, *i.e. nHT*, as calculated with Equation 3.3. Analogous methods were widely used in studies of deglutition to show the extent of hyolaryngeal approximation (Bahia & Lowell, 2023; Y.-L. Huang et al., 2009; Matsuo & Matsuyama, 2021), supporting its feasibility.

On the other hand, probe stability is crucial to ultrasound studies. To enhance probe stability for laryngeal ultrasound, Poh (2024) employed probe and head stabilizers. However, such stabilizers may restrict articulatory movements and limit the use of additional equipment, as in Moisik et al. (2014), where simultaneous laryngoscopy was performed without stabilization. Since the hyolaryngeal distance is determined by the locations of the hyoid bone and the thyroid cartilage, it may be less sensitive to probe displacement along the rostral-caudal axis as sliding along the rostral-caudal axis would minimally affect the distance between the two structures. This makes it a potentially more robust metric, particularly when the ultrasound probe is hand-held in order to reduce the interference imposed by stabilizers or to accommodate concurrent use of other recording devices.

5.1.4 Differences in Hyolarylgeal Approximation across Age Groups

Having discussed the correlation between laryngeal height and *nHT* (Section 5.1.2), the discussion now turns to significant differences in hyolaryngeal approximation across age groups. Factors related to muscular aging and age-related tonal variation are taken into consideration.

Beginning with Tone 1, a high level tone, no significant differences were found across all

doi:10.6342/NTU202503180

models (Figures 4.1, 4.2, 4.3). Moisik et al. (2014) observed laryngeal elevation during Tone 1 and interpreted this as a compensatory strategy to maintain pitch in response to decreasing subglottal pressure. Based on this interpretation, one might expect increased hyolaryngeal approximation towards the end of Tone 1 productions. The elevated laryngeal posture and the associated increase in hyolaryngeal approximation could be physiologically demanding for the elderly speakers due to age-related muscular changes. However, no significant differences in the degree of hyolaryngeal approximation between the young adult and elderly speakers were observed for any Tone 1 model. This suggests that compensating for decreasing subglottal pressure may not be sufficiently demanding to lead to measurable differences in hyolaryngeal approximation between the two groups.

By contrast, for Tone 2 productions, young adult speakers exhibited significantly greater hyolaryngeal approximation in [a] and [i] (Figures 4.1 and 4.2). Sanders (2008) showed that older speakers tended to produce Tone 2 with a rising contour, whereas younger speakers were more likely to produce Tone 2 with a dipping contour. Fon and Hsu (2007) previously examined Tone 2 production in a group of five young male speakers of Taiwan Mandarin whose parents were both native speakers of Taiwan Southern Min. It was revealed that most Tone 2 productions had a dipping contour, with the falling portion occupying approximately 40 to 50% of the production. Meanwhile, previous studies have shown that laryngeal height positively correlates with pitch (Honda, 1995; Honda et al., 1999; Moisik et al., 2014), and differences in tone contours may lead to different laryngeal movement patterns, hence affecting the correlated *nHT* (Section 5.1.1). In other words, the second halves of both dipping and rising Tone 2 productions may correspond to a rise in pitch.

Significantly lower hyolaryngeal approximation among elderly speakers was observed in

doi:10.6342/NTU202503180

regions overlapping with the second halves of [a] and [i] productions (Figures 4.1b and 4.2b). The rising portion again was associated with laryngeal elevation (Honda, 1995; Moisik et al., 2014), and a higher larynx position correlates with greater hyolaryngeal approximation (*i.e.* lower *nHT*). This posture may be more demanding to the elderly speakers due to age-related changes in their musculartur e.

In Tone 1 productions, although laryngeal elevation may help compensate for decreasing subglottal pressure (Moisik et al., 2014), it alone may not significantly affect hyolaryngeal approximation. In comparison, in Tone 2, speakers needed to raise their pitch aside from compensating the reduced subglottal pressure. This may place greater load on the muscles, contributing to the significantly lower hyolaryngeal approximation observed in elderly speakers during the production of Tone 2 [a] and [i].

With regard to Tone 3, significant reduction in hyolaryngeal approximation among elderly speakers was observed in Tone 3 [a] and [i] (figures 4.1, 4.2). Moisik et al. (2014) observed two strategies to pronouncing Mandarin Tone 3, described as larynx lowering and larynx elevation with constriction. Despite the varied patterns observed in the raw data, it helps to further explore how different production strategies may have an effect on hyolaryngeal gestures.

A heuristic post-hoc classification was applied to Tone 3 productions with data from all 150 frames to determine whether a production involved a raised or lowered laryngeal gesture. The *xT* values between 0% and 100% *nor_time* were extracted for each production and subtracted with the *xT* value at time 0% in the same production. The result was the laryngeal movement along the x-axis relative to the laryngeal position at time 0%, henceforth referred to as *xT'*. *xT'* values for each production were summed.

doi:10.6342/NTU202503180

Recalling that the x-axis in the ultrasound frames pointed towards the caudal direction (Section 3.5.1), a summed value greater than 0 indicated overall caudal movement, and the production was therefore classified as a "lowered" production. On the contrary, those productions with summed *xT'* values below 0 were classified as "raised" productions. The post-hoc grouping result was shown in Table 5.1. The elderly speakers in this dataset seemed to prefer raised laryngeal gestures in Tone 3 productions, while young adult speakers preferred lowered laryngeal gestures.

	Young Adult Speakers		Elderly Speakers	
	Raised	Lowered	Raised	Lowered
[a]	27	61	25	15
[i]	22	72	28	16
[u]	24	69	35	19

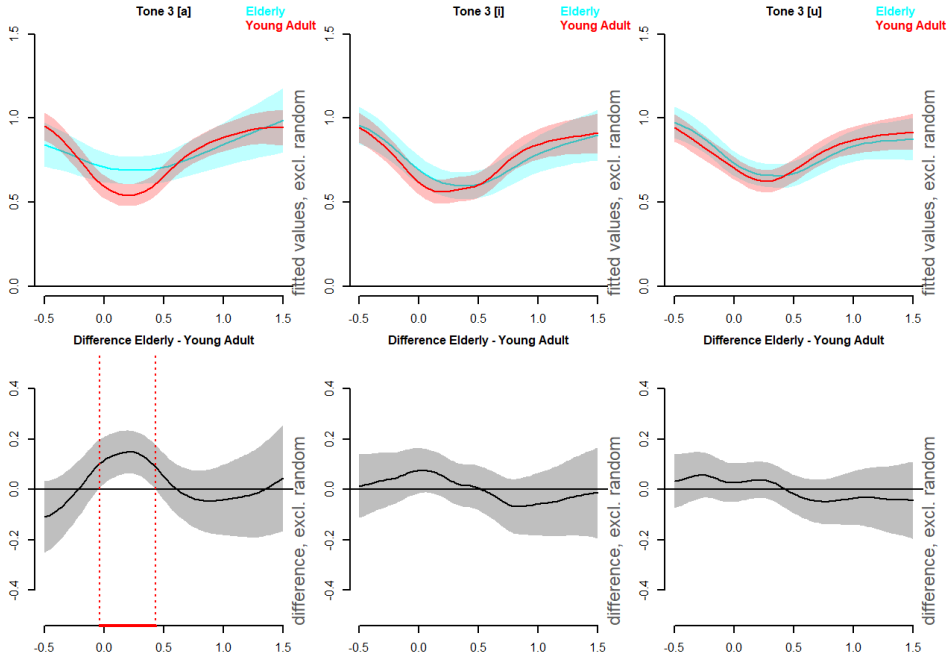
Table 5.1: Post-hoc classification of Tone 3 productions.

The subgroup data were modeled with generalized additive mixed modeling for each vowel condition. The models were the same as described in Formula 4.1. The *Time* was modeled with *abs_time* and *nor_time*. The model summaries, including parametric coefficients, approximate significance of smooth terms, AIC, and BIC values, are presented in Appendix B.

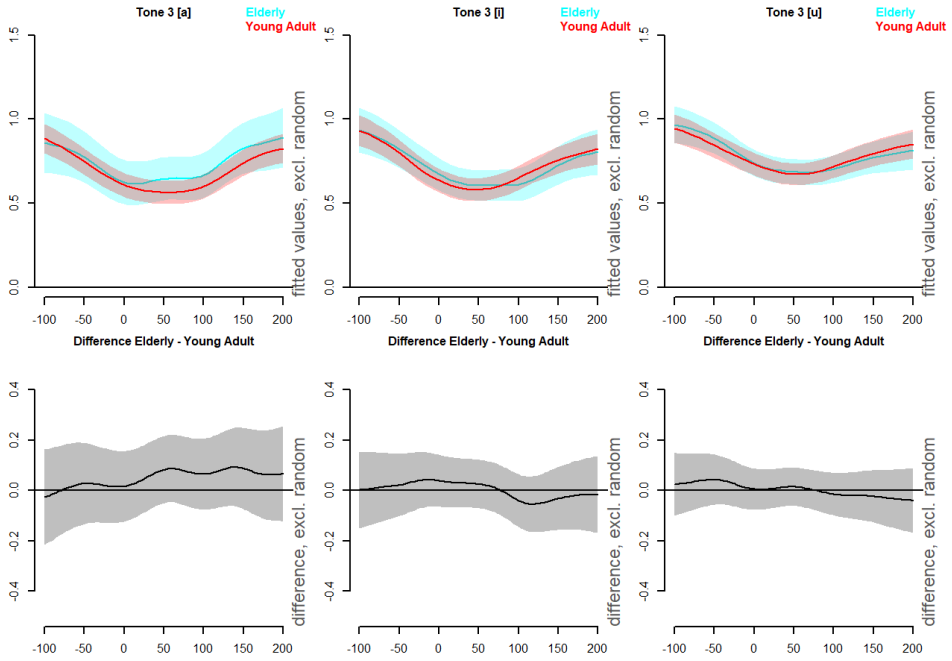
Figure 5.2 showed the results for the "raised" Tone 3 subgroup. While the *abs_time* models showed a significant time window for [a], it was not observed in the *not_time* models. Meanwhile, for the "lowered" Tone 3 subgroup, a consistent window of significant difference was found for both *abs_time* and *nor_time* models for [a] (Figure 5.3).

This result may imply that young adult speakers exhibited greater hyolaryngeal approximation during larynx-lowering Tone 3 [a] productions and potentially also during larynx-

doi:10.6342/NTU202503180



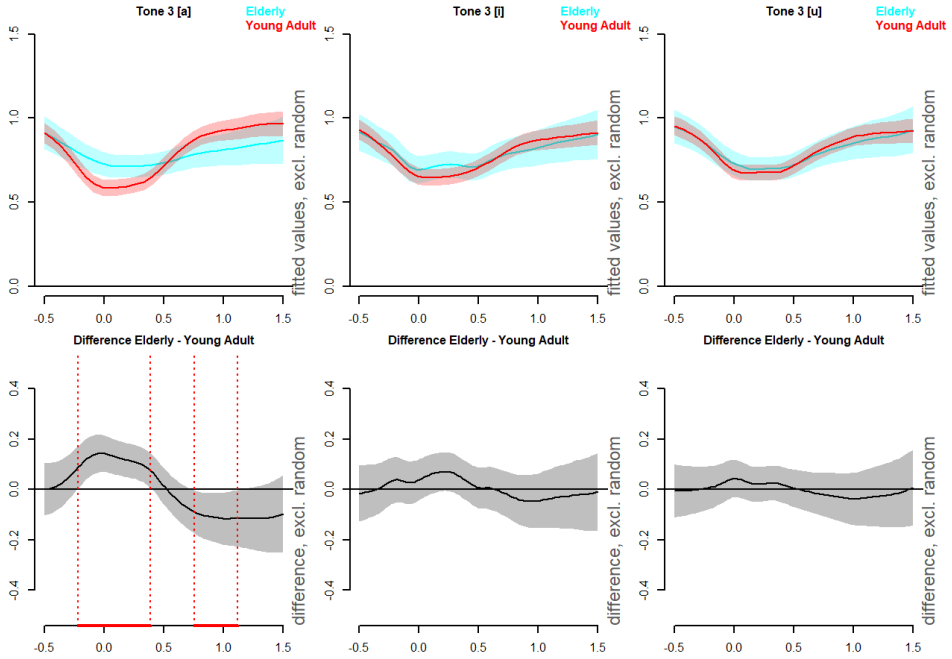
(a) Raised Tone 3 models with *abs_time* reference.



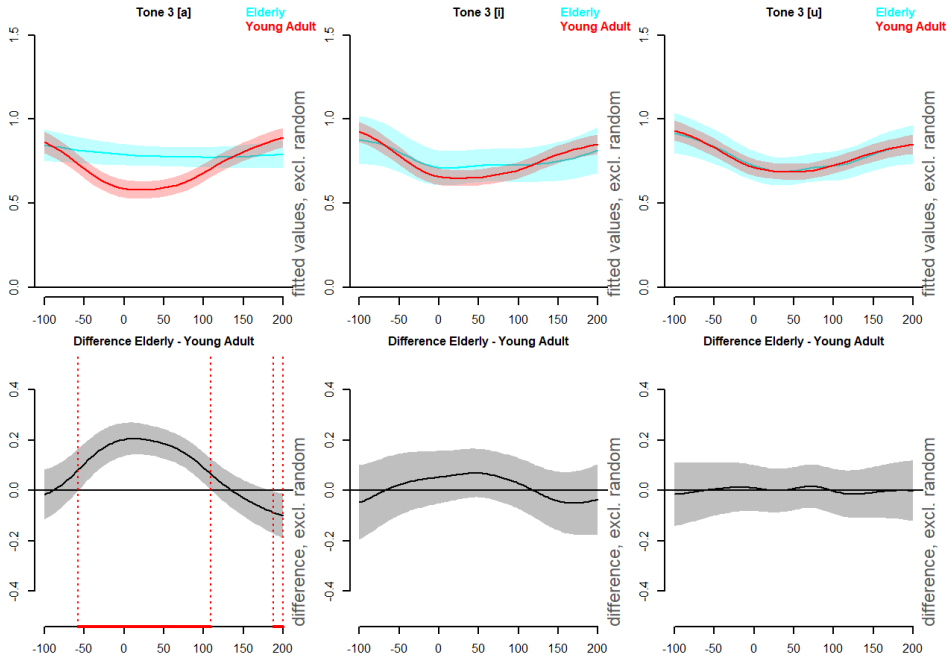
(b) Raised Tone 3 models with *nor_time* reference.

Figure 5.2: Generalized additive mixed model results for raised Tone 3 productions

doi:10.6342/NTU202503180



(a) Lowered Tone 3 models with *abs_time* reference.



(b) Lowered Tone 3 models with *nor_time* reference.

Figure 5.3: Generalized additive mixed model results for lowered Tone 3 productions

doi:10.6342/NTU202503180

raising Tone 3 [a] productions. This appeared to be a vowel-specific effect as similar patterns were not observed in [i] and [u] productions. Further studies are warranted to more comprehensively investigate Tone 3 production patterns and better understand the contributions of multiple factors to speech production.

Regarding Tone 4, a falling tone, significant differences in the degree of hyolaryngeal approximation overlapped with the voice onset (Figure 4.1b, 4.2b). The onset of Tone 4 is associated with a high pitch target. Since higher pitch is associated with laryngeal elevation (Honda, 1995; Moisik et al., 2014) and greater hyolaryngeal approximation (Section 5.1.1), the onset of Tone 4 is linked to a high laryngeal position and may impose greater demands on the aging musculature. On the other hand, Hsu and Tse (2009) showed that elderly speakers tended to have a larger pitch range for Tone 4. Nevertheless, since the gross contour of the tone does not change, laryngeal and *nHT* trends would be largely similar.

In addition to tone-specific differences, vowel-specific differences were also observed. In contrast to the [a] and [i] environments, there was only one short time window of signifcant difference in the Tone 2 [u] *abs_time* model. This effect was not observed in the corresponding model fitted with *nor_time* and was not aligned to the time lock point, which may suggest that the effect was not robust or that the effect may not be sufficiently characterized by the statistical models and hypotheses in this study.

These vowel-specific differences may be partially explained by physiological factors. Ewan and Krones (1974) reported a general low laryngeal position for [u] while a relatively higher laryngeal position was found for [a] and [i]. The laryngeal position for [i] was reportedly higher than [a], though the reverse pattern was also observed. The higher laryngeal position may pose greater challenge to the aging extralaryngeal musculature, leading to more

doi:10.6342/NTU202503180

significant differences showing reduced hyolaryngeal approximation in the elderly speakers in vowel environments [a] and [i].

Exploratory models were fitted to see if there was any interaction between vowel context and age group. However, the interaction was found to be insignificant and removed. The results of these models are presented in Appendix C.

A significant difference was observed in time segments that did not correspond to the voicing phase of tone production for Tone 3 [a] productions (Figure 4.1a). Specifically, elderly speakers showed greater hyolaryngeal approximation. The difference occurred apporximately 0.7-0.9s after voice onset. This time range was most likely after the voicing had ended, judging from the avergae duration of Tone 3 production of this dataset (Table 5.2). A summary of the duration of each tone targets is presented in Table 5.2.

	Young Adult		Elderly	
	Mean	SD	Mean	SD
Tone 1	0.4985	0.1796	0.4579	0.2212
Tone 2	0.5034	0.1548	0.5336	0.2236
Tone 3	0.4540	0.1764	0.5589	0.2074
Tone 4	0.3641	0.1401	0.3208	0.1296

Table 5.2: Characteristic of the duration of each tone for both age groups in seconds. *SD* refers to standard deviation.

The elderly speakers tended to produce longer Tone 3s than the young adult speakers. Longer productions suggest that elderly speakers sustained articulatory engagement for a longer time. Hence, the hyolaryngeal approximation may have been sustained longer and the relaxation of the articulators may have occurred later, leading to the significantly greater hyolaryngeal approximation observed in Tone 3 [a] productions approximately 0.7-0.9s after voice onset.

doi:10.6342/NTU202503180

While not statistically significant, a similar tendency was observed in the *nHT* model plots for other vowel-tone combinations where the hyolaryngeal approximation seemed to be greater after the voicing ended (after 100% in Figures 4.1b, 4.2b, and Tones 1 and 4 in Figure 4.3b). A potential contributing factor would be the reduced relaxation rate associated with muscle aging (Callahan & Kent-Braun, 2011; Hunter et al., 2016). After the voicing has ended, it might require longer time for elderly speakers to return to the resting position.

Another factor worthy of considering is the reduced capacity for fine motor control. Age-related increases in motor-unit size secondary to motor neuron apoptosis and/or alterations in common synaptic input may exert differential effects on fine motor control. (Hunter et al., 2016). This would potentially increase the instability of articulatory movements and modulate hyolaryngeal approximation.

5.2 Electromyographic Patterns and Their Potential Correlates in Speech Production

This section discusses how the electromyography patterns may correlate with Taiwan Mandarin tone and vowel production. The limitation of surface electromyography is discussed afterwards. We first focus on the results of the young adult speakers.

Regarding the results of young adult speakers, both tonal effects and vowel effects were observed (Figures 4.4, 4.5, 4.6). One apparent effect was found in Tone 2 productions across all three vowel environments. Signals from both submental and hyoid-thyroid channels remained robust for a longer time while lower neck channel signals diminished shortly after the voice onset. The submental and hyoid-thyroid channels included activation from the geniohyoid muscle and the thyrohyoid muscles, respectively, which were both reported to be

doi:10.6342/NTU202503180

associated with pitch raising (Honda, 1983, 1995; Sawashima et al., 1973). This was consistent with pitch raising that could be observed in both raising and dipping contours in Tone 2 productions.

On the other hand, Tone 4 productions from young adult speakers consistently showed a peak in the lower neck across all vowels. The peak occurred shortly after the voice onset, as annotated with the arrows in Figures 4.4, 4.5, and 4.6. This is in accordance with the findings in Boysson-Bardies et al. (1986) where the activation of the sternohyoid muscle peaked slightly after the F0 began to decrease in Tone 4 targets. The lower neck channel captured activation from the sternohyoid and sternothyroid muscles, which were reported to be correlated with pitch lowering (Boysson-Bardies et al., 1986; Erickson, 1993; Honda, 1995).

Turning to vowel effects, [a] productions showed a local peak approximately 0.2s prior to the voice onset in all three channels (the arrowheads in Figure 4.4). This activation is possibly associated with jaw opening, which is a distinguishing articulatory feature of [a] compared to the high vowels [i] and [u]. Previous literature revealed that the activation of submental muscles, including the anterior digastric, thyrohyoid, sternohyoid muscles were correlated with jaw opening, either involved in mandible depression or hyoid stabilization (Bérzin, 1995; Honda, 1995; Moore et al., 2018; Sawashima et al., 1973).

The discussion now turns to the electromyography results for elderly speakers (Figures 4.7, 4.8, 4.9). Tonal effects appeared less consistent and less pronounced across the three vowel environments. A sustained activation of submental and hyoid-thyroid channels was seen in Tone 2 [a], which may reflect laryngeal elevation assocaited with the raising or dipping pitch contour. A similar prolonged activation of submental and hyoid-thyroid channels was seen in

doi:10.6342/NTU202503180

Tone 3 [a], which may reflect a greater use of dipping contours in elderly speakers (Sanders, 2008). For Tone 4 [a], the lower neck channel peaked slightly after the voice onset, which again reflects the falling pitch contour.

For the high vowels [i] and [u], tone-specific patterns were not as obvious. A possible explanation would be that tone-specific patterns were complicated by the lingual activation patterns, which showed a slower and potentially less coordinated relation with the laryngeal gestures.

On the other hand, for vowel effects, local peaks in the submental and hyoid-thyroid channels were observed prior to the voice onset in [i] and [u] productions (arrowheads in Figures 4.8 and 4.9). The two channels captured activity from muscles located in the submental region and the supero-lateral neck. Previous studies revealed that the posterior genioglossus and geniohyoid muscles play a role in producing high vowels (Gick et al., 2013; Honda, 1983; Waltl & Hoole, 2008). These contributions may be reflected in the activity in the submental channel, while the signals from the hyoid-thyroid channel might have represented the stabilizing function from the thyrohyoid muscles. Interestingly, this high vowel effect was not observed in the young adult speakers, suggesting that the two age groups may have employed different articulatory strategies for the production task.

Previous studies have shown slower articulator movements and reduced inter-articulator synchronization in elderly speakers. With electromagnetic articulography, Mücke et al. (2020) compared German speakers of different ages and showed that the duration from lingual gesture onset to target attainment was longer for the older speakers. Bilodeau-Mercure and Tremblay (2016), on the other hand, examined the accuracy of nasal vowel production in French speakers of different ages. The target sequences in the study required speakers to alternate

doi:10.6342/NTU202503180

rapidly between oral consonants and nasal vowels. Elderly participants produced significantly more errors. The authors interpreted the results as indicative of reduced velum control and decreased synchronization of articulators during speech.

Returning to the current investigation, the earlier activation in the high vowel targets might reflect the extended duration required for the execution of the lingual postures. The lack of visible electromyography activation corresponding to lingual postures may result from the fact that the low tongue posture in [a] was associated with other sets of articulatory gestures involving the hyoglossus and jaw-opening movements (Gick et al., 2013; Waltl & Hoole, 2008). The hyoglossus muscle was unlikely to be taken up by the electromyography electrodes used in the current experiment. Jaw-opening-related signals may be recorded by the submental channel. Nevertheless, such an effect was not observed at the same time point as the high-vowel-related signals, suggesting that tongue elevation and jaw-opening are organized differently. These findings suggest that elderly speakers may have initiated lingual and laryngeal gestures with greater temporal asynchrony, whereas young adult speakers may have engaged these gesturesin a more temporally coordinated fashion.

5.3 Comparison with Deglutition Studies and Implications

A direct comparison of age-related effects on hyolaryngeal approximation was reported in Bahia and Lowell (2023), where healthy individuals were divided into younger and older groups. No significant differences in hyolaryngeal approximation were observed between age groups during normal saliva swallowing and the maximum relative change in approximation during swallowing approached 50-60%.

In contrast, the current study examined the temporal profile of hyolaryngeal approxima-

doi:10.6342/NTU202503180

tion during speech production as opposed to maximum approximation. From a functional standpoint, swallowing may place greater demands on cervical and submental musculature, given its vital role in airway protection and bolus propulsion. This functional distinction may partly explain why, in our study, model-fitted hyolaryngeal approximation remained below 50% across all vowel-tone combinations, regardless of age. Nevertheless, significantly reduced approximation was observed in the elderly group in specific vowel-tone contexts.

One potential explanation for the discrepancy lies in methodological differences. Bahia and Lowell (2023) focused on maximum approximation, while the current study examined dynamic changes across the entire speech production timeline. This contrast highlights the potential importance of temporal resolution in detecting subtler age-related changes. However, it is also possible that fitting productions into statistical models may have attenuated extreme values of hyolaryngeal approximation, as speakers may reach maximum approximation at different time points during the production. Thus, while the current approach offers a more comprehensive temporal perspective, it may underestimate peak articulatory extremes.

Future studies may explore how aging may differentially affect speech production and deglutition, with a particular focus on temporal dynamics and the precision of motor control within subcomponents of each task. Such investigations could advance our understanding of their shared neuromuscular mechanisms. The findings of this study may have clinical implications; comparative studies for speech production in nondysphagic versus dysphagic individuals may help identify disease-related changes in hyolaryngeal gestures during speech.

doi:10.6342/NTU202503180

5.4 Limitations of the Present Study

5.4.1 Limitations of Ultrasound Imaging

As described in the methodology (Section 3.7.1), only productions for which all 150 frames were retained were included in the analysis. This process, while necessary to ensure annotation quality and the integrity of the generalized additive mixed models, may introduce an unintended bias.

The figures in Appendix D illustrate where and how severe each vowel-tone combination was affected by frames that could not be annotated. The discarded frames generally occurred in the mid-portion, overlapping with the onset of articulatory movements. Despite the biased pattern, this was not unexpected, as rapid movement, overlapping of the hyoid bone and the thyroid cartilage, and structures moving out of frame were major causes of such frame loss other than poor image quality. All of the listed issues are presumably more likely to happen during speech production.

Critical to the current study was the frame loss due to overlapping of the hyoid bone and the thyroid cartilage in which the hyoid bone cast an acoustic shadow obscuring deeper structures. This may selectively remove productions with greater hyolaryngeal approximation, potentially resulting in an underestimation of hyolaryngeal approximation during speech production. Moreover, the frame loss appeared to be more pronounced in the elderly group, suggesting an uneven effect.

This frame loss was at least partially a result of the inherent nature of ultrasound imaging. Alternative imaging methods, including computed tomography, videofluorography, or magnetic resonance imaging, may be considered to overcome some of the listed limitations to

doi:10.6342/NTU202503180

provide a more comprehensive understanding of hyolaryngeal approximation during speech production.

On the other hand, the placement of the ultrasound probe in the current study was defined as deviating within 30 degrees from the midsagittal line (Figure 3.1). This definition did not rely on anatomical landmarks, and may introduce some variability in probe placement, which may modulate the hyolaryngeal measurements. The lesser cornu of the hyoid bone may be a potential anatomical marker for more consistent ultrasound probe placement.

5.4.2 Limitation of Surface Electromyography

Despite the findings discussed above, it should be noted that surface electromyography results may not be associated with the activation of specific muscles but reflects all muscular activation in the vicinity of the electrodes. Moreover, electromyography activation does not equate actual movements of the articulators. In addition, the surface electromyography results presented in this study represents data summed from multiple tokens and were normalized to facilitate visual inspection. Therefore, the visualized amplitudes do not reflect raw signal magnitude and are intended only to illustrate relative timing and pattern differences across conditions.

doi:10.6342/NTU202503180

Chapter 6 Conclusion

This study used ultrasound to examine whether the extent of hyolaryngeal approximation varies across age groups under different vowel-tone combinations in Taiwan Mandarin. In addition, surface electromyography was applied to examine the activation patterns of the musculature related to hyolaryngeal movements.

The ultrasound results revealed hyolaryngeal approximation during speech production. Importantly, significant differences in the extent of hyolaryngeal approximation were found between young adult and elderly speakers. These differences may reflect the combined effects of physiological aging and cross-generational variation in tonal production. The surface electromyography results revealed recurring tone- and vowel-related patterns. However, the patterns were not entirely consistent between young adult and elderly speakers. Notably, the elderly speakers exhibited patterns suggestive of greater temporal asynchrony between lingual and laryngeal gestures.

Ultrasound imaging and surface electromyography allowed for non-invasive examination of hyolaryngeal gestures and muscle activation patterns. However, limitations including ultrasound frame loss due to structure overlap or rapid movement may lead to an underestimation of hyolaryngeal approximation in the elderly group. On the other hand, surface electromyo-

doi:10.6342/NTU202503180

graphy may not precisely identify the activation of specific muscles and the activation does not necessarily result in actual movements of articulators.

This work contributes to the understanding of hyolaryngeal gestures in speech production and their age-related variability. It also informs future comparisons with deglutition, and highlights the exploratory potential of speech gestures in clinical contexts.

doi:10.6342/NTU202503180

Bibliography

Abd-El-Malek, S. (1955). The part played by the tongue in mastication and deglutition. *Journal of anatomy*, *89*(Pt 2), 250.

Ahn, S. Y., Cho, K. H., Beom, J., Park, D. J., Jee, S., & Nam, J. H. (2015). Reliability of ultrasound evaluation of hyoid–larynx approximation with positional change. *Ultrasound in Medicine & Biology*, *41*(5), 1221–1225. https://doi.org/10.1016/j.ultrasmedbio.2014.12.010

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Ambrosi, D., & Lee, Y.-T. (2021). Rehabilitation of swallowing disorders. In M. Cifu David X. (Ed.), *Braddom's physical medicine and rehabilitation* (Sixth Edition, 53–67.e2). https://doi.org/10.1016/B978-0-323-62539-5.00003-5

Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *The Journal of the Acoustical Society of America*, *63*(1), 211–222. https://doi.org/10.1121/1.381716

Bahia, M. M., & Lowell, S. Y. (2023). Hyolaryngeal movement during normal and effortful swallows determined during ultrasonography. *Journal of Speech, Language, and*

doi:10.6342/NTU202503180

Hearing Research, *66*(10), 3856–3870. https://doi.org/10.1044/2023_JSLHR-23-00088

Belafsky, P. C., Mouadeb, D. A., Rees, C. J., Pryor, J. C., Postma, G. N., Allen, J., & Leonard, R. J. (2008). Validity and reliability of the eating assessment tool (eat-10) [PMID: 19140539]. *Annals of Otology, Rhinology & Laryngology*, *117*(12), 919–924. https://doi.org/10.1177/000348940811701210

Belyk, M., Eichert, N., & McGettigan, C. (2021). A dual larynx motor networks hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1840), 20200392. https://doi.org/10.1098/rstb.2020.0392

Bérzin, F. (1995). Electromyographic analysis of the sternohyoid muscle and anterior belly of the digastric muscle in jaw movements. *Journal of Oral Rehabilitation*, *22*(6), 463–467. https://doi.org/10.1111/j.1365-2842.1995.tb00802.x

Bilodeau-Mercure, M., & Tremblay, P. (2016). Age differences in sequential speech production: Articulatory and physiological factors. *Journal of the American Geriatrics Society*, *64*(11), e177–e182. https://doi.org/10.1111/jgs.14491

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

Boysson-Bardies, B. d., Arabia-Guidet, C., Hallé, P. A., & Sagart, L. (1986). Tone production in modern standard chinese : An electromyographic investigation. *Cahiers de Linguistique - Asie Orientale*, *15*(2), 205–221. https://doi.org/10.3406/clao.1986.1204

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

doi:10.6342/NTU202503180

Callahan, D. M., & Kent-Braun, J. A. (2011). Effect of old age on human skeletal muscle force-velocity and fatigue properties [PMID: 21868683]. *Journal of Applied Physiology*, *111*(5), 1345–1352. https://doi.org/10.1152/japplphysiol.00367.2011

Carlson, A. (2021). *Measuring thyrohyoid space with ultrasonography: A feasibility study* [Master's thesis] [Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-06-21]. https://www.proquest.com/dissertations-theses/measuring-thyrohyoid-space-with-ultrasonography/docview/2549719406/se-2

Castillo-Allendes, A., Searl, J., Vergara, J., Ballentine, N., Ebdah, S., Rameau, A., & Hunter, E. J. (2025). Voice meets swallowing: A scoping review of therapeutic connections. *American Journal of Speech-Language Pathology*, *34*(2), 877–907. https://doi.org/10.1044/2024_AJSLP-24-00194

Chao, Y. R. (1930). Ə sistim əv "toun-letəz". *Le Maître Phonétique*, *8 (45)*(30), 24–27. Retrieved May 10, 2025, from http://www.jstor.org/stable/44704341

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken chinese*. Harvard University Press.

Clancy, E. A., Morin, E. L., Hajian, G., & Merletti, R. (2023). Tutorial. surface electromyogram (semg) amplitude estimation: Best practices. *Journal of Electromyography and Kinesiology*, *72*, 102807. https://doi.org/10.1016/j.jelekin.2023.102807

Dan, D., Feng, S., & Shinan, L. (2006). The contrast on tone between putonghua and taiwan mandarin. *ACTA ACUSTICA*, *31*(6), 536–541. https://doi.org/10.15949/j.cnki.0371-0025.2006.06.011

Dusan, S. (2007). Vocal tract length during speech production. *INTERSPEECH*, 1366–1369.

doi:10.6342/NTU202503180

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology*, *23*(2), 157–191. https://doi.org/10.1017/S095267570600087X

Erickson, D. (1993). Laryngeal muscle activity in connection with thai tones. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, *27*, 135–149.

Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, *2*(4), 327–335. https://doi.org/10.1016/S0095-4470(19)31302-6

Fon, J., & Chiang, W.-Y. (1999). What does chao have to say about tones? —a case study of taiwan mandarin / 赵氏声调系统与声学之联结及量化－以台湾地区国语为例. *Journal of Chinese Linguistics*, *27*(1), 13–37. Retrieved May 10, 2025, from http://www.jstor.org/stable/23756742

Fon, J., & Hsu, H.-j. (2007). Positional and phonotactic effects on the realization of dipping tones in taiwan mandarin. *Phonology and Phonetics, Tones and Tunes*, *2*, 239–269.

Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory phonetics*. John Wiley & Sons Inc.

Groher, M. E. (2016). Chapter 2 - normal swallowing in adults. In M. E. Groher & M. A. Crary (Eds.), *Dysphagia (second edition)* (Second Edition, pp. 19–40). Mosby. https://doi.org/10.1016/B978-0-323-18701-5.00002-6

Hiiemae, K. M., & Palmer, J. B. (2003). Tongue Movements in Feeding and Speech [PMID: 14656897]. *Critical Reviews in Oral Biology & Medicine*, *14*(6), 413–429. https://doi.org/10.1177/154411130301400604

doi:10.6342/NTU202503180

Hiiemae, K. M., Palmer, J. B., Medicis, S. W., Hegener, J., Scott Jackson, B., & Lieberman, D. E. (2002). Hyoid and tongue surface movements in speaking and eating. *Archives of Oral Biology*, *47*(1), 11–27. https://doi.org/10.1016/S0003-9969(01)00092-9

Honda, K. (1983). Relationship between vowel articulation and pitch control. *SPEECH RESEARCH*, 269.

Honda, K. (1995). Laryngeal and extra-laryngeal mechanisms of f0 control. In *Producing speech: Contemporary issues* (pp. 215–232). American Institute of Physics New York.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in f0 control [PMID: 10845244]. *Language and Speech*, *42*(4), 401–411. https://doi.org/10.1177/00238309990420040301

Hoole, P., & Kroos, C. (1998). Control of larynx height in vowel production. *ICSLP*.

Hsu, H.-j., & Tse, J. K.-p. (2009). The tonal leveling of taiwan mandarin: A study in taipei. *Concentric: Studies in linguistics*, *35*(2), 225–244.

Huang, P.-S. (2023). *Reliability and validity of the mandarin version of eating assessment tool-10* [Master's thesis, 臺北市立大學].

Huang, Y.-L., Hsieh, S.-F., Chang, Y.-C., Chen, H.-C., & Wang, T.-G. (2009). Ultrasonographic Evaluation of Hyoid–Larynx Approximation in Dysphagic Stroke Patients. *Ultrasound in Medicine & Biology*, *35*(7), 1103–1108. https://doi.org/10.1016/j.ultrasmedbio.2009.02.006

Huang, Y.-H., & Fon, J. (2011). Investigating the effect of min on dialectal variations of mandarin tonal realization. *ICPhS*, 918–921.

doi:10.6342/NTU202503180

Hunter, S. K., Pereira, H. M., & Keenan, K. G. (2016). The aging neuromuscular system and motor performance [PMID: 27516536]. *Journal of Applied Physiology*, *121*(4), 982–995. https://doi.org/10.1152/japplphysiol.00475.2016

Jones, B. (2003). Aging and neurological disease. In B. Jones (Ed.), *Normal and abnormal swallowing: Imaging in diagnosis and therapy* (pp. 227–242). Springer New York. https://doi.org/10.1007/978-0-387-22434-3_12

Kang, B.-S., Oh, B.-M., Kim, I. S., Chung, S. G., Kim, S. J., & Han, T. R. (2010). Influence of aging on movement of the hyoid bone and epiglottis during normal swallowing: A motion analysis. *Gerontology*, *56*(5), 474–482. https://doi.org/10.1159/000274517

Khoo, H.-l. (2020). A preliminary study of the tonal features of central taiwan mandarin. *Taiwan Journal of Linguistics*, *18*(1).

Kim, Y., & McCullough, G. H. (2008). Maximum hyoid displacement in normal swallowing. *Dysphagia*, *23*(3), 274–279. https://doi.org/10.1007/s00455-007-9135-y

Kuang, J. (2017). Creaky voice as a function of tonal categories and prosodic boundaries. *Interspeech*, 3216–3220.

Kubler, C. C. (1985). The influence of southern min on the mandarin of taiwan. *Anthropological Linguistics*, *27*(2), 156–176. Retrieved May 10, 2025, from http://www.jstor.org/stable/30028064

Kuhl, V., Eicke, B. M., Dieterich, M., & Urban, P. P. (2003). Sonographic analysis of laryngeal elevation during swallowing. *Journal of Neurology*, *250*(3), 333–337. https://doi.org/10.1007/s00415-003-1007-2

Kuo, Y.-H. (2005). *New dialect formation: The case of taiwanese mandarin* [Doctoral dissertation, University of Essex].

doi:10.6342/NTU202503180

Ladefoged, P. (1968). *A phonetic study of west african languages: An auditory-instrumental survey*. Cambridge University Press.

Leonard, R. J., Kendall, K. A., McKenzie, S., Gonçalves, M. I., & Walker, A. (2000). Structural Displacements in Normal Swallowing: A Videofluoroscopic Study. *Dysphagia*, *15*(3), 146–152. https://doi.org/10.1007/s004550010017

Li-Jessen, N. Y. K., & Ridgway, C. (2020). Neuroanatomy of voice and swallowing. In P. A. Weissbrod & D. O. Francis (Eds.), *Neurologic and neurodegenerative diseases of the larynx* (pp. 21–40). Springer International Publishing. https://doi.org/10.1007/978-3-030-28852-5_3

Logemann, J. A., Pauloski, B. R., Rademaker, A. W., Colangelo, L. A., Kahrilas, P. J., & Smith, C. H. (2000). Temporal and biomechanical characteristics of oropharyngeal swallow in younger and older men. *Journal of Speech, Language, and Hearing Research*, *43*(5), 1264–1274. https://doi.org/10.1044/jslhr.4305.1264

Logemann, J. A., Pauloski, B. R., Rademaker, A. W., & Kahrilas, P. J. (2002). Oropharyngeal swallow in younger and older women. *Journal of Speech, Language, and Hearing Research*, *45*(3), 434–445. https://doi.org/10.1044/1092-4388(2002/034)

Lukezic, A., Vojir, T., Cehovin Zajc, L., Matas, J., & Kristan, M. (2017). Discriminative correlation filter with channel and spatial reliability. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Matsuo, T., & Matsuyama, M. (2021). Detection of poststroke oropharyngeal dysphagia with swallowing screening by ultrasonography. *PLOS ONE*, *16*(3), 1–12. https://doi.org/10.1371/journal.pone.0248770

doi:10.6342/NTU202503180

McFarland, P., David H; Tremblay. (2006). Clinical implications of cross-system interactions. *Semin Speech Lang*, *27*(04), 300–309. https://doi.org/10.1055/s-2006-955119

Menon, K. M. N., & Shearer, W. M. (1971). Hyoid positions during repeated syllables. *Journal of Speech and Hearing Research*, *14*(4), 858–864. https://doi.org/10.1044/jshr.1404.858

Miller, N. A., Gregory, J. S., Semple, S. I., Aspden, R. M., Stollery, P. J., & Gilbert, F. J. (2012). The effects of humming and pitch on craniofacial and craniocervical morphology measured using mri. *Journal of Voice*, *26*(1), 90–101. https://doi.org/10.1016/j.jvoice.2010.10.017

Miloro, K. V., Pearson, W. G., & Langmore, S. E. (2014). Effortful pitch glide: A potential new exercise evaluated by dynamic mri. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1243–1250. https://doi.org/10.1044/2014_JSLHR-S-13-0168

Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (sllus). *Journal of the International Phonetic Association*, *44*(1), 21–58. https://doi.org/10.1017/S0025100313000327

Moore, K. L., Agur, A. M. R., & Dalley, A. F. (2018). *Clinically oriented anatomy* (Eighth edition.). Wolters Kluwer.

Mücke, D., Thies, T., Mertens, J., & Hermes, A. (2020). Age-related effects of prosodic prominence in vowel articulation. *Proceedings of the 12th International Seminar on Speech Production*, 126–129. https://hal.science/hal-03510924

Niimi, S., Horiguchi, S., & Kobayashi, N. (1988). The physiological role of the sternothyroid muscle in phonation. an electromyographic observation. *Ann Bull RILP*, *22*, 163–170.

doi:10.6342/NTU202503180

Palmer, J. B., & Hiiemae, K. M. (1997). Integration of oral and pharyngeal bolus propulsion: A new model for the physiology of swallowing [Publisher: The Japanese Society of Dysphagia Rehabilitation]. *The Japanese Journal of Dysphagia Rehabilitation*, *1*(1), 15–30.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Poh, Z. Y. (2024). Articulatory characterisation of singapore mandarin tones using ultrasound. https://doi.org/10.32657/10356/174858

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Ramsay, J. (2025). *Fda: Functional data analysis* [R package version 6.3.0]. https://doi.org/10.32614/CRAN.package.fda

Riordan, C. J. (1977). Control of vocal-tract length in speech. *The Journal of the Acoustical Society of America*, *62*(4), 998–1002. https://doi.org/10.1121/1.381595

Robbins, J. A. (1996). Normal swallowing and aging. *Seminars in neurology*, *16*(04), 309–317.

Sanders, R. (2008). Tonetic sound change in taiwan mandarin: The case of tone 2 and tone 3 citation contours. *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, *1*, 87–107.

Sawashima, M., Kakita, Y., & Hiki, S. (1973). Activity of the extrinsic laryngeal muscles in relation to japanese word accent. *Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo)*, *7*, 19–25.

doi:10.6342/NTU202503180

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. Retrieved May 17, 2025, from http://www.jstor.org/stable/2958889

Shaw, J. A., Chen, W.-r., Proctor, M. I., & Derrick, D. (2016). Influences of tone on vowel articulation in mandarin chinese. *Journal of Speech, Language, and Hearing Research*, *59*(6), S1566–S1574. https://doi.org/10.1044/2015_JSLHR-S-15-0031

Stone, M., & Lundberg, A. (1996). Three-dimensional tongue surface shapes of english consonants and vowels. *The Journal of the Acoustical Society of America*, *99*(6), 3728–3737. https://doi.org/10.1121/1.414969

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2022). itsadug: Interpreting time series and autocorrelated data using gamms [R package version 2.4.1].

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waltl, S., & Hoole, P. (2008). An emg study of the german vowel system. *8th International Seminar on Speech Production. Strasbourg, France: INRIA*, 445–448.

Whalen, D., & Levitt, A. G. (1995). The universality of intrinsic f0 of vowels. *Journal of Phonetics*, *23*(3), 349–366. https://doi.org/10.1016/S0095-4470(95)80165-0

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between l1 and l2 speakers of english. *Journal of Phonetics*, *70*, 86–116.

doi:10.6342/NTU202503180

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wu, E.-C., Fon, J., et al. (2010). The effect of min proficiency on the realization of mandarin tones in mandarin-min bilinguals. *5th International Conference on Speech Prosody Chicago*.

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

Appendix A GAMM Summary

A.1 AIC and BIC

Table A.1: Model comparison for GAM Models

target	dist	time	AIC	BIC
A1	nHT	abs_time	-68617.07	-65731.10
A2	nHT	abs_time	-69209.71	-66333.50
A3	nHT	abs_time	-64402.64	-61708.82
A4	nHT	abs_time	-59073.85	-56870.79
I1	nHT	abs_time	-76896.50	-73945.41
I2	nHT	abs_time	-68036.40	-65423.97
I3	nHT	abs_time	-72806.09	-69865.96
I4	nHT	abs_time	-67838.30	-65071.91
U1	nHT	abs_time	-88091.53	-84806.49
U2	nHT	abs_time	-77934.00	-74921.01

Continued on next page

doi:10.6342/NTU202503180

Table A.1: Model comparison for GAM Models

Target	Distance Reference	Temporal Reference	AIC	BIC
U3	nHT	abs_time	-77660.44	-74628.50
U4	nHT	abs_time	-78681.48	-75472.47
A1	nHT	nor_time	-68596.13	-65799.87
A2	nHT	nor_time	-69194.68	-66433.52
A3	nHT	nor_time	-64465.92	-61791.53
A4	nHT	nor_time	-59041.40	-56807.68
I1	nHT	nor_time	-76813.09	-73975.29
I2	nHT	nor_time	-68030.21	-65459.85
I3	nHT	nor_time	-72842.72	-70029.58
I4	nHT	nor_time	-67651.83	-65088.75
U1	nHT	nor_time	-88029.86	-84894.52
U2	nHT	nor_time	-77967.03	-74988.98
U3	nHT	nor_time	-77595.17	-74750.49
U4	nHT	nor_time	-78565.85	-75579.93
A1	aHT	abs_time	109791.37	112755.62
A2	aHT	abs_time	108362.72	111463.41
A3	aHT	abs_time	110109.20	113017.24
A4	aHT	abs_time	101775.33	104418.67

Continued on next page

doi:10.6342/NTU202503180

Table A.1: Model comparison for GAM Models

Target	Distance Reference	Temporal Reference	AIC	BIC
I1	aHT	abs_time	116309.02	119450.27
I2	aHT	abs_time	106641.53	109366.44
I3	aHT	abs_time	117103.93	120257.33
I4	aHT	abs_time	109433.38	112348.29
U1	aHT	abs_time	123440.86	126810.98
U2	aHT	abs_time	121200.77	124374.52
U3	aHT	abs_time	120934.06	124130.73
U4	aHT	abs_time	120049.69	123366.13
A1	aHT	nor_time	109865.79	112762.63
A2	aHT	nor_time	108425.79	111429.02
A3	aHT	nor_time	110084.05	113007.42
A4	aHT	nor_time	101968.35	104501.04
I1	aHT	nor_time	116514.14	119527.48
I2	aHT	nor_time	106667.06	109368.87
I3	aHT	nor_time	117271.29	120243.15
I4	aHT	nor_time	109836.51	112501.76
U1	aHT	nor_time	123536.07	126753.02
U2	aHT	nor_time	121430.25	124501.21

Continued on next page

doi:10.6342/NTU202503180

Table A.1: Model comparison for GAM Models

Target	Distance Reference	Temporal Reference	AIC	BIC
U3	aHT	nor_time	121100.13	124072.14
U4	aHT	nor_time	120356.85	123376.32

A.2 *nHT-abs_time* Models

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83971	0.02126	39.50	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	16.83	23.24	3.513	< .001
s(Time):ageYoung Adult	19.18	26.29	10.718	< .001
s(Time,participant):ageElderly	85.79	108.00	143.207	< .001
s(Time,participant):ageYoung Adult	150.06	198.00	56.730	< .001
s(id)	91.13	129.00	7.171	< .001
R-sq.(adj) = 0.871				
Deviance explained = 87.4%				

Table A.2: Summary of GAM Model: *nHT-abs_time* Tone 1 [a] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83500	0.01537	54.34	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	11.80	16.35	3.696	< .001
s(Time):ageYoung Adult	19.72	27.11	10.972	< .001
s(Time,participant):ageElderly	88.41	117.00	37.051	< .001
s(Time,participant):ageYoung Adult	142.13	207.00	16.389	< .001
s(id)	98.94	130.00	8.553	< .001
R-sq.(adj) = 0.825 Deviance explained = 82.8%				

Table A.3: Summary of GAM Model: *nHT-abs_time* Tone 2 [a] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82541	0.01506	54.80	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	8.05	10.94	2.676	< .01
s(Time):ageYoung Adult	19.11	26.28	11.368	< .001
s(Time,participant):ageElderly	76.23	99.00	94.022	< .001
s(Time,participant):ageYoung Adult	130.51	198.00	6.705	< .01
s(id)	105.24	127.00	11.866	< .001
R-sq.(adj) = 0.839 Deviance explained = 84.1%				

Table A.4: Summary of GAM Model: *nHT-abs_time* Tone 3 [a] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83549	0.01419	58.86	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	12.80	17.94	2.662	< .001
s(Time):ageYoung Adult	19.38	26.82	14.346	< .001
s(Time,participant):ageElderly	64.63	90.00	15.103	< .001
s(Time,participant):ageYoung Adult	82.27	180.00	1.525	< .001
s(id)	100.07	117.00	8.415	< .001
R-sq.(adj) = 0.768 Deviance explained = 77.2%				

Table A.5: Summary of GAM Model: *nHT-abs_time* Tone 4 [a] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83499	0.01684	49.60	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	12.02	16.72	3.439	< .001
s(Time):ageYoung Adult	16.80	23.22	9.249	< .001
s(Time,participant):ageElderly	95.87	126.00	40.592	< .001
s(Time,participant):ageYoung Adult	138.56	198.00	33.286	< .001
s(id)	104.06	140.00	8.056	< .001
R-sq.(adj) = 0.830 Deviance explained = 83.3%				

Table A.6: Summary of GAM Model: *nHT-abs_time* Tone 1 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83628	0.01823	45.88	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	8.89	12.23	2.604	< .01
s(Time):ageYoung Adult	18.43	25.49	8.194	< .001
s(Time,participant):ageElderly	81.02	108.00	17.961	< .001
s(Time,participant):ageYoung Adult	127.76	180.00	23.954	< .001
s(id)	92.59	127.00	6.092	< .001
R-sq.(adj) = 0.814 Deviance explained = 81.7%				

Table A.7: Summary of GAM Model: *nHT-abs_time* Tone 2 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.81572	0.01582	51.57	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	11.37	15.79	3.348	< .001
s(Time):ageYoung Adult	17.81	24.62	9.001	< .001
s(Time,participant):ageElderly	94.78	126.00	42.908	< .001
s(Time,participant):ageYoung Adult	140.89	198.00	16.901	< .001
s(id)	101.69	138.00	6.840	< .001
R-sq.(adj) = 0.826 Deviance explained = 82.9%				

Table A.8: Summary of GAM Model: *nHT-abs_time* Tone 3 [i] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83287	0.01280	65.07	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	14.24	19.80	2.033	< .01
s(Time):ageYoung Adult	20.34	27.86	10.332	< .001
s(Time,participant):ageElderly	94.17	117.00	37.777	< .001
s(Time,participant):ageYoung Adult	124.15	180.00	10.252	< .001
s(id)	94.21	129.00	5.709	< .001
R-sq.(adj) = 0.814 Deviance explained = 81.7%				

Table A.9: Summary of GAM Model: *nHT-abs_time* Tone 4 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.87255	0.01377	63.39	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	15.45	21.36	3.398	< .001
s(Time):ageYoung Adult	15.15	20.91	5.752	< .001
s(Time,participant):ageElderly	99.99	126.00	92.264	< .001
s(Time,participant):ageYoung Adult	155.26	207.00	32.044	< .001
s(id)	118.35	155.00	9.460	< .001
R-sq.(adj) = 0.831 Deviance explained = 83.4%				

Table A.10: Summary of GAM Model: *nHT-abs_time* Tone 1 [u] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.84228	0.01560	53.99	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	18.16	25.09	4.911	< .001
s(Time):ageYoung Adult	19.33	26.63	6.494	< .001
s(Time,participant):ageElderly	83.37	117.00	20.703	< .001
s(Time,participant):ageYoung Adult	140.34	198.00	27.211	< .001
s(id)	111.20	147.00	7.514	< .001
R-sq.(adj) = 0.799 Deviance explained = 80.2%				

Table A.11: Summary of GAM Model: *nHT-abs_time* Tone 2 [u] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.84156	0.01508	55.79	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	14.30	19.80	4.472	< .001
s(Time):ageYoung Adult	16.02	22.12	8.558	< .001
s(Time,participant):ageElderly	85.62	117.00	28.692	< .01
s(Time,participant):ageYoung Adult	141.56	198.00	30.951	< .001
s(id)	118.10	146.00	12.406	< .001
R-sq.(adj) = 0.823 Deviance explained = 82.6%				

Table A.12: Summary of GAM Model: *nHT-abs_time* Tone 3 [u] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.88204	0.01241	71.06	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	18.85	26.02	3.307	< .001
s(Time):ageYoung Adult	15.89	21.99	5.265	< .001
s(Time,participant):ageElderly	90.28	117.00	25.834	< .01
s(Time,participant):ageYoung Adult	147.79	207.00	9.274	< .001
s(id)	124.63	146.00	12.741	< .001
R-sq.(adj) = 0.820 Deviance explained = 82.3%				

Table A.13: Summary of GAM Model: *nHT-abs_time* Tone 4 [u] model

doi:10.6342/NTU202503180

A.3 *nHT-nor_time* Models

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.84658	0.01671	50.67	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	24.25	32.17	2.888	< .001
s(Time):ageYoung Adult	25.12	33.32	9.919	< .001
s(Time,participant):ageElderly	79.93	108.00	137.187	< .001
s(Time,participant):ageYoung Adult	127.46	198.00	29.734	< .001
s(id)	94.68	129.00	7.849	< .001
R-sq.(adj) = 0.871 Deviance explained = 87.3%				

Table A.14: Summary of GAM Model: *nHT-nor_time* Tone 1 [a] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83416	0.01479	56.41	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	12.75	17.57	2.768	< .001
s(Time):ageYoung Adult	27.87	36.98	10.917	< .001
s(Time,participant):ageElderly	83.79	117.00	44.888	< .001
s(Time,participant):ageYoung Adult	122.18	207.00	12.604	< .001
s(id)	99.88	130.00	8.659	< .001
R-sq.(adj) = 0.821 Deviance explained = 82.5%				

Table A.15: Summary of GAM Model: *nHT-nor_time* Tone 2 [a] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.81875	0.01718	47.65	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	7.93	10.68	1.132	≥ .1
s(Time):ageYoung Adult	27.95	36.72	6.375	< .001
s(Time,participant):ageElderly	71.92	99.00	122.238	< .001
s(Time,participant):ageYoung Adult	123.91	198.00	14.328	≥ .1
s(id)	104.69	127.00	14.985	< .001
R-sq.(adj) = 0.850 Deviance explained = 85.3%				

Table A.16: Summary of GAM Model: *nHT-nor_time* Tone 3 [a] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83063	0.01566	53.05	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	21.80	29.67	2.778	< .001
s(Time):ageYoung Adult	23.94	32.14	8.006	< .001
s(Time,participant):ageElderly	58.42	90.00	22.850	< .001
s(Time,participant):ageYoung Adult	78.14	180.00	2.005	< .001
s(id)	100.73	117.00	9.068	< .001
R-sq.(adj) = 0.769 Deviance explained = 77.3%				

Table A.17: Summary of GAM Model: *nHT-nor_time* Tone 4 [a] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83826	0.01558	53.79	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	15.48	21.24	2.771	< .001
s(Time):ageYoung Adult	20.73	28.15	7.509	< .001
s(Time,participant):ageElderly	89.23	126.00	29.453	< .001
s(Time,participant):ageYoung Adult	122.60	198.00	22.434	< .001
s(id)	104.83	140.00	7.631	< .001
R-sq.(adj) = 0.824 Deviance explained = 82.7%				

Table A.18: Summary of GAM Model: *nHT-nor_time* Tone 1 [i] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83348	0.01733	48.08	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	9.40	12.99	2.037	< .05
s(Time):ageYoung Adult	22.58	30.61	6.088	< .001
s(Time,participant):ageElderly	77.10	108.00	12.354	< .001
s(Time,participant):ageYoung Adult	119.97	180.00	22.179	< .001
s(id)	94.16	127.00	6.381	< .001
R-sq.(adj) = 0.812 Deviance explained = 81.5%				

Table A.19: Summary of GAM Model: *nHT-nor_time* Tone 2 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.81858	0.01506	54.34	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	13.77	18.98	2.075	< .01
s(Time):ageYoung Adult	26.86	35.26	6.150	< .001
s(Time,participant):ageElderly	83.46	126.00	30.625	< .001
s(Time,participant):ageYoung Adult	123.67	198.00	16.559	< .001
s(id)	102.67	138.00	6.959	< .001
R-sq.(adj) = 0.831 Deviance explained = 83.4%				

Table A.20: Summary of GAM Model: *nHT-nor_time* Tone 3 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82981	0.01460	56.85	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	9.93	13.34	1.394	≥ .1
s(Time):ageYoung Adult	26.15	34.63	8.296	< .001
s(Time,participant):ageElderly	88.48	117.00	35.732	< .001
s(Time,participant):ageYoung Adult	103.61	180.00	13.286	< .001
s(id)	92.25	129.00	5.287	< .001
R-sq.(adj) = 0.794 Deviance explained = 79.8%				

Table A.21: Summary of GAM Model: *nHT-nor_time* Tone 4 [i] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.87508	0.01380	63.42	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	19.12	25.75	2.079	< .01
s(Time):ageYoung Adult	21.30	28.54	4.509	< .001
s(Time,participant):ageElderly	91.77	126.00	84.019	< .001
s(Time,participant):ageYoung Adult	134.46	207.00	30.745	< .001
s(id)	118.55	155.00	9.310	< .001
R-sq.(adj) = 0.826 Deviance explained = 82.9%				

Table A.22: Summary of GAM Model: *nHT-nor_time* Tone 1 [u] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.84553	0.01609	52.54	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	21.48	28.97	5.147	< .001
s(Time):ageYoung Adult	20.78	28.20	5.042	< .001
s(Time,participant):ageElderly	79.99	117.00	28.261	< .001
s(Time,participant):ageYoung Adult	134.64	198.00	32.362	< .001
s(id)	111.25	147.00	8.063	< .001
R-sq.(adj) = 0.806 Deviance explained = 80.9%				

Table A.23: Summary of GAM Model: *nHT-nor_time* Tone 2 [u] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.84436	0.01513	55.79	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	15.46	21.01	2.603	< .001
s(Time):ageYoung Adult	22.57	30.25	4.780	< .001
s(Time,participant):ageElderly	69.90	117.00	25.931	< .001
s(Time,participant):ageYoung Adult	126.22	198.00	34.200	< .001
s(id)	117.95	146.00	11.826	< .001
R-sq.(adj) = 0.817 Deviance explained = 82.0%				

Table A.24: Summary of GAM Model: *nHT-nor_time* Tone 3 [u] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.87881	0.01453	60.48	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	14.17	19.43	1.935	< .01
s(Time):ageYoung Adult	19.96	27.05	4.174	< .001
s(Time,participant):ageElderly	87.90	117.00	40.690	< .001
s(Time,participant):ageYoung Adult	129.09	207.00	15.025	< .001
s(id)	118.57	146.00	11.142	< .001
R-sq.(adj) = 0.808 Deviance explained = 81.1%				

Table A.25: Summary of GAM Model: *nHT-nor_time* Tone 4 [u] model

doi:10.6342/NTU202503180

doi:10.6342/NTU202503180

Appendix B Summary of Models for Raised and Lowered Tone 3

B.1 AIC and BIC

Table B.1: Model comparison for GAM Models

Target	Distance Reference	Temporal Reference	AIC	BIC
A3 raised	nHT	abs_time	-25939.71	-24685.69
I3 raised	nHT	abs_time	-27155.08	-25838.30
U3 raised	nHT	abs_time	-29855.48	-28557.41
A3 raised	nHT	nor_time	-25950.76	-24716.53
I3 raised	nHT	nor_time	-27177.33	-25869.73
U3 raised	nHT	nor_time	-29831.75	-28557.14
A3 lowered	nHT	abs_time	-38548.24	-36803.15
I3 lowered	nHT	abs_time	-45780.35	-43837.66

Continued on next page

doi:10.6342/NTU202503180

Table B.1: Model comparison for GAM Models

Target	Distance Reference	Temporal Reference	AIC	BIC
U3 lowered	nHT	abs_time	-47921.15	-46013.03
A3 lowered	nHT	nor_time	-38563.64	-36840.11
I3 lowered	nHT	nor_time	-45828.75	-43968.26
U3 lowered	nHT	nor_time	-47897.96	-46108.27

B.2 Model Summary

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.81267	0.02237	36.33	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	5.353	6.98	2.172	< .05
s(Time):ageYoung Adult	12.603	16.96	8.678	< .001
s(Time,participant):ageElderly	51.776	63.00	180.868	< .001
s(Time,participant):ageYoung Adult	63.854	117.00	3.444	< .001
s(id)	43.526	51.00	18.326	< .001
R-sq.(adj) = 0.879				
Deviance explained = 88.2%				

Table B.2: Summary of GAM Model: *nHT-abs_time* Raised Tone 3 [a] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.79708	0.02391	33.34	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	9.531	12.97	3.588	< .001
s(Time):ageYoung Adult	12.657	17.21	6.137	< .001
s(Time,participant):ageElderly	60.092	81.00	43.302	< .001
s(Time,participant):ageYoung Adult	66.292	117.00	4.178	< .01
s(id)	38.171	49.00	9.406	< .001
R-sq.(adj) = 0.879 Deviance explained = 88.2%				

Table B.3: Summary of GAM Model: *nHT-abs_time* Raised Tone 3 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83080	0.02288	36.32	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	10.430	14.06	3.809	< .001
s(Time):ageYoung Adult	9.783	13.13	6.269	< .001
s(Time,participant):ageElderly	57.058	81.00	38.332	< .05
s(Time,participant):ageYoung Adult	52.595	99.00	3.248	< .01
s(id)	49.872	58.00	17.214	< .001
R-sq.(adj) = 0.835 Deviance explained = 83.8%				

Table B.4: Summary of GAM Model: *nHT-abs_time* Raised Tone 3 [u] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.78441	0.02664	29.44	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	14.092	17.48	2.482	< .001
s(Time):ageYoung Adult	20.165	25.87	4.435	< .001
s(Time,participant):ageElderly	38.544	63.00	155.663	< .001
s(Time,participant):ageYoung Adult	58.709	117.00	5.549	< .01
s(id)	41.968	51.00	15.555	< .001
R-sq.(adj) = 0.880 Deviance explained = 88.3%				

Table B.5: Summary of GAM Model: *nHT-nor_time* Raised Tone 3 [a] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.79853	0.02246	35.56	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	12.422	16.47	2.283	< .01
s(Time):ageYoung Adult	16.667	21.96	3.802	< .001
s(Time,participant):ageElderly	53.885	81.00	34.448	< .01
s(Time,participant):ageYoung Adult	64.404	117.00	8.959	< .01
s(id)	37.993	49.00	10.955	< .001
R-sq.(adj) = 0.891				
Deviance explained = 89.3%				

Table B.6: Summary of GAM Model: *nHT-nor_time* Raised Tone 3 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83267	0.02337	35.63	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	10.603	14.18	2.354	< .01
s(Time):ageYoung Adult	10.491	13.92	3.811	< .001
s(Time,participant):ageElderly	52.311	81.00	31.150	< .01
s(Time,participant):ageYoung Adult	53.602	99.00	8.031	< .05
s(id)	49.240	58.00	16.528	< .001
R-sq.(adj) = 0.833				
Deviance explained = 83.6%				

Table B.7: Summary of GAM Model: *nHT-nor_time* Raised Tone 3 [u] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82571	0.01683	49.07	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	6.107	7.89	2.989	< .01
s(Time):ageYoung Adult	16.516	21.98	12.296	< .001
s(Time,participant):ageElderly	40.254	63.00	97.875	< .01
s(Time,participant):ageYoung Adult	107.177	171.00	5.753	< .01
s(id)	64.082	75.00	16.115	< .001
R-sq.(adj) = 0.843				
Deviance explained = 84.7%				

Table B.8: Summary of GAM Model: *nHT-abs_time* Lowered Tone 3 [a] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82459	0.01884	43.76	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	13.118	17.70	1.892	< .05
s(Time):ageYoung Adult	14.586	19.58	9.336	< .001
s(Time,participant):ageElderly	40.650	63.00	45.771	< .001
s(Time,participant):ageYoung Adult	124.897	180.00	25.845	< .001
s(id)	61.581	87.00	7.142	< .001
R-sq.(adj) = 0.817 Deviance explained = 82.1%				

Table B.9: Summary of GAM Model: *nHT-abs_time* Lowered Tone 3 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.84387	0.01687	50.02	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	8.739	11.74	2.785	< .01
s(Time):ageYoung Adult	16.362	21.88	7.903	< .001
s(Time,participant):ageElderly	33.571	54.00	7.046	< .05
s(Time,participant):ageYoung Adult	124.622	180.00	33.525	< .001
s(id)	67.767	87.00	10.118	< .001
R-sq.(adj) = 0.823 Deviance explained = 82.7%				

Table B.10: Summary of GAM Model: *nHT-abs_time* Lowered Tone 3 [u] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82948	0.01851	44.82	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	4.489	5.57	1.577	≥ .1
s(Time):ageYoung Adult	21.949	27.86	8.077	< .001
s(Time,participant):ageElderly	44.038	63.00	100.213	< .01
s(Time,participant):ageYoung Adult	99.327	171.00	13.406	< .05
s(id)	61.384	75.00	16.359	< .001
R-sq.(adj) = 0.847 Deviance explained = 85.1%				

Table B.11: Summary of GAM Model: *nHT-nor_time* Lowered Tone 3 [a] model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82899	0.01829	45.33	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	11.746	15.63	1.517	≥ .1
s(Time):ageYoung Adult	22.968	28.76	6.896	< .001
s(Time,participant):ageElderly	38.729	63.00	37.607	< .001
s(Time,participant):ageYoung Adult	108.886	180.00	24.662	< .001
s(id)	61.916	87.00	7.048	< .001
R-sq.(adj) = 0.821 Deviance explained = 82.5%				

Table B.12: Summary of GAM Model: *nHT-nor_time* Lowered Tone 3 [i] model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.85113	0.01763	48.27	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	11.665	15.49	1.605	< .1
s(Time):ageYoung Adult	21.037	27.00	4.563	< .001
s(Time,participant):ageElderly	26.794	54.00	4.598	< .1
s(Time,participant):ageYoung Adult	108.772	180.00	32.401	< .001
s(id)	66.694	87.00	9.587	< .001
R-sq.(adj) = 0.818 Deviance explained = 82.1%				

Table B.13: Summary of GAM Model: *nHT-nor_time* Lowered Tone 3 [u] model

doi:10.6342/NTU202503180

Appendix C Exploratory Tone Models

Additional exploratory models are presented in this chapter. These models were designed to explore potential age and vowel interaction given the same tone. The maximally tested model is described as Formula C.1. This model was similar to Formula 3.4, with important differences. To account for the potential interaction between age group (young adult and elderly) and vowel ([a], [i], [u]), the interaction term was added as a parametric effect. An additional factor smooth was specified for each vowel within each age group, allowing for non-linear, group-specific vowel-level variability over time.

Another important difference was the data included. In Formula 3.4, only data with the same vowel-tone combination were included. However, in for Formula C.1, all data with the same tone were included.

$$
\begin{aligned}
y \sim\ & \text{age} * \text{vowel} + s(\text{Time}, \text{by} = \text{age}) + s(\text{Time}, \text{participant}, \text{by} = \text{age}, \text{bs} = \text{'fs'}, m = 1) + \\
& s(\text{Time}, \text{vowel}, \text{by} = \text{age}, \text{bs} = \text{'fs'}, m = 1) + s(\text{id}, \text{bs} = \text{'re'}), \\
& \text{method} = \text{'fREML'},\ \text{Rho},\ \text{AR.start}
\end{aligned}
\tag{C.1}
$$

As with Formula 3.4, in Formula C.1, time was modeled with either of the two temporal references (*abs_time* or *nor_time*) and y was modeled with either of the two distance refer-

doi:10.6342/NTU202503180

ences (*aHT* or *nHT*). Insignificant terms in the models were removed to improve parsimony. The models were also examined with AIC (Akaike, 1974) and BIC (Schwarz, 1978) to compare model performance.

aHT models consistently showed larger values for both AIC and BIC compared to the *nHT* models and were not presented. For the *nHT* models, the interaction between age and vowel was not significant, nor was the parametric term age. Hence, the two were removed, resulting in Formula C.2. The AIC and BIC of the Formula C.2 models are presented in Table C, followed by the GAMM plots of the *nHT* models (Figure C.1).

The plots in Figure C.1 are generally arranged in the same way as the figures in Section 4.2 with the exception that from left to right, the plots represents results for Tones 1 to 4. The summaries of the *nHT* models come after Figure C.1.

$$
\begin{aligned}
y \sim \text{vowel} + s(\text{Time}, \text{by} = \text{age}) + s(\text{Time}, \text{participant}, \text{by} = \text{age}, \text{bs} = \text{'fs'}, m = 1) + \\
s(\text{Time}, \text{vowel}, \text{by} = \text{age}, \text{bs} = \text{'fs'}, m = 1) + s(\text{id}, \text{bs} = \text{'re'}), \\
\text{method} = \text{'fREML'}, \ \text{Rho}, \ \text{AR.start}
\end{aligned} \tag{C.2}
$$

Table C.1: Model comparison for GAM Models

Tone	Distance Reference	Temporal Reference	AIC	BIC
1	*nHT*	*abs_time*	-232542.99	-226327.76
2	*nHT*	*abs_time*	-214858.18	-208940.32

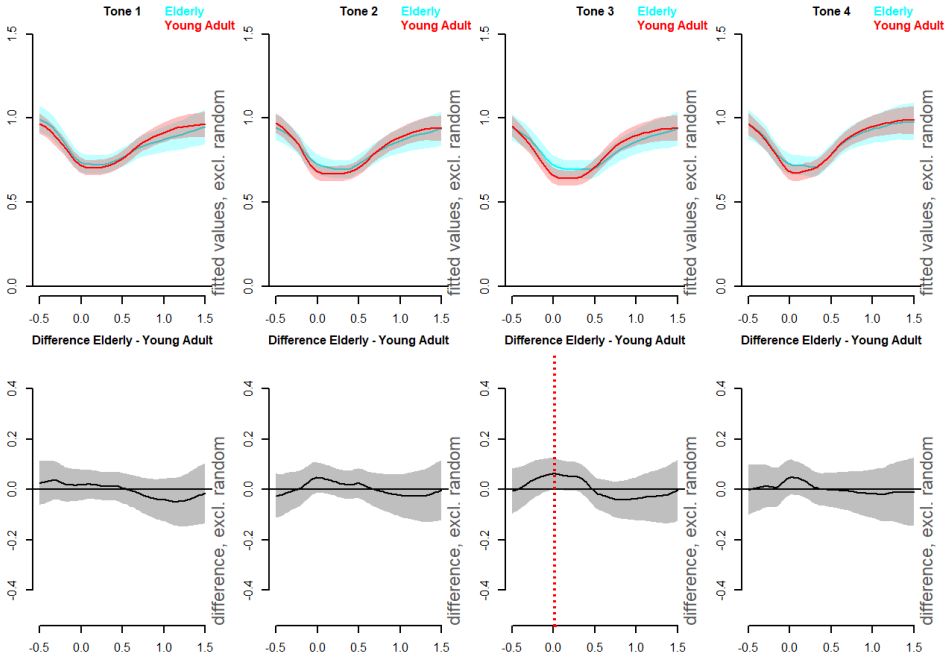
Continued on next page

doi:10.6342/NTU202503180

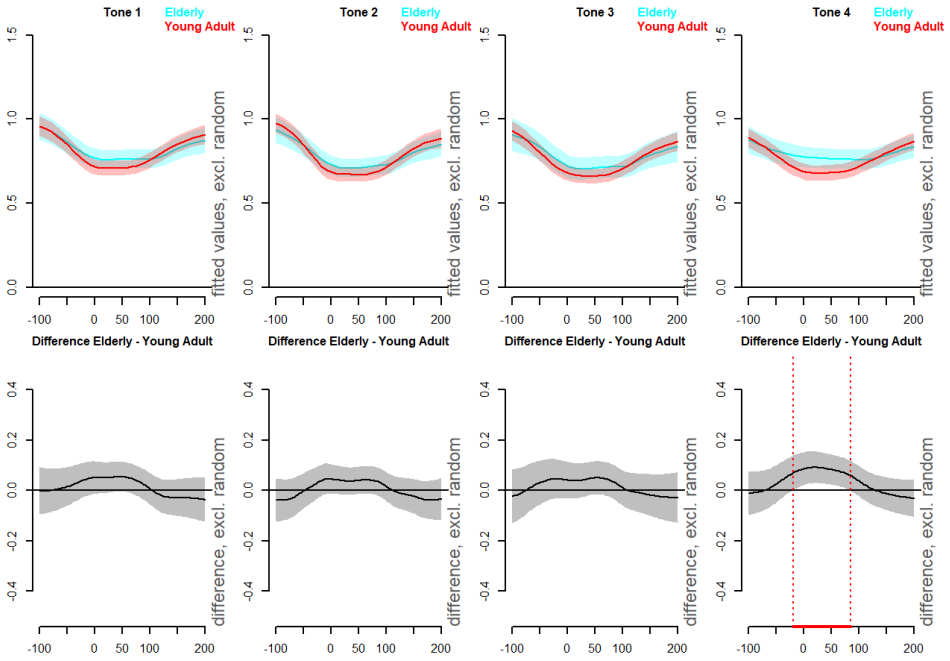
Table C.1: Model comparison for GAM Models

Target	Distance Reference	Temporal Reference	AIC	BIC
3	*nHT*	*abs_time*	-214467.93	-208362.00
4	*nHT*	*abs_time*	-204939.26	-199050.57
1	*nHT*	*nor_time*	-232420.23	-226329.36
2	*nHT*	*nor_time*	-214825.18	-208957.28
3	*nHT*	*nor_time*	-214329.22	-208321.18
4	*nHT*	*nor_time*	-204658.95	-198915.04
1	*aHT*	*abs_time*	351052.73	357336.48
2	*aHT*	*abs_time*	336781.41	342819.16
3	*aHT*	*abs_time*	348938.95	355107.63
4	*aHT*	*abs_time*	332333.48	338287.83
1	*aHT*	*nor_time*	351304.06	357496.76
2	*aHT*	*nor_time*	336970.25	342973.75
3	*aHT*	*nor_time*	349358.64	355426.66
4	*aHT*	*nor_time*	333002.28	338759.96

doi:10.6342/NTU202503180



(a) *abs_time* models



(b) *nor_time* models

Figure C.1: Generalized additive mixed model results for Tones 1 to 4

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83703	0.01583	52.88	< .001
vowelI	-0.00040	0.01159	-0.03	≥ .1
vowelU	0.03550	0.01135	3.13	< .01
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	20.77	28.77	3.112	< .001
s(Time):ageYoung Adult	22.19	30.55	5.007	< .001
s(Time,participant):ageElderly	105.98	126.00	74.713	< .001
s(Time,participant):ageYoung Adult	161.60	207.00	37.262	< .001
s(Time,vowel):ageElderly	12.50	27.00	3.868	< .001
s(Time,vowel):ageYoung Adult	14.86	27.00	8.151	< .001
s(id)	341.68	424.00	6.177	< .001
R-sq.(adj) = 0.787				
Deviance explained = 78.9%				

Table C.2: Summary of GAM Model: *nHT-abs_time* Tone 1 model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83143	0.01555	53.45	< .001
vowelI	0.00347	0.01274	0.27	≥ .1
vowelU	0.01650	0.01243	1.33	≥ .1
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	16.43	23.08	3.232	< .001
s(Time):ageYoung Adult	25.75	34.92	6.151	< .001
s(Time,participant):ageElderly	101.10	126.00	29.132	< .001
s(Time,participant):ageYoung Adult	159.98	207.00	27.850	< .001
s(Time,vowel):ageElderly	12.44	27.00	2.830	< .001
s(Time,vowel):ageYoung Adult	13.61	27.00	4.363	< .001
s(id)	320.54	404.00	5.504	< .001
R-sq.(adj) = 0.760				
Deviance explained = 76.2%				

Table C.3: Summary of GAM Model: *nHT-abs_time* Tone 2 model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82741	0.01560	53.05	< .001
vowelI	-0.01149	0.01364	-0.84	≥ .1
vowelU	0.01281	0.01365	0.94	≥ .1
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	17.51	24.48	3.255	< .001
s(Time):ageYoung Adult	22.16	30.48	6.083	< .001
s(Time,participant):ageElderly	103.80	126.00	69.844	< .001
s(Time,participant):ageYoung Adult	161.82	198.00	27.287	< .001
s(Time,vowel):ageElderly	10.46	27.00	1.552	< .001
s(Time,vowel):ageYoung Adult	13.64	27.00	5.086	< .001
s(id)	340.02	411.00	7.413	< .001
R-sq.(adj) = 0.784 Deviance explained = 78.7%				

Table C.4: Summary of GAM Model: *nHT-abs_time* Tone 3 model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.84654	0.01498	56.50	< .001
vowelI	-0.01895	0.01710	-1.11	≥ .1
vowelU	0.03384	0.01679	2.02	< .05
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	22.39	30.86	2.770	< .001
s(Time):ageYoung Adult	24.29	33.19	5.039	< .001
s(Time,participant):ageElderly	101.68	126.00	27.283	< .001
s(Time,participant):ageYoung Adult	151.28	207.00	7.727	< .001
s(Time,vowel):ageElderly	10.62	27.00	2.438	< .001
s(Time,vowel):ageYoung Adult	14.85	27.00	5.423	< .001
s(id)	323.28	392.00	6.316	< .001
R-sq.(adj) = 0.748 Deviance explained = 75.0%				

Table C.5: Summary of GAM Model: *nHT-abs_time* Tone 4 model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83797	0.01407	59.58	< .001
vowelI	0.00243	0.01160	0.21	≥ .1
vowelU	0.03871	0.01136	3.41	< .001
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	25.77	34.56	2.343	< .001
s(Time):ageYoung Adult	27.72	36.41	4.590	< .001
s(Time,participant):ageElderly	100.24	126.00	75.114	< .001
s(Time,participant):ageYoung Adult	138.98	207.00	26.234	< .001
s(Time,vowel):ageElderly	14.55	27.00	10.194	< .001
s(Time,vowel):ageYoung Adult	14.76	27.00	13.477	< .001
s(id)	343.73	424.00	6.199	< .001
R-sq.(adj) = 0.786				
Deviance explained = 78.8%				

Table C.6: Summary of GAM Model: *nHT-nor_time* Tone 1 model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.82902	0.01524	54.38	< .001
vowelI	0.00620	0.01291	0.48	≥ .1
vowelU	0.02091	0.01259	1.66	< .1
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	20.21	27.73	4.134	< .001
s(Time):ageYoung Adult	31.61	41.40	6.244	< .001
s(Time,participant):ageElderly	95.75	126.00	32.109	< .001
s(Time,participant):ageYoung Adult	148.64	207.00	27.466	< .001
s(Time,vowel):ageElderly	10.79	27.00	2.485	< .001
s(Time,vowel):ageYoung Adult	13.12	27.00	5.691	< .001
s(id)	324.17	404.00	5.878	< .001
R-sq.(adj) = 0.761				
Deviance explained = 76.4%				

Table C.7: Summary of GAM Model: *nHT-nor_time* Tone 2 model

doi:10.6342/NTU202503180

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.81135	0.01565	51.84	< .001
vowelI	0.00576	0.01384	0.42	≥ .1
vowelU	0.03013	0.01386	2.17	< .05
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	20.97	28.64	2.140	< .001
s(Time):ageYoung Adult	32.22	41.64	4.474	< .001
s(Time,participant):ageElderly	93.60	126.00	58.249	< .001
s(Time,participant):ageYoung Adult	144.51	198.00	27.611	< .001
s(Time,vowel):ageElderly	11.80	27.00	3.067	< .001
s(Time,vowel):ageYoung Adult	13.32	27.00	7.896	< .001
s(id)	342.11	411.00	7.631	< .001
R-sq.(adj) = 0.784				
Deviance explained = 78.6%				

Table C.8: Summary of GAM Model: *nHT-nor_time* Tone 3 model

Parametric Coefficients	Estimate	Std. Error	t value	p-value
(Intercept)	0.83047	0.01698	48.92	< .001
vowelI	0.00008	0.01858	0.00	≥ .1
vowelU	0.04749	0.01830	2.59	< .01
Approximate significance of smooth terms	edf	Ref.df	F	p-value
s(Time):ageElderly	22.76	30.96	2.569	< .001
s(Time):ageYoung Adult	30.40	39.93	4.711	< .001
s(Time,participant):ageElderly	97.72	126.00	32.914	< .001
s(Time,participant):ageYoung Adult	133.82	207.00	13.001	< .001
s(Time,vowel):ageElderly	14.18	27.00	3.867	< .001
s(Time,vowel):ageYoung Adult	13.17	27.00	6.150	< .001
s(id)	320.19	392.00	6.235	< .001
R-sq.(adj) = 0.738				
Deviance explained = 74.1%				

Table C.9: Summary of GAM Model: *nHT-nor_time* Tone 4 model

doi:10.6342/NTU202503180

Appendix D Retained Frames

The figures show distribution of speaker contributions across frames. The x-axis represents frame indices, and the y-axis shows the number of production contributing to a given frame. Pink bars indicate that the frame is lacking input from at least one speaker compared to the first frame.

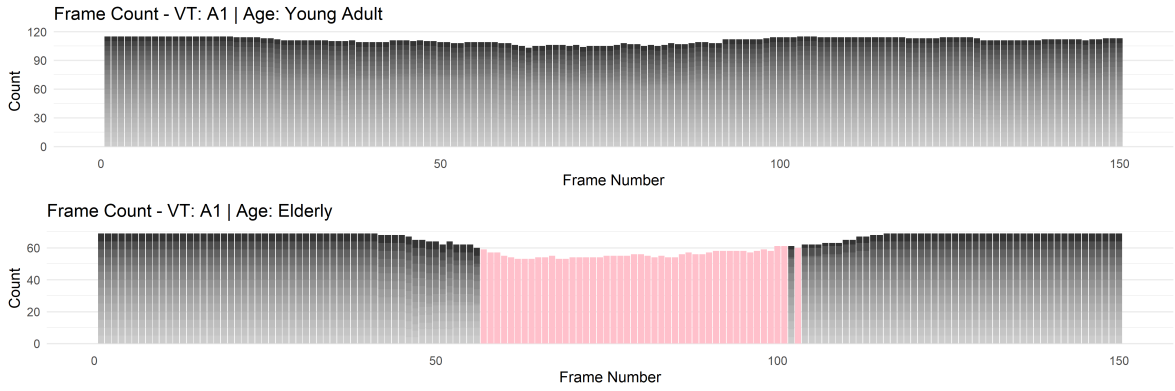


Figure D.1: Frame contribution: Tone 1 [a]

doi:10.6342/NTU202503180

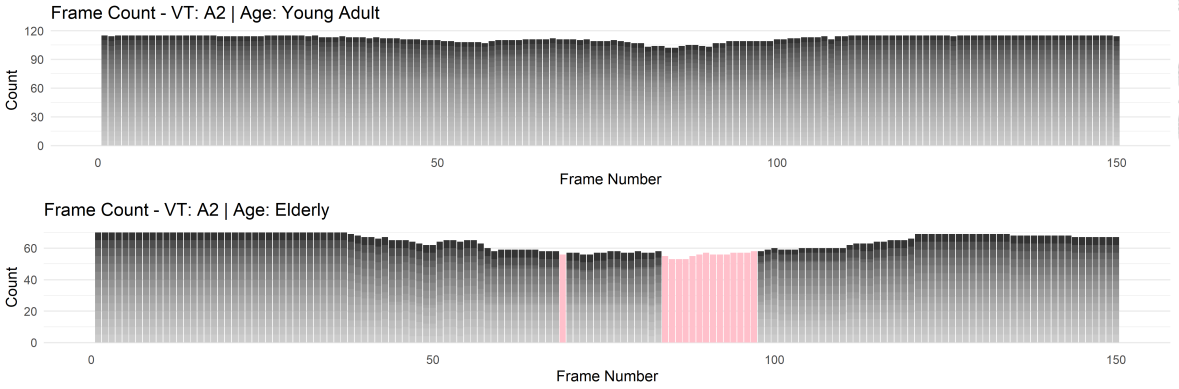


Figure D.2: Frame contribution: Tone 2 [a]

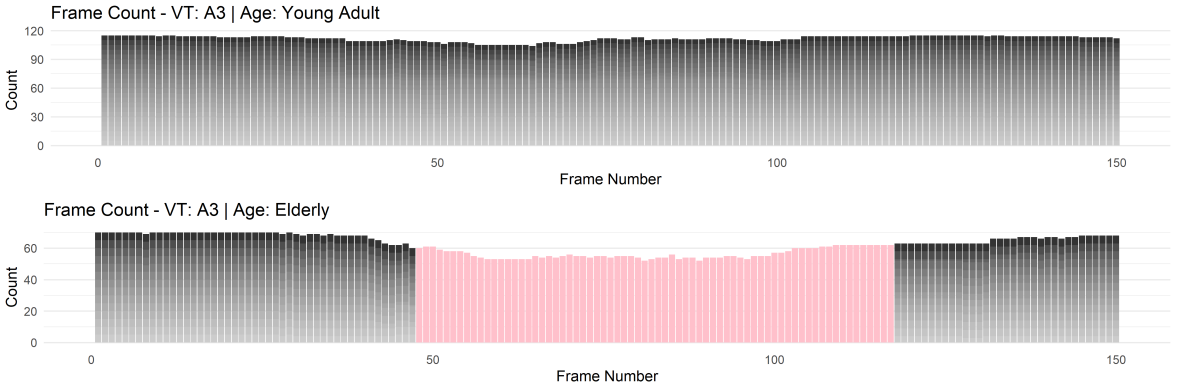


Figure D.3: Frame contribution: Tone 3 [a]

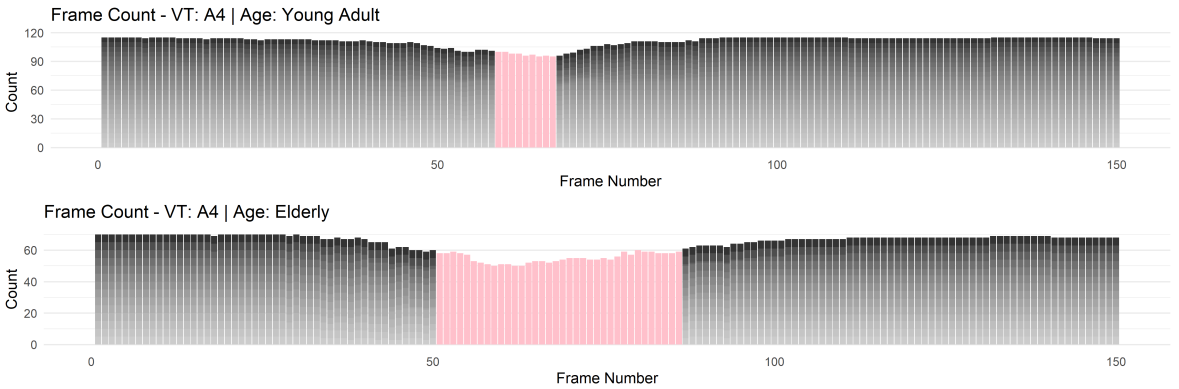


Figure D.4: Frame contribution: Tone 4 [a]

doi:10.6342/NTU202503180

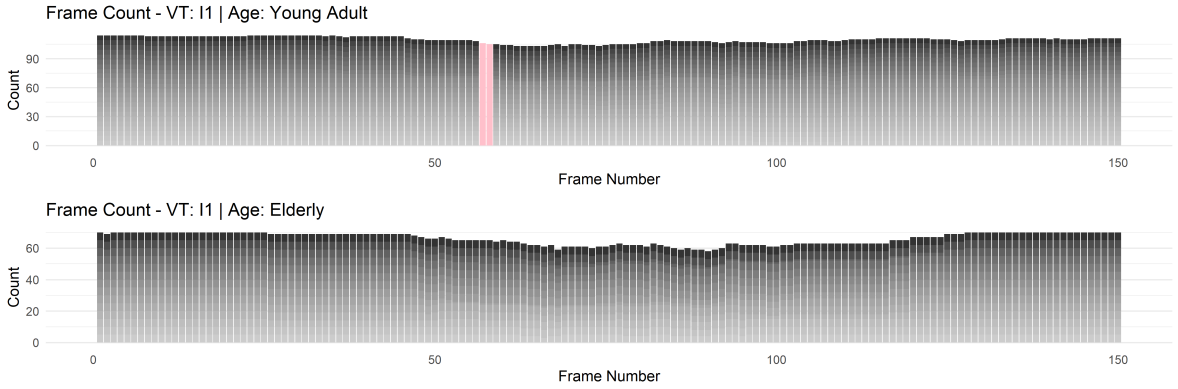


Figure D.5: Frame contribution: Tone 1 [i]

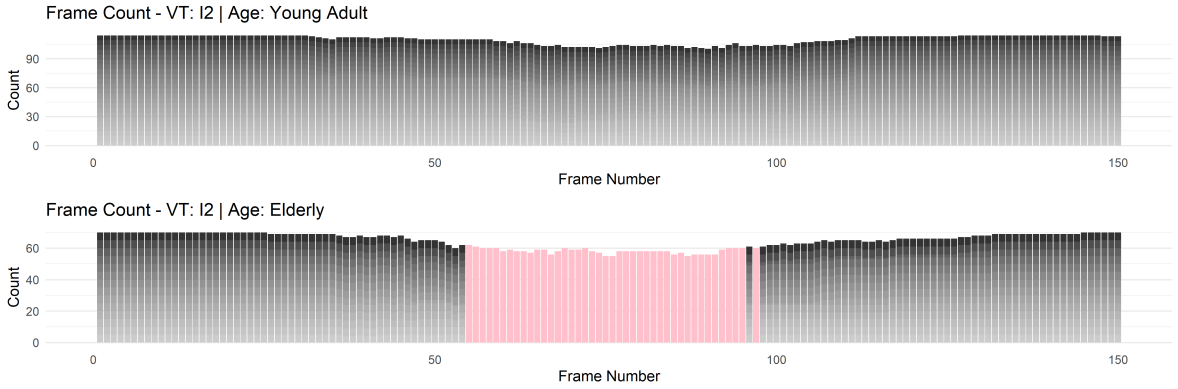


Figure D.6: Frame contribution: Tone 2 [i]

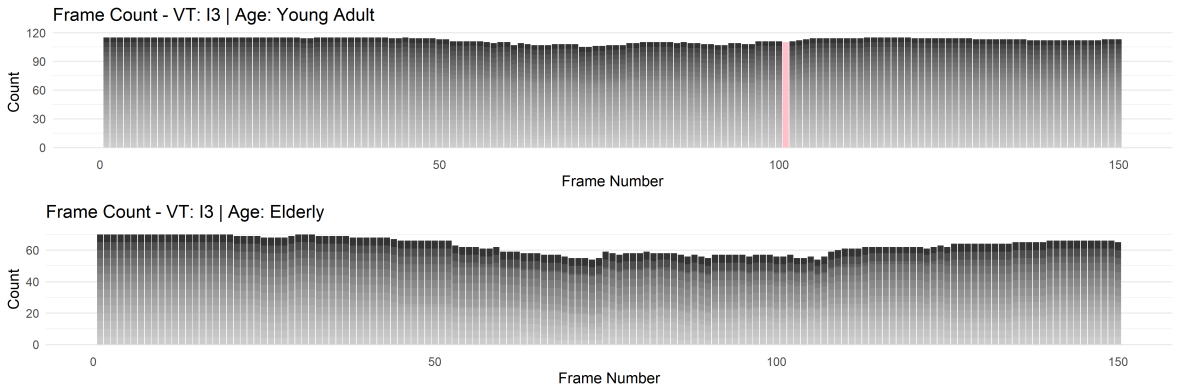


Figure D.7: Frame contribution: Tone 3 [i]

doi:10.6342/NTU202503180

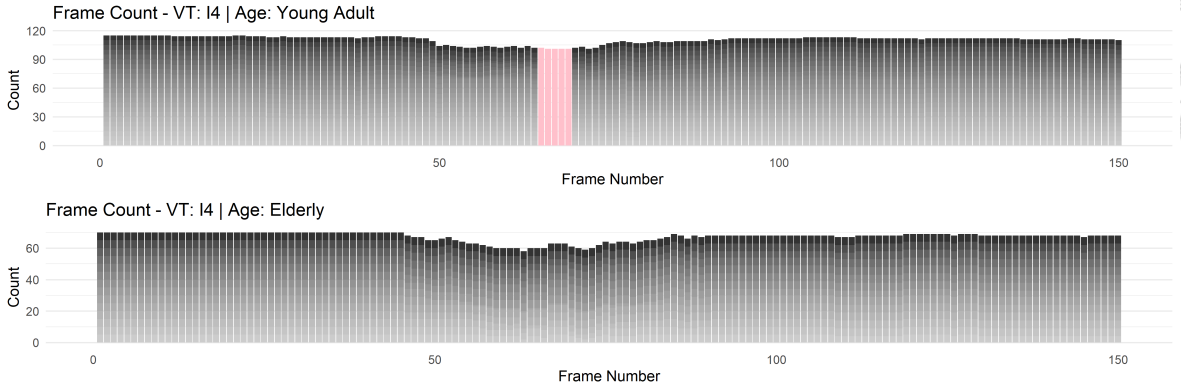


Figure D.8: Frame contribution: Tone 4 [i]

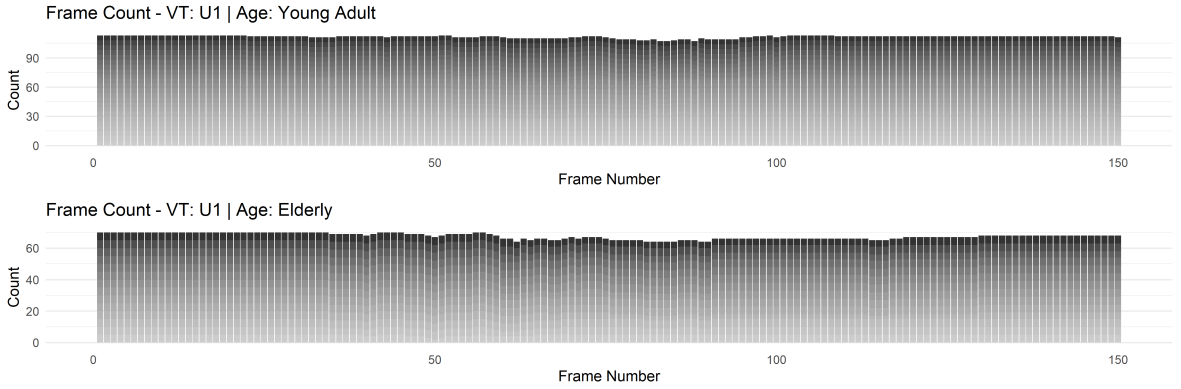


Figure D.9: Frame contribution: Tone 1 [u]

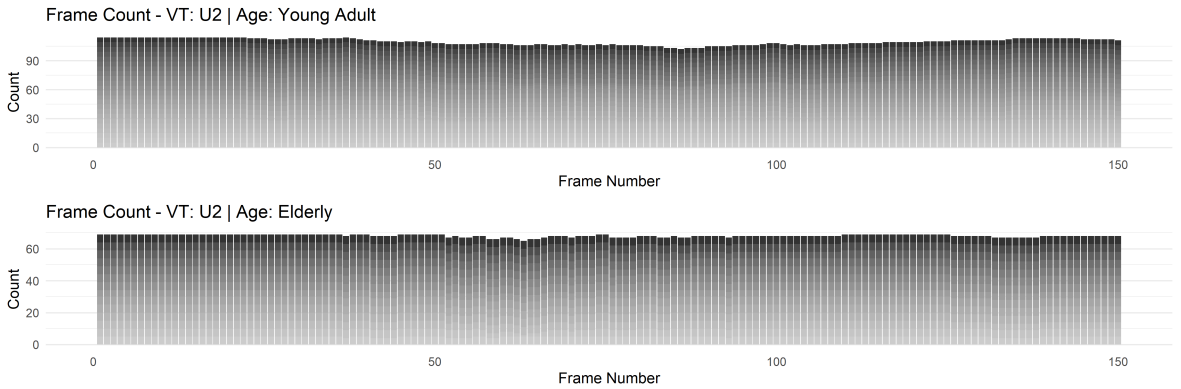


Figure D.10: Frame contribution: Tone 2 [u]

doi:10.6342/NTU202503180

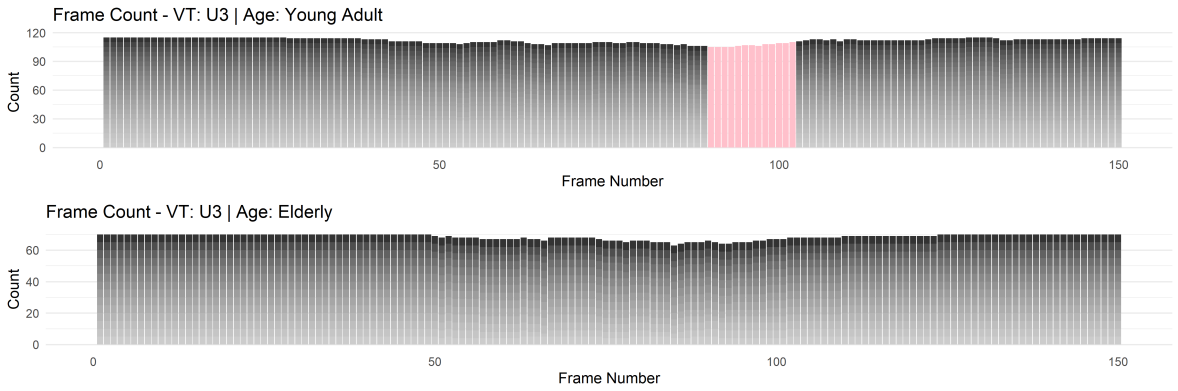


Figure D.11: Frame contribution: Tone 3 [u]

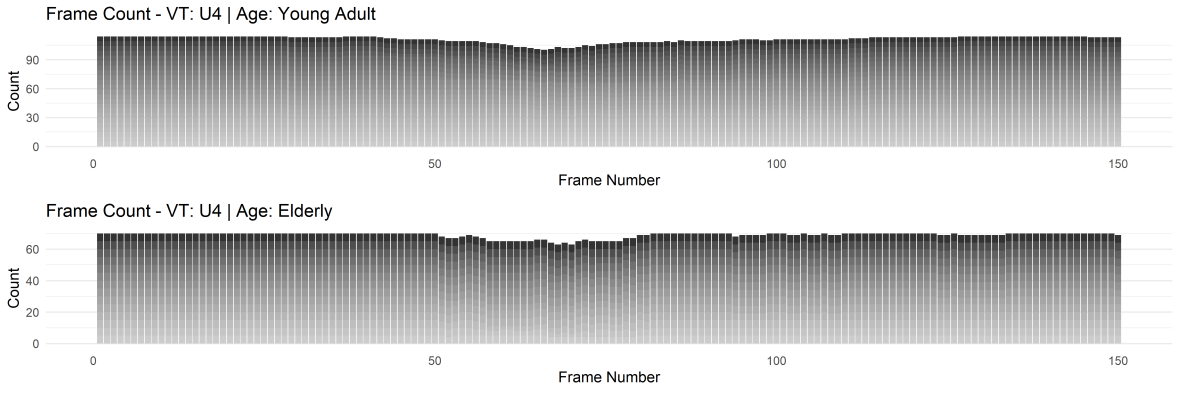


Figure D.12: Frame contribution: Tone 4 [u]

doi:10.6342/NTU202503180