The IF0 effect in the Hong Kong Cantonese tone system

by

Bingqing Yu

B.A. (Classical Philology), Zhejiang University. 2019

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Linguistics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Summer 2022

Declaration of Committee

Name:	**Bingqing Yu**
Degree:	**Master of Arts**
Title:	**The IF0 effect in the Hong Kong Cantonese tone system**
Committee:	**Chair: Maite Taboada** Professor, Linguistics
	Murray Munro Supervisor Professor, Linguistics
	Henny Yeung Committee Member Assistant Professor, Linguistics
	Stephen Winters Examiner Assistant Professor, Linguistics University of Calgary

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

This study investigates intrinsic fundamental frequency (IF0) in Hong Kong Cantonese tones. Given the tension between exaggerating IF0 to enhance vowel contrasts and restraining IF0 to distinguish confusable tone pairs, the investigation 1) examines the relationship between IF0 and tone merging and 2) tests the mixed physiological-enhancement hypothesis (MPEH) that IF0 is physiologically determined, but adjusted to enhance perception of contrasts. When F0 measurements of vowels are fit to Generalized Additive Mixed Models (GAMM), the results suggest that the confusable tone pairs are at the beginning of a tone-merging process. The size of the IF0 effect varies as follows: T2 > T1/T5 > T3/T6, which supports MPEH to some extent. However, IF0 of T4 and individual speaker performance cannot be fully explained by MPEH. Future research is needed to elucidate the IF0 mechanism and the relationship between IF0 and tone merging.

Keywords: Hong Kong Cantonese; tone-merging; intrinsic F0; mixed physiological-enhancement hypothesis

Acknowledgments

I still remember the time when I first arrived at the Linguistics Department at SFU three years ago as a new international grad student. My heart was filled with excitement, and I thought finishing this master's degree would be a piece of cake. However, as an international student who came from another field, my master thesis did not progress as smoothly as I expected, especially during the COVID period. I encountered many difficulties, from emotion management to data analyses to writing. But luckily, I met so many nice people along the way. With their help, I overcame the difficulties, successfully completed my thesis, and grew a lot.

I would like to thank my supervisor, Dr. Murray Munro, for his guidance and support both academically and mentally. His comments and suggestions always helped me further develop my thoughts. There was a long period when I felt lost in what I wanted to do, and I would especially thank him for tolerating my immature thoughts and goals that change repeatedly. As a non-native speaker, my academic writing was sometimes hard to read, and I thank him for helping me refine my writing patiently many times. When I was very depressed and didn't feel confident, his encouragement helped me adjust my mood and move on. I also thank him for writing me countless reference letters, which gave me many opportunities to try out different disciplines and different programs.

I would like to thank many other professors who supported me complete this thesis. First, I would thank Dr. Marc Brunelle at the University of Ottawa for teaching me courses on sound change, helping me develop the topic of this thesis, and giving me suggestions on the introduction and methods sections. His love for the field of sound change has always influenced and encouraged me. I would not have pursued a graduate career in linguistics if it weren't for him, who noticed my enthusiasm for the phonetics-phonology interface and has helped and advised me along the way. I would also thank Dr. Marion Caldecott, who taught me graduate phonology. Through discussing with her, I gained lots of insights into the phonetics-phonology interface. I would also thank Dr. Henny Yeung and Dr. Yue Wang, who offered me RA positions and taught me how to conduct good lab research. Dr. Bryan Gick at the University of British Columbia and Dr. Suzy Ahn at the University of Ottawa also taught me research methods on experimental phonetics and phonology. I would also thank Prof. Ian

Bercovitz for helping me with statistics during the summer vacation, and Dr. Koichi Haseyama for helping me with my writing. I would also love to thank all the other professors and staff at SFU, including Dr. Heather Bliss, Dr. Maite Taboada, and so on, who taught me courses and supported me throughout my masters.

I would also thank the Cantonese speakers who spent their valuable time participating in this study. I would especially thank Yvonne and Mina, who also invited their friends to participate.

I would also like to thank my linguistic peers at SFU, UOttawa, and the University of Toronto, who gave me lots of help and emotional support. Thank you, Keith Leung, for helping me with recruiting participants and writing Praat scripts. Thank you, Tan Ta, Jiang Song, and Zhanao Fu, for helping me with the GAMM models. Thank you, Irina Presnyakova and Trevor Block, for proofreading my thesis and revising my grammatical mistakes. I would also thank Sylvia Cho, Dasha Gluhareva, Yuchen Li, Lydia Castro, Xizi Deng, Yifang Yuan, Lucas Chambers, Zachary Gilkison, Rosemary Webb, Ronald Nederveen, and Kelli Finney for being wonderful peers in the department and for making my time at SFU warm and fun.

I would also thank my friends outside of linguistics, who supported me and brought me lots of fun: Sitong Zhai, Ziwei Yan, Lelei Zhang, Jiazheng Zhang, Yiyi Shi, Yizhou Sun, and so on.

In the end, many thanks to my family members, who have always been there for me in my life.

Again, I would give my most sincere thanks to all the kind people that I mentioned and not mentioned above. I will try my best to pass on your wonderful attributes and contribute to academia.

Table of Contents

List of Tables

List of Figures

List of Acronyms

HK	Hong Kong
MPEH	Mixed physiological-enhancement hypothesis

Chapter 1

Introduction

1.1 The intrinsic F0 effect and the tension within the IF0 effect in tonal languages

1.1.1 Mechanism of the intrinsic F0 effect

It is generally agreed that the intrinsic F0 (IF0) effect is the tendency for high vowels to have higher F0 than low vowels, which was first reported in German by Meyer (1896-7). Until now, the mechanism of the IF0 effect is still under debate. According to Van Hoof and Verhoeven (2011), three hypotheses may account for the mechanism: the physiological hypothesis, the enhancement hypothesis, and the mixed physiological-enhancement hypothesis.

The general idea of the physiological hypothesis is that the IF0 effect is intrinsic and biomechanical. It is essentially an automatic by-product of vowel articulation, as anatomical links between the tongue and the larynx can affect F0 in coarticulation (Honda, 1983). Detailed discussions of the various explanations that have been put forward in this perspective, including the tongue pull hypothesis, can be found in Dyhr (1990), Fischer-Jørgensen (1990) and Sapir (1989). The thrust of the enhancement hypothesis is that speakers of certain languages may actively and independently control IF0 in order to perceptually enhance vowel contrasts (Diehl, 1991; Diehl & Kluender, 1989a, b; Kingston, 1992; Kingston & Diehl, 1994). This hypothesis is based on research from Syrdal and Gopal (1986) and Traunmüller (1981), who found that listeners judge vowel height not only based on F1 frequency, but also on the auditory distance between F1 and F0: the smaller this distance, the greater the perceived vowel height. Supporters of this hypothesis argue that IF0 might be caused by speakers actively raising F0 in the production of high vowels (thereby diminishing the distance between F1 and F0) and lowering F0 in low vowels (thereby enlarging the distance between F1 and F0), with the purpose of enhancing the perceptual distinctiveness of high and low vowels.

Researchers favouring the physiological hypothesis (Kingston, 2007; Whalen and Levitt, 1995; Whalen et al., 2007) have stressed that the IF0 effect is found in almost all languages. They have also emphasized that the IF0 effect is found even in infants who have not developed concepts of vowel categories. In contrast, proponents of the enhancement hypothesis have pointed out that laryngectomized speakers also show an IF0 effect which is comparable in size to that of the laryngeal participants (Gandour & Weinberg, 1980; Pettorino, 1987). This finding suggests that tongue pull on the phonatory system should be ruled out as the cause of IF0 in oesophageal speech, and IF0 can be actively and deliberately produced. This implies that IF0 might indeed function as a perceptual cue to vowel height. This implication has been supported by Petersen (1986, as cited in Van Hoof & Verhoeven, 2011), who found that IF0 plays an important role in the perception of vowels with ambiguous formant values. However, none of these arguments can support one hypothesis over the other, as there is evidence for both.

Van Hoof and Verhoeven (2011), who noticed that the *physiological hypothesis* and the *enhancement hypothesis* are not necessarily mutually exclusive, proposed the *mixed physiological-enhancement hypothesis (MPEH)* which assumes that IF0 is physiologically determined but may be enhanced or restrained by speakers to help with the perception of phonological contrasts in a language. To be more specific, it is proposed that IF0 is language-universal, but its size is language-specific. The size of the IF0 effect may depend on the extent to which a particular language exaggerates or supresses it to enhance the phonological contrasts in the language. This hypothesis can accommodate experimental evidence supporting both the physiological hypothesis and the enhancement hypothesis.

The *MPEH* is also supported by experimental results found by Van Hoof and Verhoeven (2011). They measured the average F0 difference between high vowels and the low vowels and compared it among native speakers of Moroccan Arabic and Dutch (a language with 12 vowels). Since the vowel inventory is much larger in Dutch than in Arabic, there may be a greater need to perceptually enhance vowel contrasts in Dutch than in Arabic. IF0 would be at least one of the options available to Dutch speakers to maximize the perceptual difference between high and low vowels, thus freeing up more perceptual space for the vowels with intermediate degrees of opening. The results of this study showed that no spectral differences were found between the corresponding vowels of Arabic and Dutch, which implies that the physiological coarticulatory effect of IF0 should be similar in these two languages. The results also showed that the IF0 effect existed in both languages, with native speakers of Dutch showing a larger IF0 effect than native speakers of Arabic. These results support the hypothesis that although the IF0 is automatic and occurs across languages, it may be exaggerated more in languages with a larger vowel inventory and more levels of vowel height to enhance vowel contrasts in speech perception. The *MPEH* is further supported by Hoole and Honda (2011) and Hoole et al. (2006), who found that IF0 emerges as an automatic consequence of vowel articulation, and that speakers sometimes control their cricothyroid muscles, which can tilt the thyroid forward to help tense the vocal cords, to enhance the mechanically-driven IF0 effect. High vowels showed greater CT activity compared to low vowels, which may be a strategy that speakers adopt based on the mechanical coarticulatory effect to enhance phonological vowel height distinctions. Bang et al. (2018) studied the relationship between the size of the IF0 effect and the role of F0 within a single language, Seoul Korean, in which a relevant phonetic change is currently in progress: the primary cue to a stop contrast in phrase-initial position is shifting from VOT to F0 (Kang & Han, 2013). The results showed that the size of the IF0 is being attenuated over historical time, as F0 gradually becomes a primary cue for the stop contrast. Additionally, the attenuation of the IF0 effect was larger for aspirated stops, the stop category that is affected most by the sound change. As the listeners gradually use F0 as the primary cue to perceive voicing contrasts, the IF0 effect may create a challenge for them because now the F0 serves two different phonological purposes simultaneously. So, the speakers may need to supress the IF0 effect to maintain the F0 difference between the voicing categories. Therefore, the findings of Bang et al. (2018) also support the *MPEH*.

1.1.2 The tension within the IF0 effect in tonal languages

According to Van Hoof and Verhoeven's (2011) *MPEH*, IF0 is physiologically determined but may be enhanced or restrained by speakers to help with the perception of phonological contrasts in a language, thus causing differences in the size of the IF0 effect in different languages or different varieties of a language. Van Hoof and Verhoeven (2011) further proposed that the size of the IF0 effect may depend on the extent to which a language 1) exploits it to enhance vowel contrasts or 2) restrains it to preserve tonal contrasts. In the 2011 study, the hypothesis that the size of the IF0 may depend on the extent to which a language exploits it to enhance vowel contrasts was examined on two non-tonal languages (Arabic and Dutch), which was mentioned in the section above. However, they did not test the second proposal that the size of the IF0 effect may also depend on the extent to which a language restrains it to preserve tonal contrasts, as tonal languages, especially the ones with vowels of multiple heights, also need to enhance the vowel height contrasts. Therefore, according to the MPEH, there might be a tension in tonal languages between exaggerating IF0 to enhance the vowel height contrast and supressing IF0 to maintain the tonal contrast.

Previous cross-linguistic studies that compare the size of the IF0 effect between tonal languages and non-tonal languages have yielded evidence in favour of such a tension. Whalen and Levitt (1995) found that the mean IF0 values for tonal languages were substantially smaller than those for non-tonal languages in the survey of the IF0 effect in 31 languages, both tonal and non-tonal. Connell (2002) investigated the IF0 effect in four African tonal languages, and found that the mean IF0 values of tonal languages were generally smaller than that in intonational languages, particularly in Mambila which showed no significant IF0 effect. He also found that in the tonal languages, the tone inventory size alone did not affect the size of the IF0 effect. Instead, the nature of the tone system, especially the degree of F0 modulation used in producing tonal contrasts, was likely to be the primary determinant of the size of the IF0 effect. Sonderegger et al. (2017) conducted a cross-linguistic study of the IF0 effect in 14 languages, including both tonal and non-tonal ones. They fitted a linear mixed effects model to the dataset to compare the direction of IF0 effect (positive or negative) and the size of the IF0 effect across the 14 languages. German showed the greatest positive relation between F0 and vowel height (2.13 semitones, $p < .001$), while Vietnamese demonstrated a non-signficant relation between F0 and vowel height (0.39 st, $p = 0.26$). Sonderegger et al. (2017) reported that the IF0 effects they found in the languages were not as robust as those in Whalen and Levitt (1995). They did not find clear differences in IF0 effect between tonal and non-tonal languages. However, the tonal languages, which are Vietnamese, Thai, Hausa and Mandarin, generally showed smaller IF0 effects than non-tonal languages. Languages with a more complex tone system, which are Vietnamese and Thai, showed weaker IF0 effects than those with a less complex tone system, which are Hausa and Mandarin.

Findings of the studies cited above may imply that IF0 is restrained more in tonal languages, especially languages with complex tones, to maintain the tonal contrasts. However, the findings also show some inconsistencies, as there are also non-tonal languages that show smaller IF0 than tonal languages. Although different methodologies in different studies or different data sources in different languages may partially account for the inconsistencies, the inconsistencies may also imply the tension between exaggerating IF0 to enhance the vowel height contrast and supressing IF0 to maintain the tonal contrast. For example, if a tonal language has more levels of vowel heights than another non-tonal language, it may also exaggerate IF0 more to enhance vowel height contrasts, which may make IF0 in this language smaller than in the non-tonal one. This tension may also partially account for why the degree of F0 modulation used in producing tonal contrasts affects the size of the IF0 effect in tonal languages, while the tone inventory size does not (Connell, 2002). In other words, when comparing tonal languages to non-tonal languages, the studies cited above did not control the need for enhancing vowel height contrasts to be similar among languages. To be more specific, those studies did not control the levels of vowel height in each vowel system; nor did they control the extent to which the speakers/listeners rely on vowel height when distinguishing vowels. Since vowels can also be distinguished by other factors such as frontness, roundedness and length, it is not safe to attribute the smaller size of the IF0 effect in the tonal languages found by Connell (2002) and Whalen and Levitt (1995) only to the need to preserve tonal contrasts in perception. According to Sonderegger et al. (2017), some tonal languages show a larger IF0 than non-tonal languages. This finding may also imply that the need for suppressing the IF0 effect to maintain the tonal contrasts might not be as strong as the need for exaggerating the IF0 effect to maintain the vowel height contrasts in these tonal languages with a larger IF0 effect. To summarize, due to the inconsistent findings and lack of variable control in these cross-linguistics studies, further studies are required to determine whether IF0 is really restrained more in tonal languages, especially in those with complex tone system, or easily confusable tones. The tension between maintaining the vowel height contrast and maintaining the tonal contrast in tonal languages also requires further research.

Different from the cross-linguistic studies above, Siddins & Harrington (2015) investigated a single tonal language, Hong Kong Cantonese, to study whether the three level tones in this language can be influenced by IF0 effects. Two typical close vowels: /i/ and /y/, and two typical open vowels: /a/ and /a:/ in CVN syllables were pronounced with the three level tones to test the effect of vowel height on different ranges of each speaker’s F0. An IF0 effect was found in all the three tones, and IF0 was suppressed more in the mid-low F0 area than the mid-high area. Similarly, Whalen & Levitt (1995) found that the IF0 effect disappears at the low end of a speaker's F0 range cross-linguistically. Furthermore, Connell (2002) found that the size of the IF0 effect is generally reduced or neutralized for low tones. What's more, previous research (Hallé, 1994; Erickson, Honda, Hirai & Beckman, 1995; Honda et al., 1999) has suggested that a different mechanism is involved in the control of F0 at the low end of a speaker's range, which includes increased SH activity and/or vertical movement of the larynx. These findings imply that the IF0 effect appears to be universal and constrained by physiological factors. Therefore, although there is a stronger need to maintain the tonal contrasts in the mid-low area due to a narrower tonal space in this area, we cannot conclude from findings by Siddins & Harrington (2015) that IF0 is suppressed more in the mid-low area due to the need to maintain the tonal contrast. This is because the size of the IF0 effect in this area may be intrinsically smaller due to physiological reasons.

In summary, the MPEH and previous findings imply that in tonal languages, there may be a tension between exaggerating the IF0 effect to enhance the vowel height contrasts and restraining the IF0 effect to keep the easily confusable tone pairs contrastive. This tension needs to be tested in future experiments with better variable control. I will investigate this tension, and thus test the *MPEH* by studying the IF0 effect in the six Hong Kong Cantonese tones. This approach will be informative for the following two reasons. First, the distribution of the six tones within a speaker's normal F0 range as well as the perceptual correlates of these tones are very appropriate for testing the MPEH with good variable control. This is because Hong Kong Cantonese has 1) both a tone in the higher F0 range and tones in the mid-low F0 range, 2) both tone pairs that are easily confusable and relatively distinctive, and 3) both level tones and contour tones. Secondly, the Hong Kong Cantonese vowel /i/ and vowel /ɛ/ are both long, unrounded front vowels, differing only in height (Zee, 1999). I will discuss the distribution of these six tones in a speaker's normal F0 range and the perceptual correlates of these tones in Section 1.2.

1.2 The six tones in Hong Kong Cantonese and their perceptual correlates

1.2.1 The six tones in Hong Kong Cantonese and their F0 traces

It is commonly agreed that in Hong Kong Cantonese, each syllable (usually corresponding to a morpheme) carries a tone, and that the tone system includes six lexical tones and three allotones (Bauer & Benedict, 1997). The lexical tones appear in open syllables or syllables with nasal endings [-m, -n, -ŋ]. They are (using Chao's 1930, 1947 notation) T1 (high-level [55]), T2 (high-rising [25]), T3 (mid-level [33]), T4 (low-falling), [21]), T5 (low-rising [23]), and T6 (low-level [22]). Figure 1 taken from Mok et al. (2013, p. 343) demonstrates the F0 traces of the six lexical tones produced by Peggy Mok, who distinguishes all six tones clearly in production.

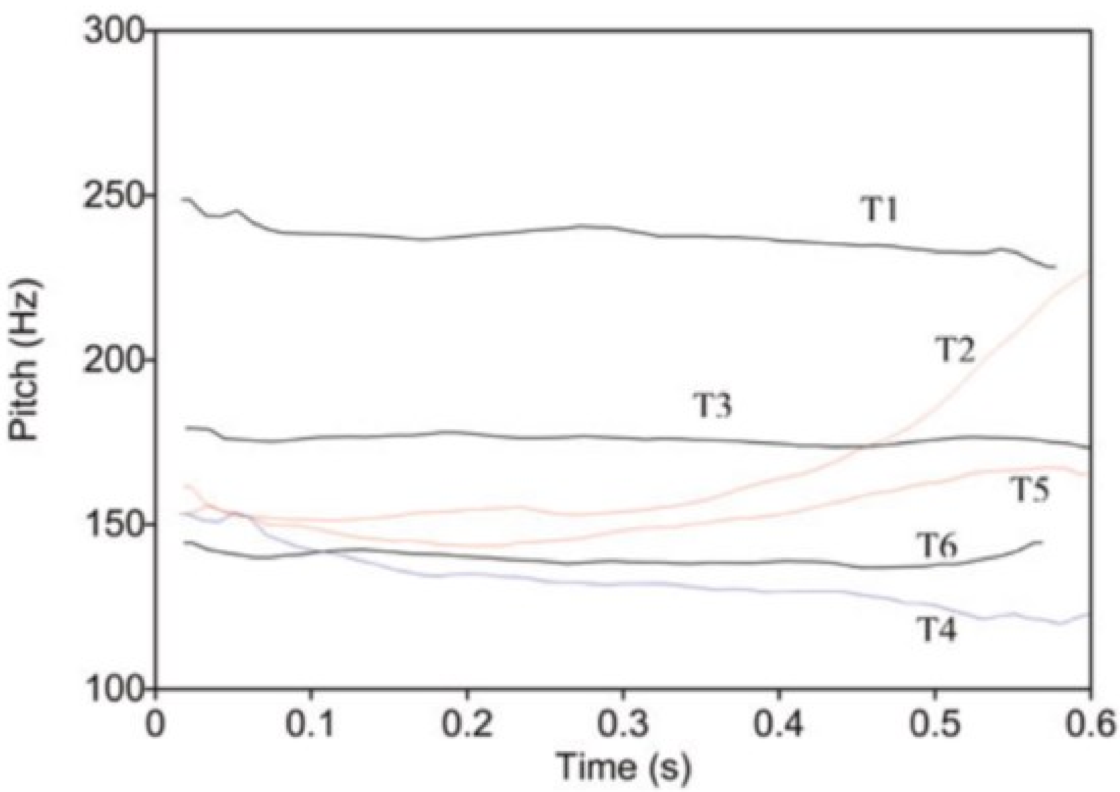


Figure 1.1 F0 traces of the six lexical tones on the syllable [ji] spoken by a female speaker (reprinted from Mok et al., 2013; permission not required)

Figure 1.1 shows that the F0 of T4 is relatively low, so if the IF0 effect is physiologically determined, the size of the IF0 effect would be smaller for T4. The tonal space is very small in the pitch range where T2, T3, T4, T5, and T6 are located, but the tonal space between T1 and the other tones is relatively large. Therefore, if the *MPEH* is correct, T1 should have less need for supressing the IF0 effect to maintain the tonal contrasts.

Previous studies have suggested that F0 is a primary and sufficient cue for tonal distinctions in Cantonese (e.g., Gandour, 1981; Khouw & Ciocca, 2007; Vance, 1976). These studies also suggested that F0 change (direction and magnitude) and relative F0 level are the acoustic cues for tone identification. Therefore, tones that are similar in these F0-based acoustic properties (having similar contours and/or having little contrast in height) may be easily confused by speakers. For example, T3 and T6 are both level tones with a very narrow tonal space between them. T2 and T5 are both rising tones with the same starting point [2]. T5 and T4 are both contour tones with only a small change in F0 from the onset to the offset, so they may be confusable with T6 which shares the same starting point [2] with them. If these tones were really easily confusable, according to the *MPEH*, each of them would have a stronger need for supressing the IF0 effect to maintain the tonal contrasts.

1.2.2 Confusable tone pairs and tonal variation in Hong Kong Cantonese

Previous studies have reported confusion in certain tone pairs in Hong Kong Cantonese. So and Varley (1991), for instance, found that some subjects in a group of 101 Hong Kong Cantonese speakers perceptually confused T2 and T5, while Kei et al. (2002) found that six of fifteen subjects did not distinguish T2 and T5 in their production Of these six subjects, two produced T2 as T5, three produced T5 as T2, and one produced tone contours with endpoints midway between T2 and T5. Bauer et al. (2003) similarly found that two male speakers of Hong Kong Cantonese confused T2 and T5 in both directions when producing tones under experimental conditions. Yiu (2009) investigated both the production and perception of this tone pair, and found that five of fifteen participants were not able to distinguish T2/T5 on some occasions in perception, while three appeared to confuse the two rising tones in production. Additionally, the participants' production performance did not predict their perception performance, and vice versa. Similarly, Whitehill, Ciocca, and Chow (2000) reported no significant difference on average F0 in T3 and T6, which implied that these two tones may be confusable as well.

Based on these findings, Mok et al. (2013) studied the production and perception of tone merging in Hong Kong Cantonese. The authors did not indicate under which condition two tones can be defined as merging (the sound change process), nor did they provide diachronic evidence that the two tones are really in a merging process. Any two tones are believed by the authors to be merging when they are produced similarly in production or easily confused in perception. In an attempt to locate merging speakers, they recruited young participants and determined whether a participant is a merging participant or a controlled participant based on the results of a simple production test: recordings of each participant were checked by native speakers of Cantonese who clearly distinguish all six tones to determine which speakers tended to merge the tones. In the production experiment, the identified merging participants were asked to read monosyllabic words carrying the six tones. Then, predictive discriminant analysis (Tabachnick & Fidell, 2018), which is a multivariate test that shows how different cases group themselves based on a set of variables, was conducted to examine whether the merging tones produced by each participant were distinct from each other. In the predictive discriminant analysis, the dependent variable, tone, was taken as the grouping variable, and four independent variables were used as the predictors: the F0 values at the second, fifth, sixth, and ninth measurement points of the pitch contour, representing the onset, middle, and offset pitch values of the tone. Results of the predictive discriminant analysis (misclassification rates) were compared across the merging speakers and across different tone pairs. Misclassification rates of the merging participants were also compared to reference (non-merging) participants, who are the two authors of the paper. According to the misclassification rates, Mok et al. (2013) concluded that the merging of all three pairs is incomplete, as most merging speakers showed only a partial overlap in the merging tone pairs. Additionally, the misclassification rates showed that the merging of tones is not symmetrical. For most speakers, T2 tokens were more often misidentified as T5 by discriminant analysis than T5 tokens as T2, and T4 tokens were more often misclassified as T6 than T6 tokens as T4. In the perception experiment on both the merging participants and the control participants using AX discrimination tasks, the merging participants could still distinguish the six lexical tone categories in perception, despite being slower in tone perception than the control group in general (not only for the identified merging tone pairs). Mok et al. (2013) also found that tones with lower type frequency were more prone to merge.

In order to study tone merging in more detail, Fung and Lee (2019) investigated production and perception of all the Cantonese tone pairs. They used the term tone merger to denote the collapse of the phonetic contrast of two tones and coined the term “tone mergerer” to refer to a person who merged two tone categories. In the production experiment, the statistical output “Pillai score” of a multivariate analysis of variance (MANOVA) models (Hay et al., 2006) was adopted to examine the degree of overlap between the trajectories of the two tones. Participants receiving a Pillai score with statistical significance (i.e., $p < 0.05$) were considered distinct speakers, while those without statistical significance (i.e., $p > 0.05$) were considered tone mergerers. In the perception experiment, participants who achieved 62.5% correct or lower in the AX discrimination task were considered as tone mergerers. The results (Table 1.2.2a, 1.2.2b, 1.2.2c) showed that three tone mergers: T2 and T5, T3 and T6, T4 and T6 were found to be tone mergers in production and/or perception. For all three tone pairs, more than half of the participants did not show confusion in either production or perception (60% of the participants for T2-T5, 53.33% for T3-T6, 68.33% for T4-T6). For T2-T5, 17.5% of the participants confused them in both production and perception, 5% confused them in only production, and 17.5% confused them only in perception. For T3-T6, none of the participants confused them in both production and perception, 46.67% confused them in only production, and none confused them only in perception. For T4-T6, 1.66% of the participants confused them in both production and perception, 13.33% confused them in only production, and 16.67% confused them only in perception.

Table 1.1 Classification of four types of participants (merged vs unmerged) of (a) T2-T5; (b) T3-T6; and (c) T4-T6 (reprinted with permission from Fung &Lee, 2019)

(a) T2-T5

Production \ Perception	Unmerged	Merged
Unmerged	60%	17.5%
Merged	5%	17.5%

(b) T3-T6

Production \ Perception	Unmerged	Merged
Unmerged	53.33%	0
Merged	46.67%	0

(c) T4-T6

Production \ Perception	Unmerged	Merged
Unmerged	68.33%	16.67%
Merged	13.33%	1.66%

To summarize, previous findings have suggested that three tone pairs in Hong Kong Cantonese (T2 and T5; T3 and T6; T4 and T6) are sometimes confused in production and/or perception in some native speakers of Hong Kong Cantonese. Bauer et al. (2003) Fung and Lee (2019) and Mok et al. (2013) all used merge/merging/merger to describe the confusion of certain tone pairs, and believed that the confusion implies that these tone pairs are undergoing tone merging, which is a sound change process. Since some speakers confuse certain tone pairs while others do not, there is variation across Hong Kong Cantonese speakers in the production and/or perception of certain tone pairs, which implies a possibility that Hong Kong Cantonese is undergoing a merging process. However, experimental results of these three studies, as well as other studies of Cantonese lexical tones, only showed cross-speaker variation in production and/or perception, and tones with lower type frequency were more prone to be confused. These findings are not sufficient to support the hypothesis that a sound change is in progress, as opposed to the observation that easily confusable tones have been very similar for a long time. Until now, no longitudinal experimental studies have compared Hong Kong Cantonese tone production/perception at two timepoints. Similarly, no cross-sectional studies focusing on tone production/perception with two generations of Hong Kong Cantonese have been conducted so far. Therefore, it remains unclear whether Cantonese tone pairs are merging or not.

What is important is that according to previous findings, no matter whether there is sound change or not, the confusable tone pairs are still phonologically contrastive for most speakers in the Hong Kong Cantonese speaking community. Mok et al. (2013)

demonstrated that although the participants who were judged to be merging showed reduced tone space, they still had six lexical tone categories in production, and could still distinguish the six lexical tone categories in perception. The results of Fung and Lee (2019) indicated that for each of the three confusable tone pairs which were confused by some participants in production and/or perception, fewer than half of the participants at the most showed confusion. These findings suggest that the confusable tone pairs are still phonologically contrastive, so if the *MPEH* is correct, each of the easily confusable tones should have a relatively stronger need for supressing the IF0 effect to make it acoustically different from the other tones.

1.2.3 Perceptual correlations of the confusable tone pairs

Previous perceptual studies have suggested that fundamental frequency (F0) is a primary and sufficient cue for tonal distinctions in Cantonese (e.g., Fok, 1974; Gandour, 1981; Vance, 1976) and that F0 change (direction and magnitude) and relative F0 level are also relevant. Following these studies, Khouw and Ciocca (2007) conducted a production experiment and a perception experiment to investigate more deeply how listeners rely on F0 change and F0 level to identify naturally produced Hong Kong Cantonese tones. These researchers showed that in both production and perception, F0 change over the later portion (the sixth and seventh sections out of eight sections) of the vocalic segment separates the six lexical tones into four groups: the high rising tone (T2), the low rising tone (T5), the three level tones (T1, T3, T6), and the low falling tone (T4). These findings suggest that both the direction and the magnitude of F0 change provide possible perceptual cues for distinguishing among contour tones. This is consistent with the findings of previous perceptual studies (Fok, 1974; Gandour, 1981; Vance, 1977). Khouw and Ciocca's finding also indicates that the production of T2 and T5 is acoustically separated based on F0 change over the later portion of the vocalic segment. This is consistent with Ciocca et al. (2002), who found that T2 has a larger magnitude of F0 increase as compared to T5, and with Kei et al. (2002), who demonstrated that listeners rely on the difference between the endpoints of the two tone contours as the salient perceptual cue.

For F0 level, Khouw and Ciocca (2007) showed that in both production and perception, mean F0 level separates the six lexical tones into four groups: T1 with the highest average F0, T4 with the lowest average F0, T3 with average F0 halfway between T1 and T4, and T2, T5, T6 with average F0 between tone T3 and T4. A relatively larger separation on average F0 was found between productions of T1 and T3, compared to the one between tones T3 and T6. This finding is consistent with that of Whitehill, Ciocca, and Chow (2000), who showed that the overall F0 of T1 is significantly higher than that of T3 and T6. However, Khouw and Ciocca (2007) found a a significant difference on average F0 between T3 and T6, while Whitehill, Ciocca, and Chow (2000) reported no a significant difference . Khouw and Ciocca (2007) found that T2 and T5 are grouped between T3 and T4, whereas Gandour (1981) found that these two tones are grouped between T1 and T3. Although there are inconsistencies, overall, it is clear that F0-based acoustic properties are cues for Cantonese lexical tones.

A question that arises is whether non-F0-based acoustic properties are also cues for Cantonese lexical tones. Gandour and Harshman (1978) relied on multidimensional analysis to extract perceptual dimensions that underlie perception of tones by listeners whose language backgrounds were Thai, Yoruba, or American English. Their results showed that the only non-F0-based perceptual dimension is duration of the vocalic segment. In Mandarin, possible non-F0-based perceptual correlates of native speakers are duration (Blicher, Diehl, & Cohen, 1990), creaky voice (Gottfried & Suiter, 1997), as well as dynamic cues, such as temporal position of turning point in the F0 contour (Liu & Samuel, 2004), and frequency difference between the onset F0 and the lowest turning point in the F0 contour (Shen & Lin, 1991). However, studies have not provided evidence that syllable duration is an important cue in Cantonese either in tone production (Vance, 1976) or in tone perception (Fok, 1974). Furthermore, Khouw and Ciocca (2007) found that T2 and T5 show similar temporal positions of turning point in the F0 contour, which implies that duration or temporal position of turning point may not be important cues to distinguish Cantonese T2 and T5. Therefore, although some non-F0-based cues, such as duration and voice intensity might cue the perception of lexical tones, they appear minimally relevant. However, Yu and Lam (2014) found that creak occurs systematically more often on T4 than other tones. Listeners identified T4 with 20% higher accuracy when it was realized with creak than when it was not and showed a higher proportion of T4 responses for creaky stimuli in a two-alternative forced choice task of identifying stimuli as T4 or T6. This finding implies that differences in voice quality contribute to the perception of Hong Kong Cantonese tones alongside F0, especially for T4.

1.3 Research questions and predictions of this study

This study will investigate the IF0 effect in tonal languages in more detail by considering the tension between exaggerating the IF0 effect to enhance the vowel height contrasts and restraining the IF0 effect to keep the easily confusable tone pairs contrastive. The mechanism of the IF0 effect will be studied by comparing the magnitude of the IF0 effect in the six Hong Kong Cantonese tones produced by native speakers. The goal is to test the *MPEH*, which states that IF0 is physiologically determined but may be enhanced or restrained by speakers to help with the perception of phonological contrasts in a language.

According to Zee (1999), in Hong Kong Cantonese, /i/ and /ɛ/ are both long, unrounded front vowels and are, respectively, close and open-mid, differing only in height. Therefore, if the *MPEH* is correct, there might be a need to exaggerate the IF0 effect to enhance the vowel height contrast in Hong Kong Cantonese. Additionally, due to the fact that the three easily confusable tone pairs (T2 and T5, T3 and T6, T4 and T6) are sometimes confused by speakers/listeners, and these tone pairs are still phonologically contrastive in the language community, there might be a stronger need to restrain the IF0 effect to preserve the tonal contrasts for these easily confusable tone pairs. These facts imply that there may be a tension between exaggerating the IF0 effect to enhance the vowel height contrasts and restraining the IF0 effect to keep the easily confusable tone pairs contrastive.

As discussed above, in Hong Kong Cantonese, T2/T5 and T3/T6 are both confusable pairs, and all these four tones are roughly in the same F0 range. Therefore, when comparing the size of the IF0 effect of these tones, especially when comparing T6 to T5 (two relatively lower tones) and T2 to T3 (two relatively higher tones), we can ensure that any observed difference in the size of the IF0 effect is due to different degree of suppression or exaggeration of IF0, but not the result of reduction or neutralization in the lower F0 range.

Previous studies have shown that the only reliable cue for distinguishing T3 and T6 is the mean F0 value, while the only reliable cue for distinguishing T2 and T5 is the magnitude of F0 change over the later portion of the vocalic segment. Most previous research on IF0 measured the F0 values of the vowel mid points (Chen et al., 2021;

Connell, 2002), and assumed IF0 to be a simple additive or subtractive effect on the vowels. Therefore, even if the IF0 is exaggerated to enhance the vowel height contrast between /i/ and /ɛ/, it should not make T2 and T5 more easily confused by listeners. However, if IF0 is exaggerated to enhance the vowel height contrast between /i/ and /ɛ/, this will make T3 and T6 more easily confused. Since the three easily confusable tone pairs reported in previous studies of Hong Kong Cantonese are T3 and T6, T2 and T5, and T4 and T6, T2 should not be confused with any other tones except T5. Furthermore, T5 should not be confusable with any other tones except T2. Thus, there is not a strong need to enhance the contrast between T2 and any other tones except T5, and there is not a strong need to enhance the contrast between T5 and any other tones except T2, either. Due to these facts, I assume if the *MPEH* is correct, T2 and T5 will show a larger IF0 effect than T3 and T6. Additionally, T6 is confusable with two tones (T3 and T4), and is lower than T3, so the IF0 effect of T6 may be smaller than that of T3.

If the *MPEH* is correct, then IF0 should be both physiologically determined and enhanced (or restrained) by speakers to help with the perception of phonological contrasts in a language. If IF0 is physiologically determined, then it is expected that the size of the IF0 should be smaller in the lower F0 range. If IF0 is also controlled by speakers to enhance the phonological contrast, a tone that is *not* easily confusable with other tones will be less influenced by the method of restraining the IF0 effect to enhance the tonal contrasts. Therefore, according to the physiological-enhancement hypothesis, since T4 is the lowest tone and is easily confusable with T6, T4 would show the smallest size of the IF0 effect among the six tones. Since T1 is the highest tone and is not easily confusable with any of the other five tones, T1 would show the largest size of the IF0 effect among all the six tones.

In summary, if the *MPEH* is correct, it is predicted that: 1) Among the six tones, T4 will show the smallest size of the IF0 effect because it is both low and confusable, while T1 will show the largest size of the IF0 effect because it is both high and not confusable. The size of the IF0 effect of the other four tones will be in the middle. 2) Among the four tones that show a middle size of the IF0 effect, T3 and T6 will show a smaller IF0 effect than T2 and T5, because T3 and T6 are level tones. T6 will show a smaller size of the IF0 effect than T3, because it is lower and confusable with more tones. Additionally, previous studies have suggested that Hong Kong Cantonese speakers vary in how much they confuse an easily confusable tone pair in production and/or perception. Therefore, for the easily confusable tone pairs, I will also analyze the acoustic data to study how easily confusable a tone pair is in production across speakers. This is because the degree of confusion for a tone pair can affect the need to maintain the tonal contrast for the two tones in this tone pair, which can further affect the size of the IF0 effect of these two tones. No perceptual experiments will be performed in the current study.

Chapter 2

Methods

2.1 Participants

The participants were twenty female native speakers of Hong Kong Cantonese who were born in Cantonese-speaking families in Hong Kong and lived mostly in Hong Kong before the age of 18. Ten of them were aged between 18-28, with the remainder between 35-45. They were recruited ethically either from Hong Kong, or from the Cantonese-speaking community in the Greater Vancouver area, to verify that they speak Cantonese on a daily basis. When queried, they reported no history of speech, language or hearing disorders.

2.2 Stimuli

The target items were all CV syllables, with V = /i/ or /ɛ/. Since different consonants may affect vowel onset F0 differently, the onset consonant of all syllables was controlled to be fricative /s/ to minimise coarticulatory effects. All six lexical tones in Hong Kong Cantonese were pronounced within these two types of CV syllables. Each of the 12 syllables has a corresponding real Cantonese word with high frequency of occurrence in contemporary speech[1]. The syllables with tones and their corresponding real Cantonese words are shown in Table 2.2. The characters chosen here do not have alternative readings when presented in isolation[2].

---

[1] "/si3/ 弒 (kill)" has a relatively lower, but moderate frequency among these 12 words, but it is also frequently used in Hong Kong Cantonese.

[2] One native speaker of Cantonese who did not participate in the experiment checked the stimuli and confirmed that all these syllables are Cantonese real word syllables with high or moderate frequency. He also confirmed that each Cantonese word that presents the corresponding syllable only has one reading. During the experiment, four of the twenty speakers read "/si3/ 弒 (kill)" incorrectly. These participants were invited to do another recording section, during which they were instructed to read "/si3/ 試 (try)", along with four other stimuli, embedded in the same sentences, for ten repetitions. "/si3/ 試 (try)" has a high frequency, but it has another reading of /si5/, so it was presented to the participants with the English gloss "try" in order to elicit the target reading. Then, for these four participants, their speech of "/si3/ 試 (try)" was analyzed instead of "/si3/ 弒 (kill)".

Table 2.1 Stimuli

	T1	T2	T3	T4	T5	T6
/si/	/si1/ 詩 (poem)	/si2/ 史 (history)	/si3/ 弒 (kill)	/si4/ 時 (time)	/si5/ 市 (city)	/si6/ 事 (thing)
/sɛ/	/sɛ1/些 (some)	/sɛ2/ 寫 (write)	/sɛ3/ 卸 (discharge)	/sɛ4/ 蛇 (snake)	/sɛ5/ 社 (society)	/sɛ6/ 射 (shoot)

2.3 Procedure

The syllables were elicited using Traditional Chinese characters shown in Table 1 above embedded in a Cantonese sentence "我讀__這個字 [I read the word __.]". Each sentence was presented to each participant 10 times, resulting in 120 stimulus sentences in total. The order of the 120 sentences was randomized. Due to constraints on conducting in-person research during the COVID-19 pandemic, the participants were instructed and monitored through live Zoom meetings. The participants were told to use Audacity (Audacity Team, 2020) to record their own speech at a sampling rate of 44100 Hz with 16-bit resolution, using their best available microphones. During the recording, the participants sat in a quiet room, viewed the stimulus sentences on their computer screens, and read the stimuli at a moderate pace. Each participant was rewarded $20 or partial course credit.

2.4 Acoustic and statistical analysis

The initial and final positions of the vocalic segment of each target syllable were demarcated and annotated using a Praat Textgrid. Following Khouw and Ciocca (2007), the initial position of the vocalic segment was set at the first upward-going zero-crossing of the voicing cycle that was closest to each calculated point in the amplitude waveform. The final position of the vocalic segment was set to the beginning of the last vocalic pulse as seen on the spectrographic display.

After annotating the initial and final positions of each vocalic segment, following Gandour (1983), Vance (1977) and Khouw & Ciocca (2007), the F0 values of each target syllable were measured at nine consecutive points of the vocalic segments (initial, 12.5%, 25%, 37.5%, 50%, 62.5%, 75%, 87.5%, final) using a Praat script.

Before conducting the statistics, outliers were removed from the dataset. For each tone by age group combination, I calculated the mean F0 value and standard deviation (SD). Then, in each case I excluded the data outside the range of the mean ± 1.96*SD. Additionally, since T4 is very low and sometimes creaky, many F0 data points measured by Audacity and Praat appeared erroneous. After removing the outliers using the method just described, there were still some data points with very high F0 values. To exclude these, F0 values greater than 250 were removed. The percentage of collected data points remaining in the analysis are shown in Table 2.2.

Table 2.2 Percentage of data points included for each tone by age group combination

	Older (age 35-45)	Younger (age 18-28)
T1	93	97
T2	96	90
T3	96	95
T4	94	88
T5	97	92
T6	97	92

After the F0 measurements were conducted in Praat, the rest of the data analysis was performed in R (R Core Team, 2020). The tone curves were compared through: 1) calculating the mean differences at each datapoint and 2) fitting the data into generalized additive mixed models (GAMMs, Wood, 2017). When calculating the mean differences, similar to Fung and Lee (2019) and Mok et al. (2013), datapoint 1 and 9 (start and end point of the vocalic segments) were excluded to avoid carryover effects, anticipatory effects, or invalid datapoints because Praat sometimes fails to extract F0 values at the beginning and end of vocalic segments. However, in GAMM analysis, these two datapoints were included to better capture time-varying, dynamic tone contours. All the figures in the results section were generated using the default plotting function built into GAMM or ggplot2 (Wickham, 2016) with 95% confidence interval. The full complete statistical outputs are listed in Appendix A.

Chapter 3

Results

In this chapter, first, in section 3.1, GAMM models will be applied to the acoustic data to determine whether each speaker clearly distinguishes between the three confusable tone pairs reported by previous studies (T2 and T5, T3 and T6, T4 and T6. The data will be split up by older and younger speakers in order to study whether there is a sound change in process. Since previous studies have reported inter-speaker variation in how much the speakers confuse the tone pairs, both overall group analysis of each tone pair by age group combination and individual analysis of each speaker will be conducted. These analyses will provide an acoustic measure of how confusable each tone pair is for each age group, with inter-speaker variation taken into consideration. Then, in section 3.2, GAMM models will be applied to the acoustic data to analyze the IF0 effect in each tone by age group combination. Similarly, in addition to the overall group analyses, inter-speaker variation based on individual analyses will also be taken into consideration. The size of the IF0 effect will be compared across the six tones to test the mixed physiological-enhancement hypothesis. The relationship between tone merging and the size of the IF0 effects will also be analyzed to study whether speakers or tones that are more advanced in the merging process will show smaller IF0.

3.1 Confusable tone pairs and tonal variation in Hong Kong Cantonese

In this section, GAMM models will be applied to the acoustic data to determine whether each speaker clearly distinguishes between the three confusable tone pairs reported by previous studies. In the overall group analyses, for each tone pair by age group combination, GAMMs were fitted to the time-varying F0 data, to assess the difference between the two tones in the confusable tone pair. The best GAMM model, which was selected by the model comparison method suggested by Wieling (2018), includes a parametric term for tone, a smooth term for normalised time, a smooth term for a normalised-time-by-tone interaction and random smooths for time-by-speaker-by-tone. The Akaike Information Criterion (AIC) value of each model without auto-correction was compared to the one with auto-correction, and the model with a lower AIC value was selected.

Results of the GAMMs are shown in Table 3.1. T2 is the baseline for T2 and T5, T3 is the baseline for T3 and T6, and T4 is the baseline for T4 and T6. Results show a significant difference s between the two tones of each tone pair by age group combination. The estimates suggest that the younger group shows a smaller difference than the older group between T2 vs T5 and T3 vs T6, and a bigger difference than the older group between T4 and T6. The GAMM results are also plotted in Figure 3.1 and Figure 3.2. The x-axis of both figures shows duration of the tones, which is normalized to be from 0 to 1. The y-axis of figure 3.1 show F0, and the y-axis of figure 3.2 shows F0 difference. In Figure 3.1, each plot contains a pink ribbon and a blue ribbon representing the F0 curves. The three plots in the first row show results for the older group, and the plots in the second row show results of the younger group. In Figure 3.2, each plot contains a grey ribbon, which represents the size of the F0 difference between the two tones. The interval between the two red vertical lines with a red horizontal line on the x-axis shows the time range where the two tones are significantly different based on the GAMM outputs. The figures show that for the younger group's T2 and T5, the T2 shows significantly higher F0 at the beginning and the end, but not in the middle. For the other tone by age group combinations, a significant difference occurs across the entire tone curves.

Table 3.1 GAMM outputs for each tone pair by age group combination

Older (age 35-45)					
	Estimate	SE	*t*	*p*	r^2
T2 and T5	-12.462	0.745	-16.720	< 0.001	0.829
T3 and T6	-12.632	0.511	-24.740	< 0.001	0.877
T4 and T6	25.260	0.802	31.480	< 0.001	0.760
Younger (age 18-28)					
	Estimate	SE	*t*	*p*	r^2
T2 and T5	-4.888	0.938	-5.210	< 0.001	0.551
T3 and T6	-3.743	0.683	-5.482	< 0.001	0.657
T4 and T6	26.869	1.415	18.980	< 0.001	0.527

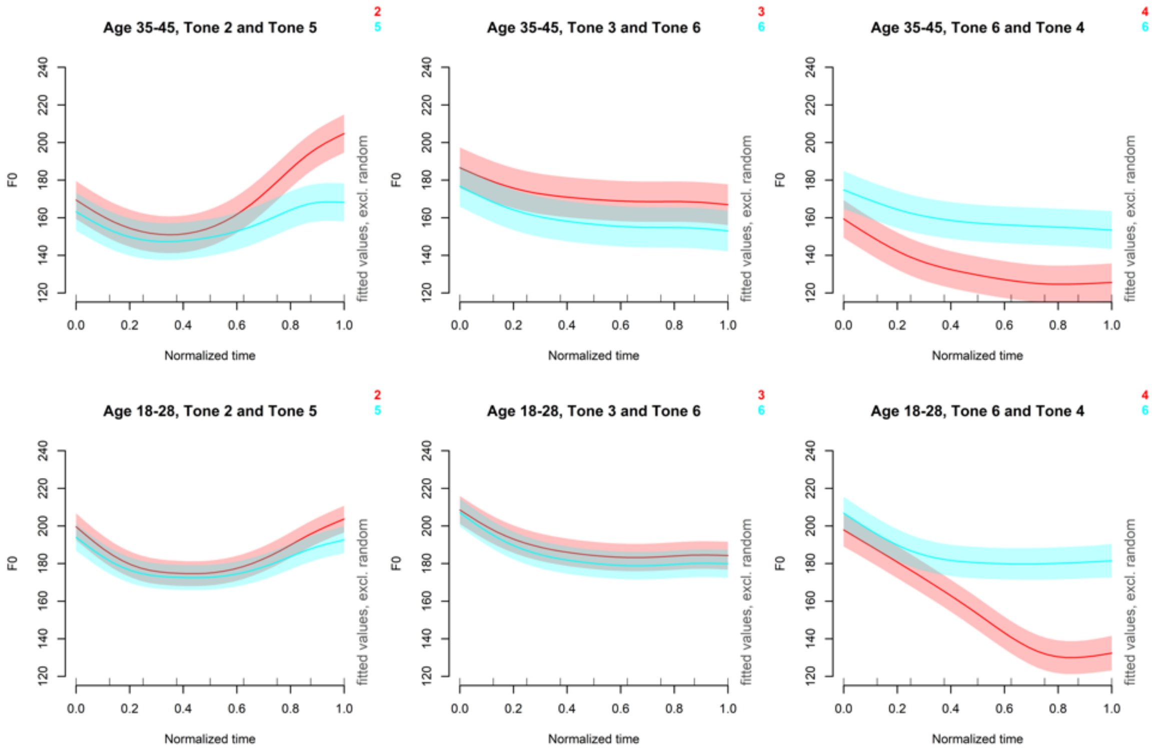


Figure 3.1 Non-linear smooths for each tone pair by age group combination

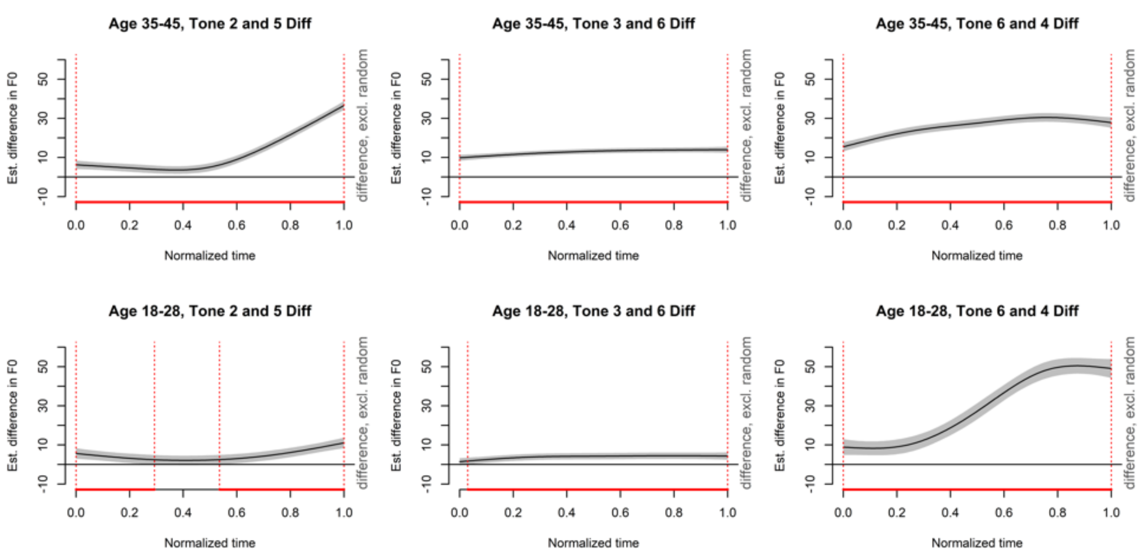


Figure 3.2 Difference smooths for each confusable tone pair by age group combination

To get a clearer picture of inter-speaker variation, in the following sections, inter-speaker variations for each tone pair by age combination will be discussed.

3.1.1 Older group, T2 and T5

Figure 3.1 and 3.2 included the non-linear smooths and difference smooths for T2 and T5 produced by the older group. It can be seen from the GAMM figures that T2 shows significantly higher F0 than T5 from the beginning to the end. The two tones are rising tones, and the difference between them increases from the middle to the end of the vocalic segment. Table 3.2 shows the mean F0 difference between T2 and T5 (T2 - T5) for each F0 measurement point (2-8) for all speakers of the older group. The mean difference is the biggest at timepoint 8, with a value of 29.4 Hz. The mean difference tends to be bigger in the later portion of the tone curves, which aligns with the previous finding that T2 and T5 are acoustically separated based on F0 change over the later portion (the sixth and seventh sections out of eight sections) of the vocalic segment (Khouw & Ciocca, 2007).

Table 3.2 Mean F0 difference between T2 and T5 (T2 - T5) for each F0 measurement point for the older group

Point	2	3	4	5	6	7	8
Mean Diff	6.3	4.0	5.1	6.3	10.9	19.4	29.3
At Point 8, Mean Diff is the biggest = 29.3							

Figure 3.3 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular. Only speaker 9 shows some data points that are noticeably far from the means.

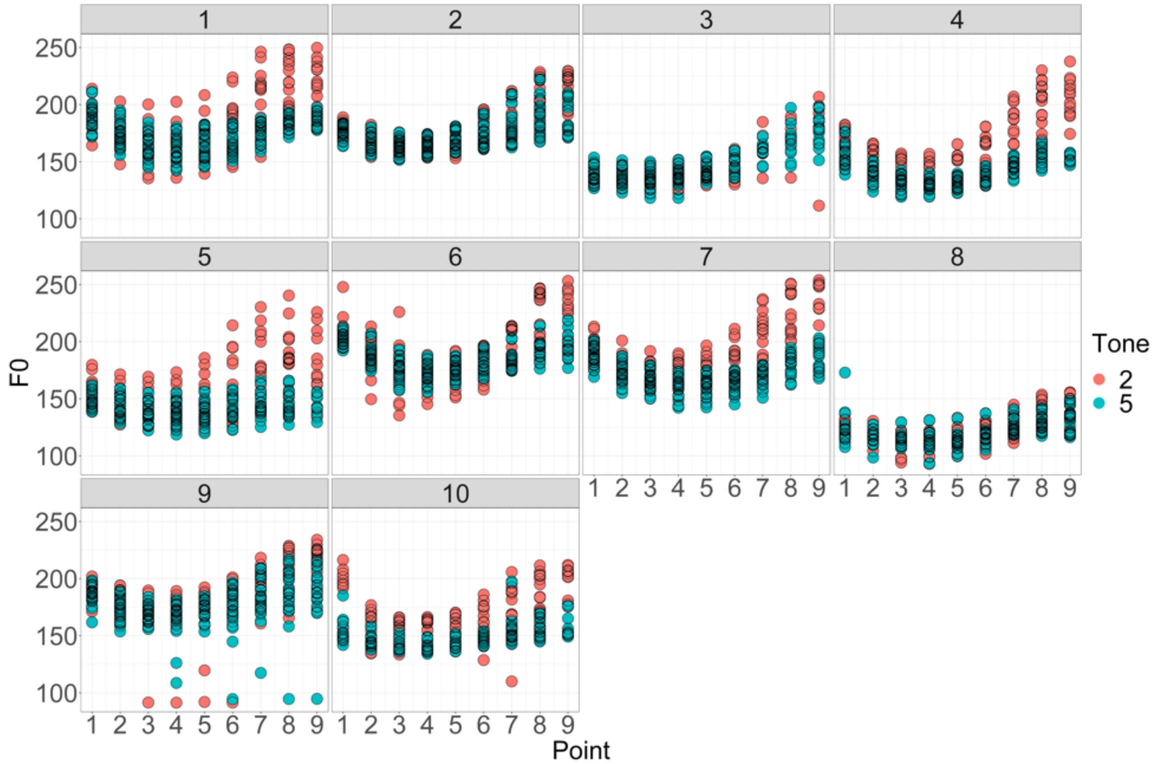


Figure 3.3 **F0 data points for T2 and T5 produced by each speaker in the older group**

Figure 3.4 shows the GAMM non-linear smooths of the two tones produced by each participant. Nine of the speakers (all except for speaker 3) appear to show higher T2 than T5. This finding is also supported by the GAMM difference smooths of the two tones produced by each participant, which are shown in Appendix B.

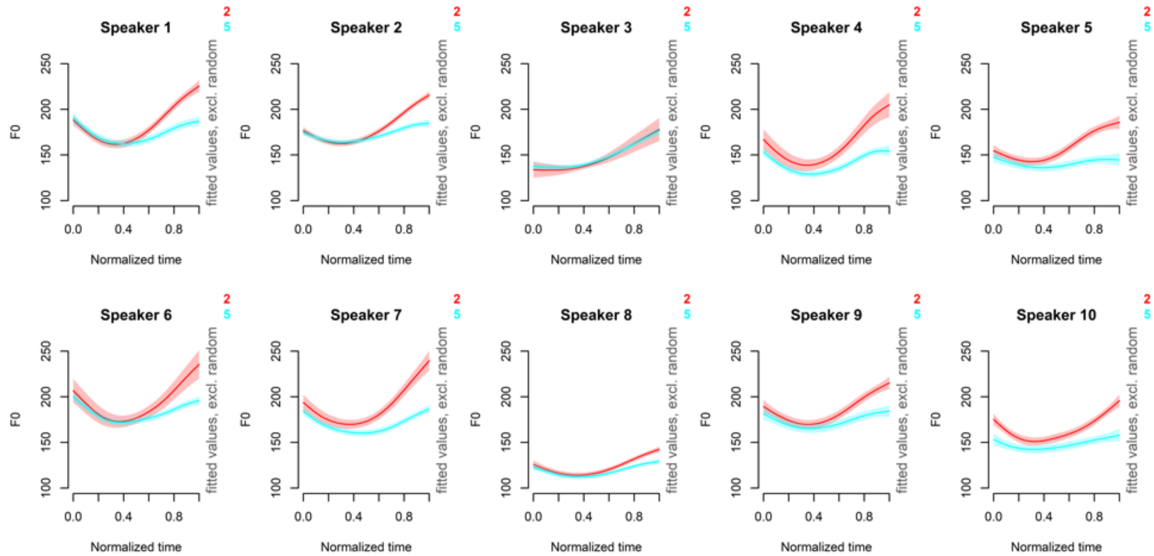


Figure 3.4 Non-linear smooths of T2 and T5 produced by each speaker in the older group

Table 3.3 shows the mean F0 difference between T2 and T5 (T2 - T5) at timepoint 8, where the two tones differ the most, for each of the nine speakers that appear to show higher T2 than T5. All nine speakers except for speaker 8 show that the F0 values of T2 are higher than T5 by about 20-45 Hz. The mean difference of the nine speakers is 30.8 Hz. Table 3.4 shows that only speaker 3 appears to confuse the two tones, and her T2 is higher than T5 by 4.6 Hz at timepoint 8.

Table 3.3 Mean F0 difference between T2 and T5 (T2 - T5) at measurement point eight for each speaker of the older group that appears to show higher T2 than T5

Speaker	1	2	4	5	6	7	8	9	10
Mean Diff	31.9	22.2	44.3	38. 9	34.4	43.8	10.0	28.2	31.9
Mean Diff for the nine speakers = 30.8									

Table 3.4 Mean F0 difference between T2 and T5 (T2 - T5) at measurement point eight for each speaker of the older group that appears to confuse T2 and T5

Speaker	3
Mean Diff	4.6

3.1.2 Older group, T3 and T6

Figure 3.1 and 3.2 included the non-linear smooths and difference smooths for T3 and T6 produced by the older group. The GAMM figures show that T3 is significantly higher than T6 from the beginning to the end. The two tones are both level tones, and the difference between them is roughly the same from the beginning to the end of the tones. Table 3.5 shows the mean F0 difference between T3 and T6 (T3 - T6) for each F0 measurement point (2-8) for all speakers of the older group. The mean difference is the biggest at timepoint 6, with a value of 13 Hz. The mean difference tends to be bigger in the middle of the tone curves. This is reasonable because T3 and T6 are both level tones, and they are acoustically separated based on mean F0 values (Khouw & Ciocca, 2007).

Table 3.5 Mean F0 difference between T3 and T6 (T3 - T6) for each F0 measurement point for the older group

Point	2	3	4	5	6	7	8
Mean Diff	9.4	10.9	11.8	12.5	13.4	13.0	12.9
At Point 6, Mean Diff is the biggest = 13.4							

Figure 3.5 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular. Again, only speaker 9 shows a few data points of T6 that are noticeably far from the means.

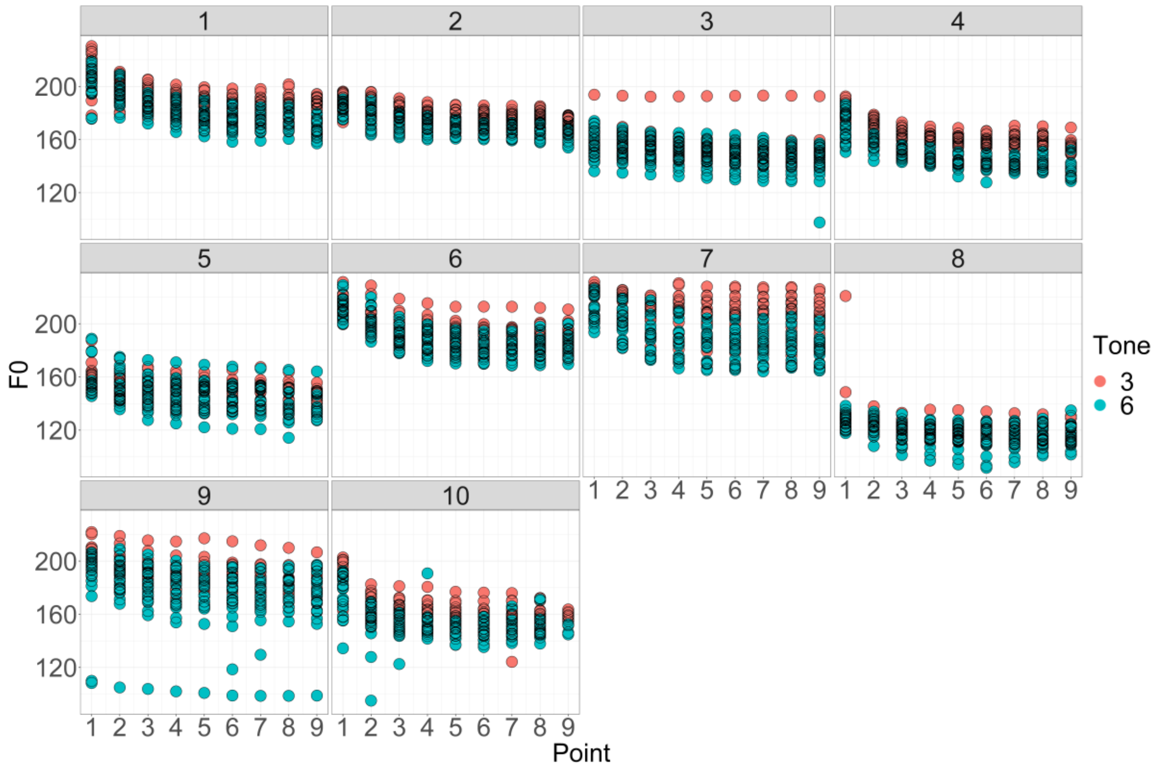


Figure 3.5 F0 data points for T3 and T6 produced by each speaker in the older group

Figure 3.6 shows the GAMM non-linear smooths of the two tones produced by each participant. Nine of the speakers (all except for speaker 3) appear to show higher T3 than T6. This finding is also supported by the GAMM difference smooths of the two tones produced by each participant, which are shown in Appendix B.

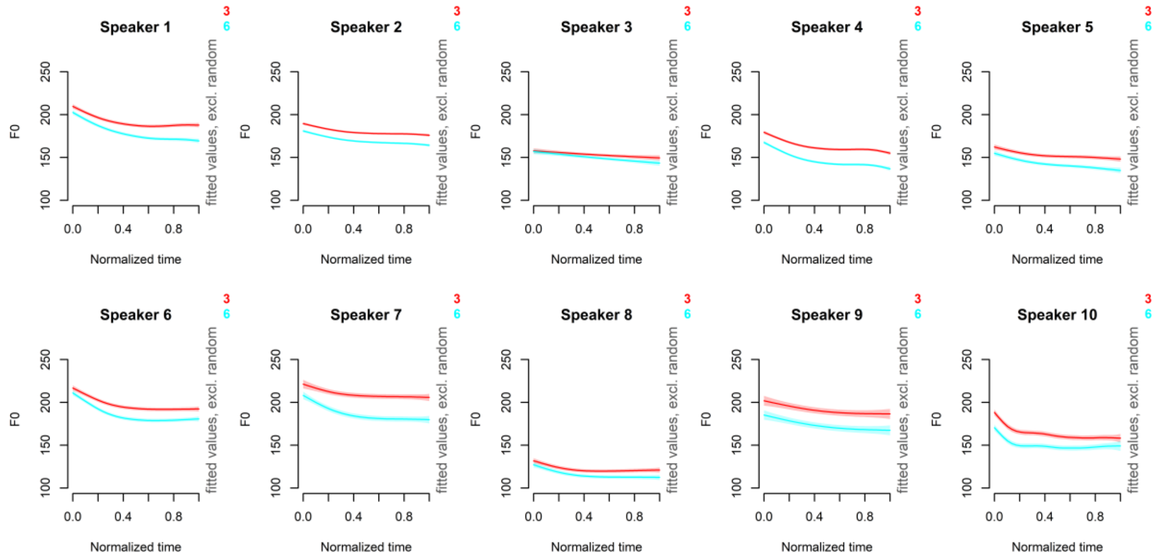


Figure 3.6 Non-linear smooths of T3 and T6 produced by each speaker in the older group

Table 3.6 shows the mean F0 difference between T3 and T6 (T3 - T6) at timepoint 6, where the two tones differ the most, for each of the nine speakers that appears to show higher T3 than T6. All nine speakers show that T3 is higher than T6 by around 10-25 Hz. The mean difference of the nine speakers is 15.5 Hz. Table 3.7 shows that only speaker 3 appears to confuse the two tones. For speaker 3, T3 is higher than T6 by 3.6 Hz at timepoint 6.

Table 3.6 Mean F0 difference between T3 and T6 (T3 - T6) at measurement point six for each speaker of the older group that appear to show higher T3 than T6

Speaker	1	2	4	5	6	7	8	9	10
Mean Diff	14.1	10.4	17.2	11.2	12.3	26.0	9.0	18.2	20.8
Mean Diff for the nine speakers = 15.5									

Table 3.7 Mean F0 difference between T3 and T6 (T3 - T6) at measurement point six for each speaker of the older group that appears to confuse T3 and T6

Speaker	3
Mean Diff	3.6

3.1.3 Older group, T4 and T6

Figure 3.1 and 3.2 included the non-linear smooths and difference smooths for T4 and T6 produced by the older group. The GAMM figures show that T6 is significantly higher than T4 from the beginning to the end, and the difference between them increases from the beginning to the second half of the tones. Table 3.8 shows the mean F0 difference between T4 and T6 (T6 - T4) for each F0 measurement point (2-8) for all speakers of the older group. The mean difference is the biggest at timepoint 6, with a value of 26 Hz. The mean difference tends to be bigger in the later portion of the tone curves. This is reasonable because T4 is a low-falling tone, and it is differentiated from T6 in both mean F0 and F0 change over the later portion (the sixth and seventh sections out of eight sections) of the vocalic segments (Khouw & Ciocca, 2007).

Table 3.8 Mean F0 difference between T4 and T6 (T6 - T4) for each F0 measurement point for the older group

Point	2	3	4	5	6	7	8
Mean Diff	15.1	19.5	20.8	22.9	26.4	26.4	24.5
At Point 6, Mean Diff is the biggest = 26.4							

Figure 3.7 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular. Many speakers (speaker 6, 7, 9, 10) sometimes creak when producing T4.

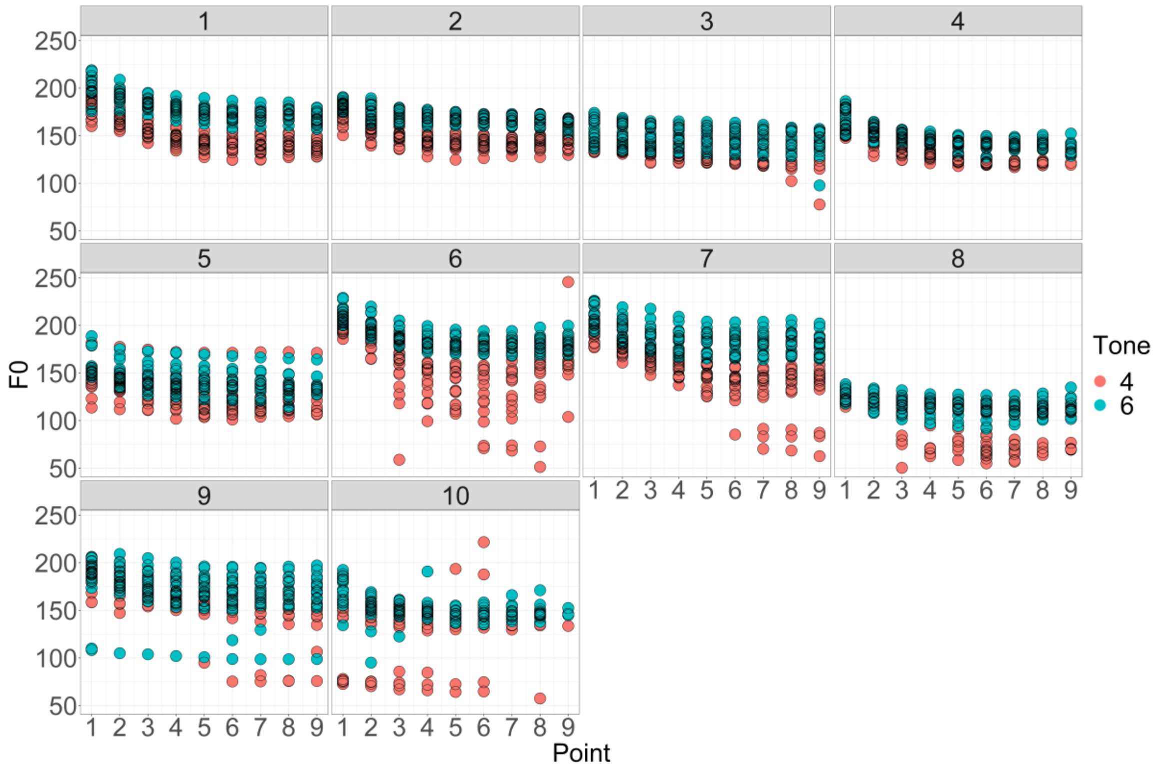


Figure 3.7 F0 data points for T4 and T6 produced by each speaker in the older group

Figure 3.8 shows the GAMM non-linear smooths of the two tones produced by each participant. All ten speakers appear to show higher T6 than T4. This finding is also supported by the GAMM difference smooths of the two tones produced by each participant, which are shown in Appendix B.

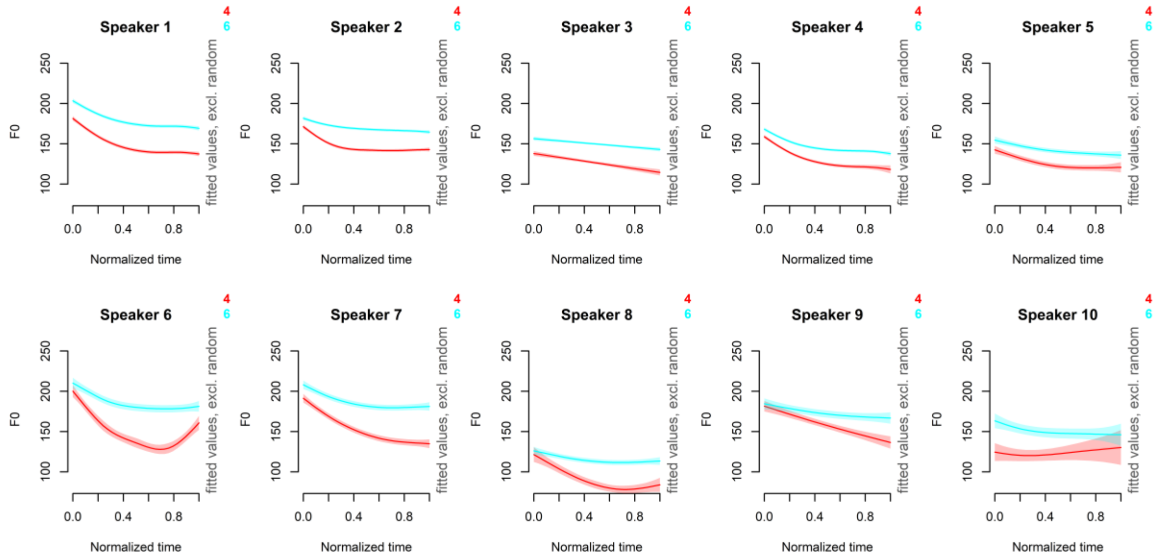


Figure 3.8 Non-linear smooths of T4 and T6 produced by each speaker in the older group

Table 3.9 shows the mean F0 difference between T4 and T6 (T6 - T4) at timepoint 6, where the two tones differ the most, for each speaker of the older group. All ten speakers show that T6 is higher than T4 by around 10-50 Hz. The mean difference of the ten speakers is 26.4 Hz.

Table 3.9 Mean F0 difference between T4 and T6 (T6 - T4) at measurement point six for each speaker of the older group that appear to show higher T6 than T4

Speaker	1	2	3	4	5	6	7	8	9	10
Mean Diff	32.3	25.8	24.4	19.4	20.3	49.6	41.5	36.1	15.1	10.5
Mean Diff for the ten speakers = 26.4										

3.1.4 Younger group, T2 and T5

Figure 3.1 and 3.2 included the non-linear smooths and difference smooths for T2 and T5 produced by the younger group. It can be seen from the GAMM figures that the T2 show a significantly higher F0 than T5 at the beginning and the later portion of the vocalic segment. The two tones are rising tones, and the difference between them is the biggest at the end of the vocalic segment. Table 3.10 shows the mean F0 difference between T2 and T5 (T2 - T5) for each F0 measurement point (2-8) for all speakers of the younger group. The mean difference is the biggest at timepoint 8, with a value of 8.2 Hz.

The mean difference tends to be bigger at the later portion especially the end of the tone curves, which aligns with the previous finding that T2 and T5 are acoustically separated based on F0 change over the later portion (the sixth and seventh sections out of eight sections) of the vocalic segment (Khouw & Ciocca, 2007).

Table 3.10 Mean F0 difference between T2 and T5 (T2 - T5) for each F0 measurement point for the younger group

Point	2	3	4	5	6	7	8
Mean Diff	4.1	3.1	2.9	3.7	4.5	5.2	8.2
At Point 8, Mean Diff is the biggest = 8.2							

Figure 3.9 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is less concentrated and regular than that of the older group. Speaker 16 creaks a lot in both T2 and T5.

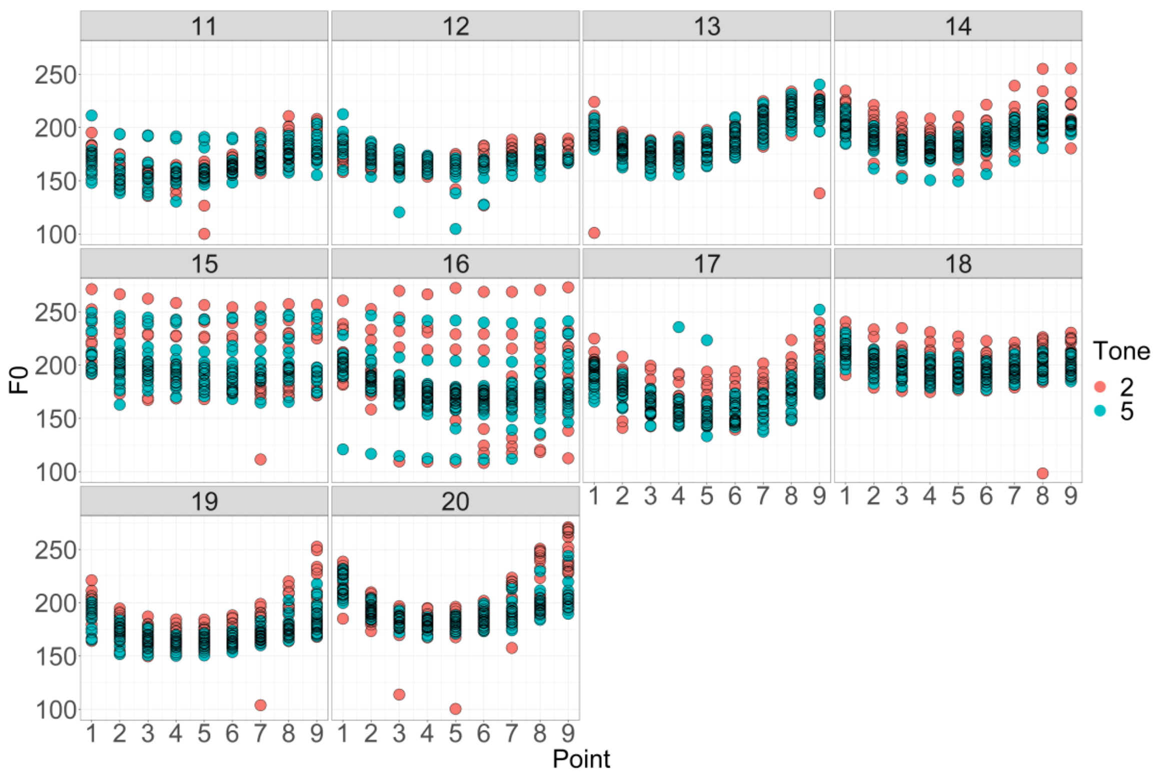


Figure 3.9 Mean F0 difference between T2 and T5 (T2 - T5) for each F0 measurement point for the younger group

Figure 3.10 shows the GAMM non-linear smooths of the two tones produced by each participant. Seven of the ten speakers (all except for speakers 13, 15, 16) appear to show higher T2 than T5. This finding is also supported by the GAMM difference smooths of the two tones produced by each participant, which are shown in Appendix B.

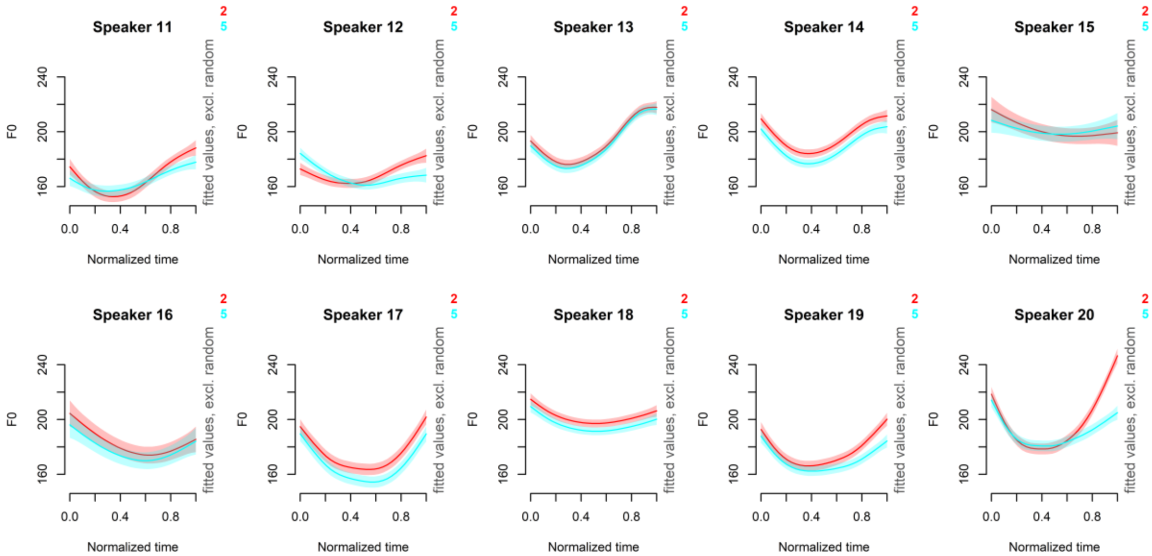


Figure 3.10 Non-linear smooths of T2 and T5 produced by each speaker in the younger group

Table 3.11 shows the mean F0 difference between T2 and T5 (T2 - T5) at timepoint 8, where the two tones differ the most, for each of the seven speakers that appear to show higher T2 than T5. All seven speakers expect for speakers 8 and 20 show that the F0 values of T2 are higher than T5 by about 10-20 Hz. The mean difference of the seven speakers is 11.4 Hz. Table 3.12 shows the results for the three speakers that appear to merge T2 and T5. For these three speakers, T2 is higher than T5 by 1.2 Hz on average.

Table 3.11 Mean F0 difference between T2 and T5 (T2 - T5) at measurement point eight for each speaker of the younger group that appears to show higher T2 than T5

Speaker	11	12	14	17	18	19	20
Mean Diff	7.6	10.8	9.1	15.2	1.3	12.4	23.1
Mean Diff for the seven speakers = 11.4							

Table 3.12 Mean F0 difference between T2 and T5 (T2 - T5) at measurement point eight for each speaker of the younger group that appears to confuse T2 and T5

Speaker	13	15	16
Mean Diff	3.1	0.8	-0.2
Mean Diff for the three speakers = 1.2			

3.1.5 Younger group, T3 and T6

Figure 3.1 and 3.2 included the non-linear smooths and difference smooths for T3 and T6 produced by the younger group. It can be seen from the GAMM figures that T3 shows significantly higher F0 than T6 from the beginning to the end. The two tones are level tones, and the difference between the two tones increases from the beginning to the middle of the tones, but then stays the same from the middle to the end of the tones. Table 3.13 shows the mean F0 difference between T3 and T6 (T3 - T6) for each F0 measurement point (2-8) for all speakers of the younger group. The mean difference is the biggest at timepoint 8, with a value of 3.6 Hz. The mean difference is similar through the whole vocalic segment, which aligns with the previous finding that T3 and T6 are acoustically separated based on mean F0 value (Khouw & Ciocca, 2007).

Table 3.13 Mean F0 difference between T3 and T6 (T3 - T6) for each F0 measurement point for the younger group

Point	1	2	3	4	5	6	7	8	9
Mean Diff	0.1	2.5	3.5	2.6	2.1	3.0	3.1	3.6	2.7
At Point 8, Mean Diff is the biggest = 3.6									

Figure 3.11 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is less concentrated and regular than that of the older group. Speakers 15 and 16 creak sometimes in both T3 and T6.

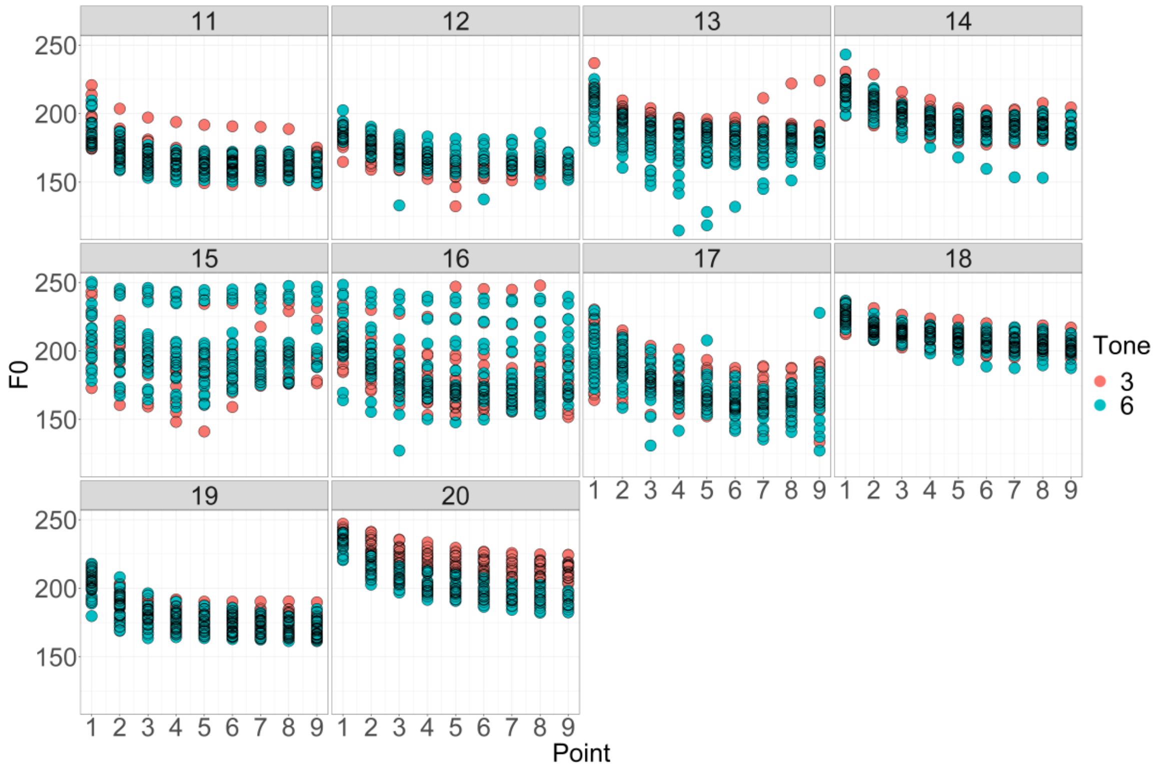


Figure 3.11 F0 data points for T3 and T6 produced by each speaker in the younger group

Figure 3.12 shows the GAMM non-linear smooths of the two tones produced by each participant. Five of the ten speakers (speakers 11, 13, 17, 19, 20) appear to show higher T3 than T6. It is noticeable that three speakers (speakers 12, 15, 16) show lower T3 than T6. These findings are also supported by the GAMM difference smooths of the two tones produced by each participant, which are shown in Appendix B.

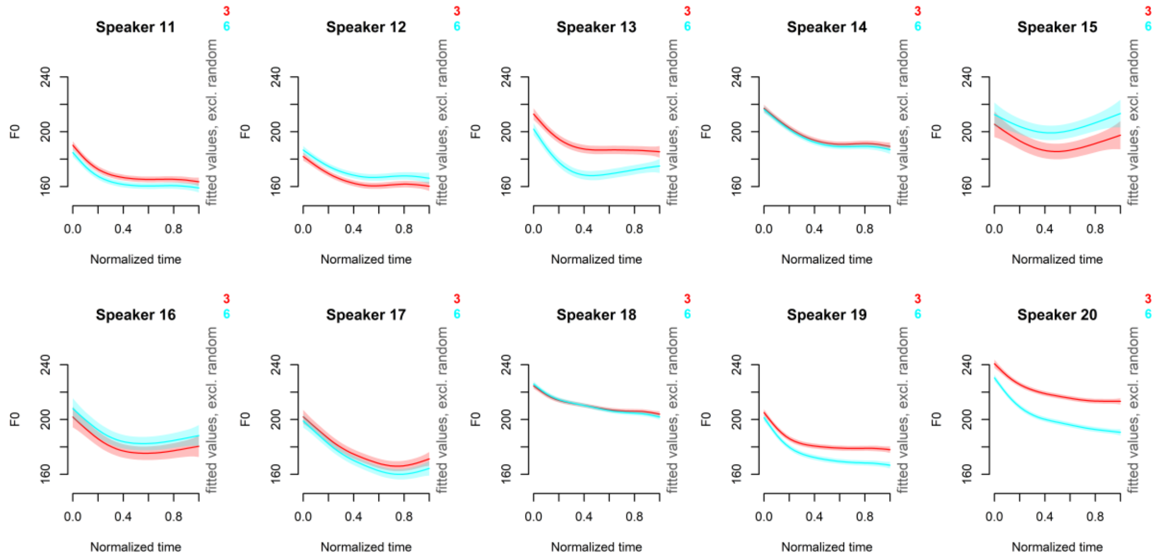


Figure 3.12 Non-linear smooths of T3 and T6 produced by each speaker in the younger group

Table 3.14 shows the mean F0 difference between T3 and T6 (T3 - T6) at timepoint 8, where the two tones differ the most, for each of the five speakers that appear to show higher T3 than T6. All five speakers except for show that the F0 values of T3 are higher than T6 by about 5-20 Hz. The mean difference of the five speakers is 11.7 Hz. Table 3.15 shows the results for the five speakers that appear to merge T3 and T6. three of them show lower T2 than T5 at timepoint 8. For these five speakers, T3 is lower than T6 by 4.2 Hz on average.

Table 3.14 Mean F0 difference between T3 and T6 (T3 - T6) at measurement point eight for each speaker of the younger group that appears to show higher T3 than T6

Speaker	11	13	17	19	20
Mean Diff	5.2	13.3	7.4	11.2	21.3
Mean Diff for the five speakers = 11.7					

Table 3.15 Mean F0 difference between T3 and T6 (T3 - T6) at measurement point eight for each speaker of the younger group that appears to merge T3 and T6

Speaker	12	14	15	16	18
Mean Diff	-6.0	3.1	-13.8	-5.3	1.1
Mean Diff for the five speakers = -4.2					

3.1.6 Younger group, T4 and T6

Figure 3.1 and 3.2 included the non-linear smooths and difference smooths for T4 and T6 produced by the younger group. It can be seen from the GAMM figures that the T6 show a significantly higher F0 than T4 from the beginning to the end. T6 is a level tone, and T4 is a low-falling tone. The difference between the two tones increases from the beginning to the end of the tones. Table 3.16 shows the mean F0 difference between T6 and T4 (T6 - T4) for each F0 measurement point (2-8) for all speakers of the younger group. The mean difference is the biggest at timepoint 7, with a value of 38.1 Hz. The mean difference tends to be bigger in the later portion of the tone curves. This is reasonable because T4 is a low-falling tone, and it is differentiated from T6 in both mean F0 and F0 change over the later portion (the sixth and seventh sections out of eight sections) of the vocalic segments (Khouw & Ciocca, 2007).

Table 3.16 Mean F0 difference between T4 and T6 (T6 - T4) for each F0 measurement point for the younger group

Point	1	2	3	4	5	6	7	8	9
Mean Diff	10.1	8.1	10.8	13.3	21.8	30.3	38.1	37.3	36.4
At Point 7, Mean Diff is the biggest = 38.1									

Figure 3.13 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is less concentrated and regular than that of the older group. It’s also less concentrated and regular than that of other tone-by-group combinations. This is also indicated by a relatively lower r^2 value (0.527, shown in Table 3.1) of the GAMM model output. The datapoints of T4 show that most of the young speakers creak a lot when producing T4.

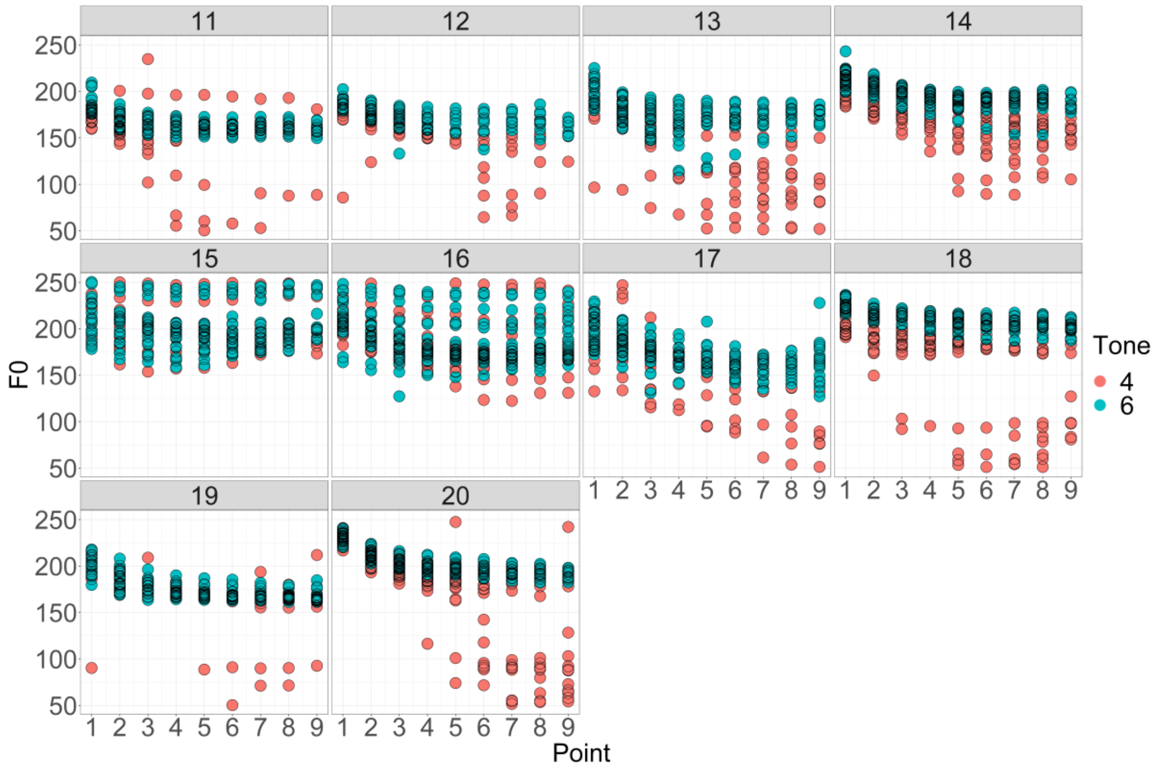


Figure 3.13 F0 data points for T4 and T6 produced by each speaker in the younger group

Figure 3.14 shows the GAMM non-linear smooths of the two tones produced by each participant. Seven speakers (all except for speakers 15, 16, 19) appear to show higher T6 than T4. Figure 3.13 implies that the reason why speaker 19 does not show significantly higher T6 than T4 is lack of valid datapoints due to creaky voice. This finding is also supported by the difference smooths of the two tones produced by each participant, which are shown in Appendix B.

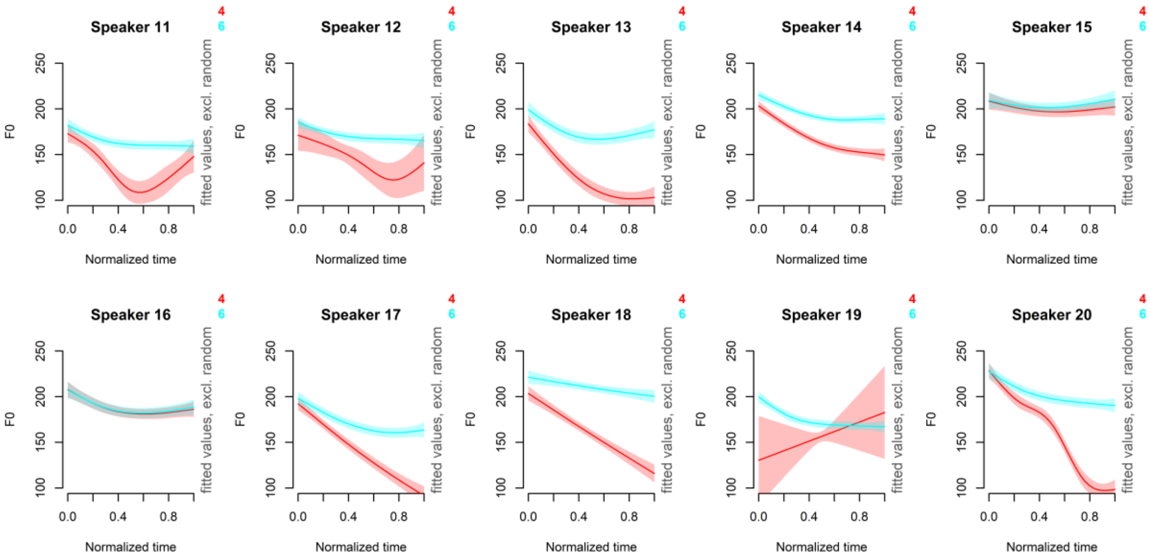


Figure 3.14 Non-linear smooths of T4 and T6 produced by each speaker in the younger group

Table 3.17 shows the mean F0 difference between T4 and T6 (T6 - T4) at timepoint 7, where the two tones differ the most, for the seven speakers that appear to show higher T6 than T4. All seven speakers show that the F0 values of T6 are higher than T4 by more than 35 Hz. The mean difference of the nine speakers is 60.1 Hz. Table 3.18 shows the results for the three speakers that appear to merge T4 and T6. For these three speakers, T6 is higher than T4 by 14.8 Hz on average.

Table 3.17 Mean F0 difference between T4 and T6 (T6 - T4) at measurement point seven for each speaker of the younger group that appears to show higher T6 than T4

Speaker	11	12	13	14	17	18	20
Mean Diff	49.3	54.1	73.7	37.3	43.8	74.7	87.8
Mean Diff for the seven speakers = 60.1							

Table 3.18 Mean F0 difference between T4 and T6 (T6 - T4) at measurement point seven for each speaker of the younger group that appears to merge T6 and T4

Speaker	15	16	19
Mean Diff	8.3	2.0	34.2
Mean Diff for the three speakers = 14.8			

3.1.7 Summary of analyses for the confusable tone pairs

Table 3.19 shows the GAMM estimates and *p* values for each tone pair by age group combination. Table 3.20 shows the mean F0 difference at the measurement point where the difference is the biggest for the speakers that appear to merge the tones. When comparing the older group to the younger group, table 3.19 shows that for both T2 vs T5 and T3 vs T6, the older group shows bigger GAMM estimates, which implies bigger differences between the confusable tone pairs than the younger group. This bigger difference is also shown in the speakers who appear to produce distinctive tones in each tone pair. Table 3.20 shows that for these speakers, the mean F0 differences at the measurement point where the difference is the biggest is greater in the older group than the younger group for both T2 vs T5 and T3 vs T6. These results suggest that at this time in history, the younger group shows a smaller F0 difference when producing the confusable tone pairs compared to the older group. What's more, Table 3.21 shows that for T3 vs T6, younger speakers who appear to show tone-merging even show slightly higher T6 than T3. This implies that tone merging is in progress, and the younger group is more advanced in the merging progress. However, for T4 vs T6, the younger group shows bigger mean F0 differences in the estimates in the GAMM outputs. For this tone pair, younger speakers who appear to produce distinctive tones also show bigger mean F0 than the older group at the measurement point where the difference is the biggest. Visual inspection of the F0 data distribution plots suggests that the bigger difference of the younger group is due to more creakiness when producing T4. It seems that the younger group creaks more than the older group when producing the very low T4. When considering the tone merging progress, these results suggest that generally, for both age groups, none of the three confusable tone pairs have fully merged.

Table 3.19 GAMM estimates and p values for each tone pair by age group combination

Group	Older			Younger		
Tone	T2 vs T5	T3 vs T6	T6 vs T4	T2 vs T5	T3 vs T6	T4 vs T6
GAMM Estimate	-12.462	-12.632	25.260	-4.888	-3.743	26.869
p	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table 3.20 Mean F0 difference at the measurement point where the difference is the biggest for the speakers that show distinctive tones

Group	Older			Younger		
Tone	T2 vs T5	T3 vs T6	T6 vs T4	T2 vs T5	T3 vs T6	T4 vs T6
Point	9	5	7	8	5	8
Num. of speakers	9	9	10	7	5	7
Mean F0 Diff	30.8	15.5	26.4	11.4	11.7	60.1

Table 3.21 Mean F0 difference at the measurement point where the difference is the biggest for the speakers appear to be tone-mergers

Group	Older			Younger		
Tone	T2 vs T5	T3 vs T6	T6 vs T4	T2 vs T5	T3 vs T6	T4 vs T6
Point	9	5	7	9	5	8
Num. of speakers	1	1	0	3	5	3
Mean F0 Diff	4.6	3.6	N/A	1.2	-4.2	14.8

However, individual analyses showed that some speakers in both age groups appear to merge some tone pairs. Table 3.22 shows the speakers that appear to merge the tone pair in each tone pair by age group combination. More speakers in the younger group appear to show merging of the confusable tone pairs than the older group in each tone pair by age group combination. Additionally, table 3.20 shows that even for the speakers who have not merged the tone pairs, younger speakers show a smaller F0 difference between T2 vs T5 and T3 vs T6 compared to the older group. We cannot see the smaller difference between T4 vs T6 in the younger group due to more creakiness in T4 produced by younger speakers. These results suggest that T2 vs T5 and T3 vs T6 are likely to be in a tone-merging process. Table 3.22 also shows that only a minority of speakers in both age groups appear to merge the tone pairs. This implies that Hong Kong Cantonese is at the beginning of the merging process. The younger group show more speakers who merge T3 and T6 compared to the other two tone pairs. For younger and older speakers who appear to show distinctive T3 and T6, the mean difference between T3 and T6 is smaller compared to the other two confusable tone pairs. These findings suggest that T3 vs T6 is more advanced in the merging process than the other two confusable tone pairs.

Table 3.22 Speakers that appear to merge the tone pair in each tone pair by age group combination

	Older (age 35-45)		Younger (age 18-28)	
	Speakers	Num.	Speakers	Num.
T2 vs T5	3	1 out of 10	13, 15, 16	3 out of 10
T3 vs T6	3	1 out of 10	12, 14, 15, 16, 18	5 out of 10
T6 vs T4	None	0 out of 10	15, 16, 19	3 out of 10

The results also show great inter-speaker variation. Some participants (speakers 3, 8, 12, 14, 15, 16, 18) tend to show a smaller F0 difference between the two tones of each of the three confusable tone pairs, while others (speakers 1, 2, 5, 6, 7, 9, 20) tend to produce all the three confusable tone pairs with substantial F0 differences. When looking at whether an individual tends to merge all the three confusable tone pairs at the same time, there is not an obvious within-speaker consistency. In other words, some participants merge all three confusable tone pairs at the same time, while others do not. For example, speaker 3, who appears merge T2 vs T5 and T3 vs T6, does not seem to merge T4 vs T6. Speaker 13, who appears to merge T2 vs T5 does not seem to merge T3 vs T6 or T4 vs T6. Speaker 12, 14, 18, who appear to merge T3 vs T6, do not seen to merge T2 vs T5 or T4 vs T6. Speaker 19, who appears to merge T4 vs T6, does not seem to merge T2 vs T5 or T3 vs T6. These examples do not support within-speaker consistency. However, speakers 15 and 16 appear to merge all three confusable tone pairs, which indicates within-speaker consistency to some extent.

3.2 The size of the IF0 effect of the Hong Kong Cantonese tones

In this section, GAMM models will be applied to the acoustic data to investigate the size of the IF0 effect for each tone by age group combination. In the overall group analysis, for the two levels of vowel height in each tone, GAMMs were fitted to the time-varying F0 data, to see whether the two levels of vowel height are significantly different or not. The best GAMM model, which was selected by the model comparison method suggested by Wieling (2018), includes a parametric term of vowel height, a smooth term of normalized time, a smooth term of a normalized-time-by-vowel-height interaction and random smooths of time-by-speaker-by-vowel-height. Auto-correction is included in the models.

Results of the GAMMs are shown in Table 3.23. High vowels are the baselines. Only speakers of the younger group show that T2 carried by high vowels is significantly ($p < 0.05$) higher in F0 than mid vowels. Similarly, speakers of the older group show that T2 carried by high vowels is nearly significantly ($p = 0.067$) higher in F0 than mid vowels. The other tone by age group combinations show no a significant difference between the two levels of vowel height. These findings suggest that there is no significant IF0 effect in any of the tone by age group combinations, except for T2 produced by the younger group. However, the estimates of all the tone by age group combinations, except for T6 produced by the younger group, are negative. This implies that for most tone by age combinations, tones carried by high vowels appear to show higher F0 than those carried by mid vowels, although the difference is statistically non-significant in most cases. Therefore, the IF0 effect may actually exist in most tone by age group combinations. The estimates also show that older speakers appear to show bigger F0 difference between the two levels of vowel heights compared to the younger group, which suggests that the IF0 effect may be generally smaller for the younger group. Additionally, the estimates show that for both the younger group and the older group, T3 and T6, which are level tones, appear to show a smaller difference between the two levels of vowel heights compared to the other four tones. Contrastively, T2, which is a contour tone, appears to show the biggest difference between the two levels of vowel heights among all of the six tones. This implies that level tones may show smaller IF0 effects than contour tones.

GAMM smooths of the older group and the younger group respectively are plotted in Figure 3.15 and Figure 3.16. Each plot in Figure 3.15 and Figure 3.16 contains a pink ribbon and a blue ribbon, which represents the F0 curves of the high vowel and the low vowel of each tone. GAMM difference smooths of the older group and the younger group respectively are plotted in Figure 3.17 and Figure 3.18. Each plot in Figure 3.17 and Figure 3.18 contains a grey ribbon, which shows the size of the F0 difference between the same tone on the high vowel and on the low vowel. The interval between two red vertical lines with a red horizontal line on the x axis shows the time range where the same tones on the two levels of vowel height are significantly different. In line with the GAMM estimates, comparison of Figure 3.15 and 3.16, as well as comparison of Figure 3.17 and 3.18 shows that the older group generally show a bigger difference between the two levels of vowel heights, with high vowels showing higher in F0 than mid vowels, compared to the younger group. Figure 3.17 and 3.18 also show that for T2, older speakers' high vowels show significantly higher F0 at the later portion of vocalic segments; younger speakers' high vowels show significantly higher F0 at both the middle and later portion of vocalic segments. Since GAMM outputs also show a significant difference s for T2 produced by the younger group and near-significant difference (p = 0.067) for T2 produced by the older group, the results suggest that there is a substantial IF0 effect for T2 produced by both age groups, especially in the later portion of the vocalic segment. For the other tone by age combinations except for T6 produced by the younger group, the figures show that high vowels show slightly higher F0 than the mid counterparts, which aligns with the negative GAMM estimates. However, this tendency is not confirmed by statistical significance. To investigate the IF0 effect in more detail, inter-speaker variation needs to be analyzed. In the following sections, inter-speaker variation for each tone by age group combination will be discussed.

Table 3.23 GAMM outputs for each tone by age group combination

Older (age 35-45)					
	Estimate	SE	t	p	r^2
T1	-5.263	16.407	-0.321	0.748	0.856
T2	-19.400	10.598	-1.830	0.067	0.873
T3	-3.333	11.929	-0.279	0.780	0.914
T4	-11.248	10.045	-1.120	0.263	0.684
T5	-4.854	9.277	-0.523	0.601	0.851
T6	-3.398	10.531	-0.323	0.747	0.848
Younger (age 18-28)					
	Estimate	SE	t	p	r^2
T1	-2.001	11.900	-0.168	0.867	0.730
T2	-15.786	6.275	-2.516	0.012	0.572
T3	-0.324	7.945	-0.041	0.967	0.756
T4	-1.976	11.175	-0177	0.860	0.556
T5	-4.398	6.189	-0.711	0.477	0.571
T6	2.980	7.598	0.392	0.695	0.652

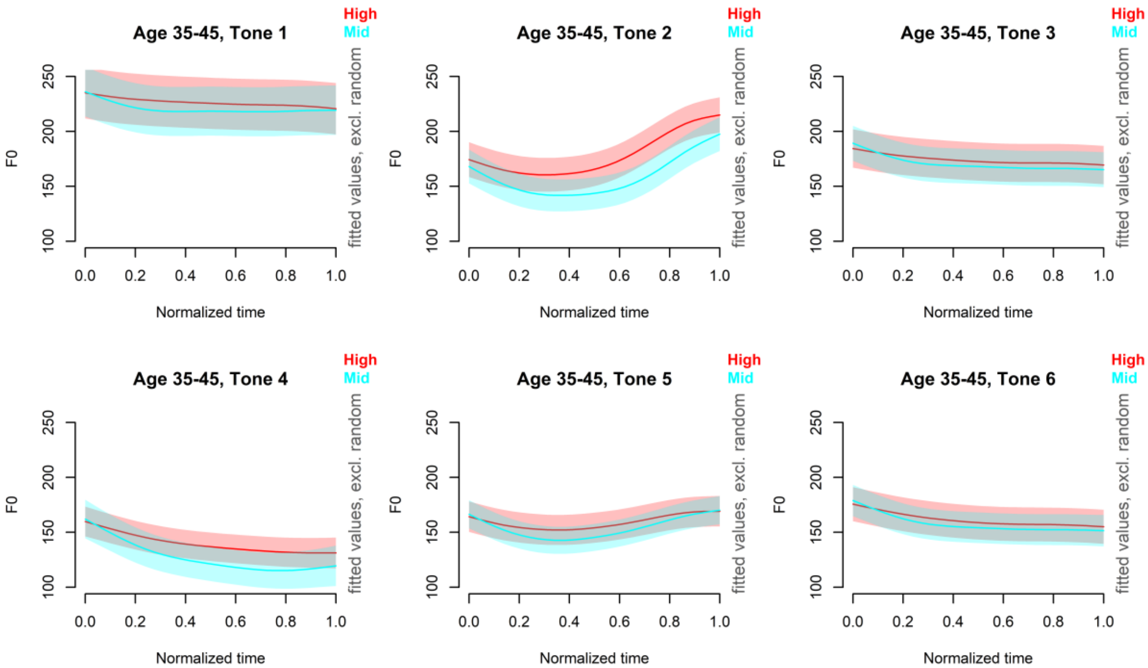


Figure 3.15 **Non-linear smooths for each tone produced by speakers of the older group**

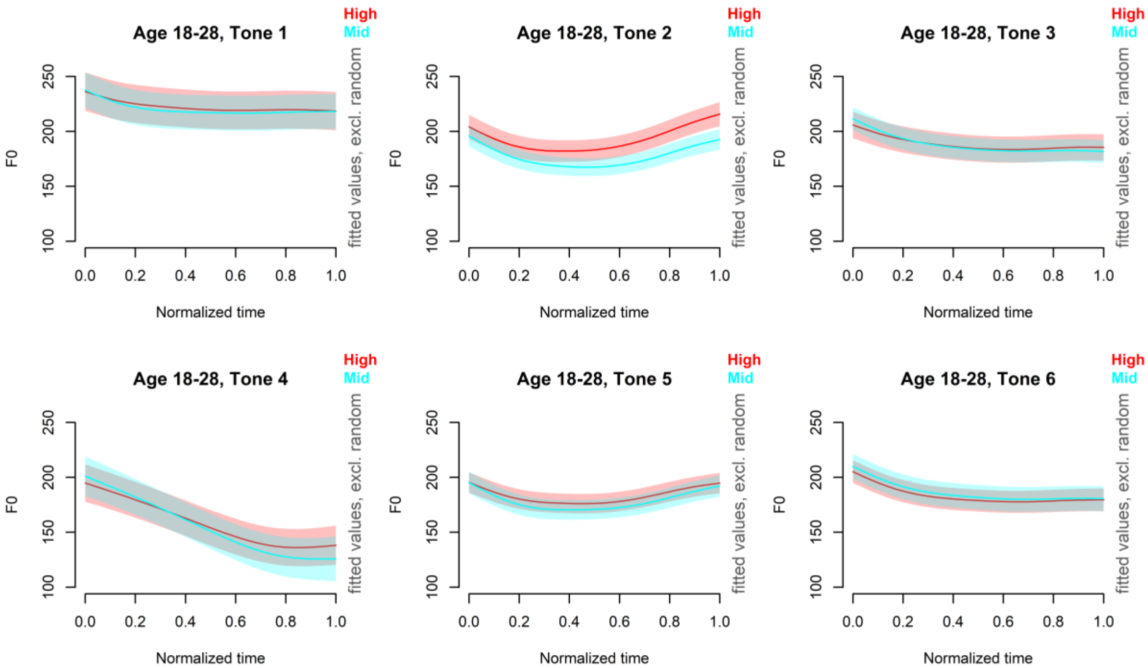


Figure 3.16 **Non-linear smooths for each tone produced by speakers of the younger group**

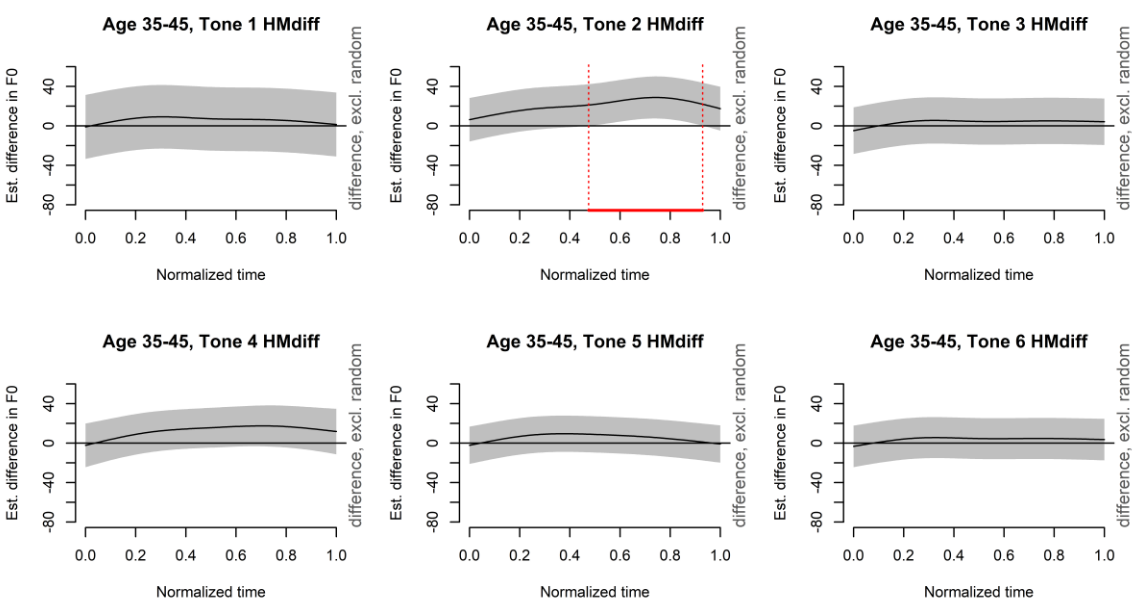


Figure 3.17 **Difference smooths for each tone produced by speakers of the older group**

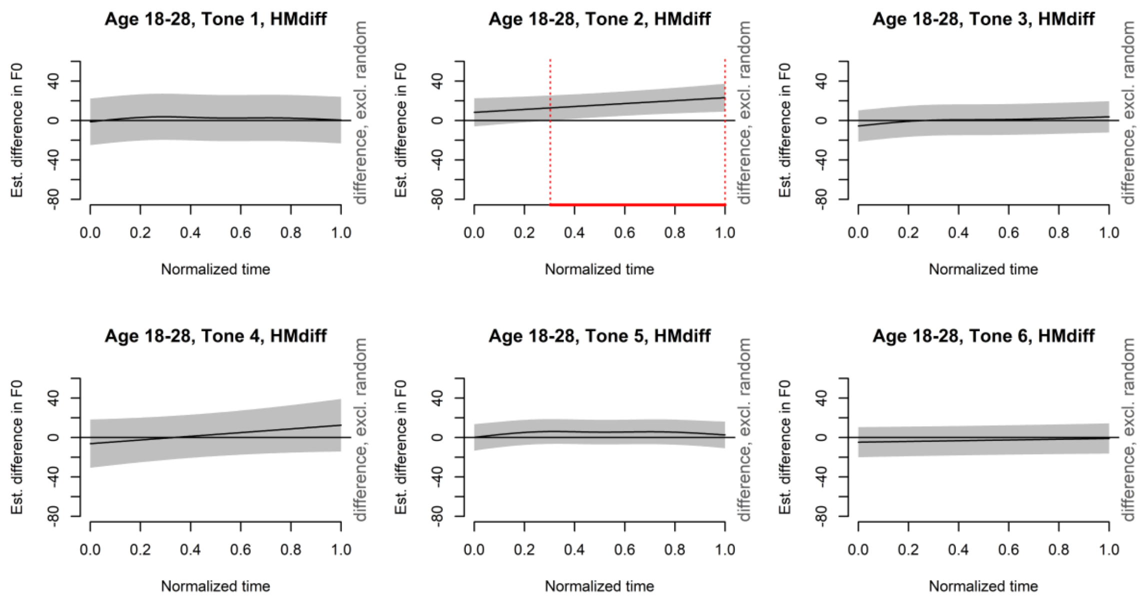


Figure 3.18 Difference smooths for each tone produced by speakers of the younger group

3.2.1 Older group, T1

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T1 produced by the older group. GAMM does not show a significant difference between the two levels of vowel heights carrying T1. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment. Table 3.24 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the older group. The mean differences are positive at all timepoints. The biggest mean difference value is 10.0 Hz, which occurs at timepoint 5.

Table 3.24 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T1 produced by the older group

Point	2	3	4	5	6	7	8
Mean Diff	6.4	9.3	9.1	10.0	8.3	7.5	6.7
At point 5, Mean Diff is the biggest = 10.0							

Figure 3.19 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular.

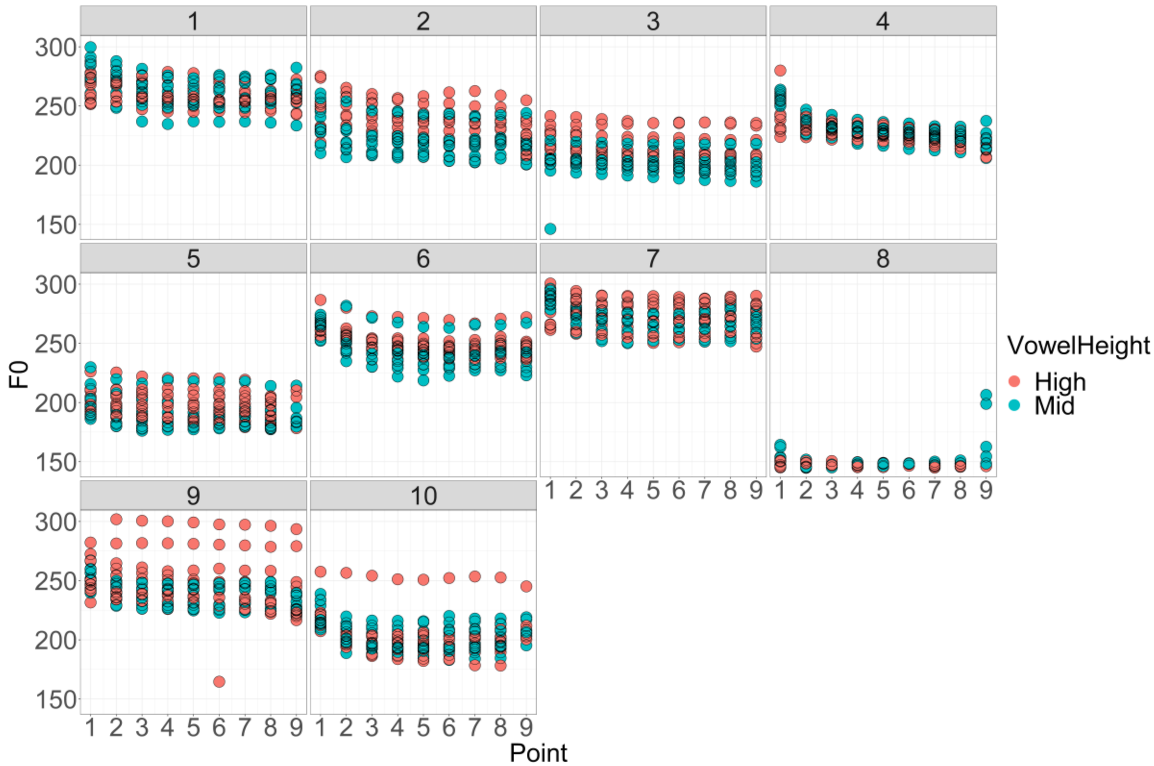


Figure 3.19 F0 data points for high vowels and mid vowels carrying T1 produced by each speaker in the older group

Figure 3.20 shows the GAMM non-linear smooths of the high and mid vowels carrying T1 produced by each participant. For six speakers (speakers 2, 3, 5, 6, 7, 9), high vowels carrying T1 appear to show higher F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

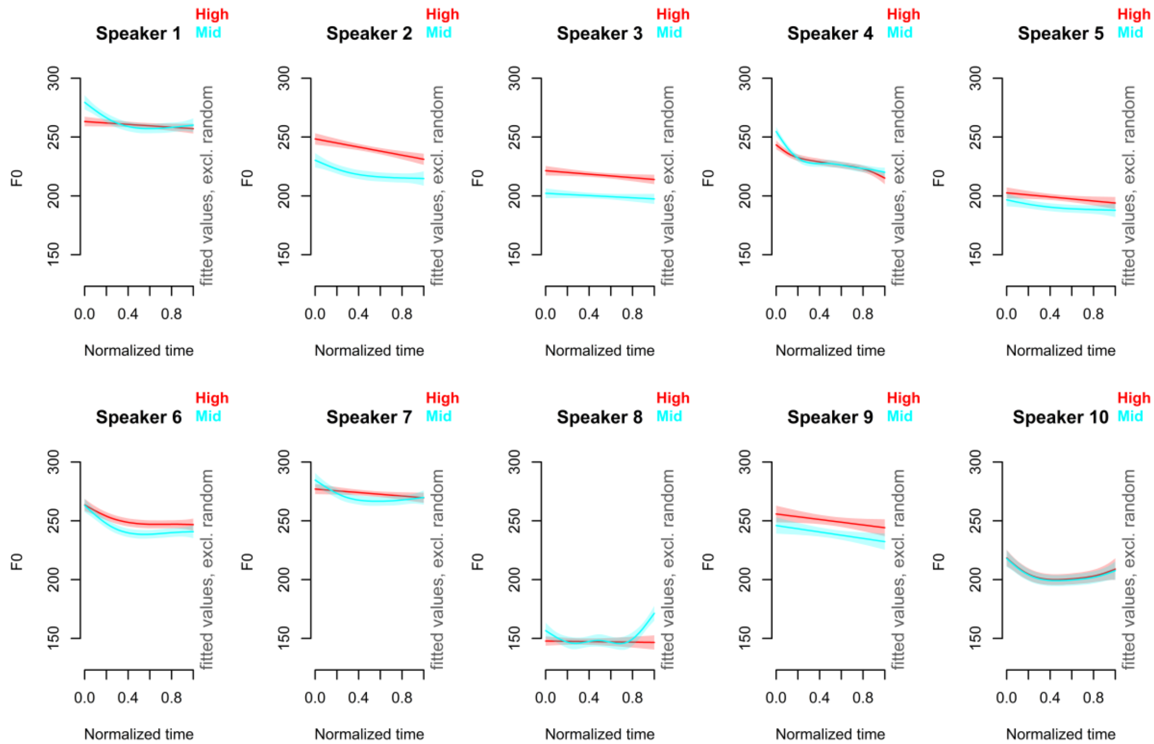


Figure 3.20 Non-linear smooths of high vowels and mid vowels carrying T1 produced by each speaker in the older group

Table 3.25 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T1 at timepoint 5, where high and mid vowels differ the most, for each of the six speakers whose high vowels appear to be higher in F0 than mid vowels. All six speakers show that the F0 values of high vowels are higher than mid vowels by about 5-20 Hz. The mean difference of the nine speakers is 12.4 Hz. Table 3.26 shows that for the four speakers who appear to show no IF0, the differences are all near zero, and the mean difference is 0.1 Hz.

Table 3.25 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T1 at measurement point five for each speaker of the older group who, on visual inspection, appears to show IF0 effect

Speaker	2	3	5	6	7	9
Mean Diff	20.8	16.7	9.4	9.6	4.9	13.0
Mean Diff for the six speakers = 12.4						

Table 3.26 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T1 at measurement point five for each speaker of the older group who, on visual inspection, appears to show no IF0 effect

Speaker	1	4	8	10
Mean Diff	0.8	-0.1	-0.9	0.6
Mean Diff for the four speakers = 0.1				

3.2.2 Older group, T2

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T2 produced by the older group. GAMM shows a significant difference between the two levels of vowel heights carrying T2 in the later portion of the vocalic segment. It can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment, and the difference is significant in the later portion of the vocalic segments. Table 3.27 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the older group. The mean differences are positive at all timepoints. The biggest mean difference value is 29.7 Hz, which occurs at timepoint 8. The mean difference is greater in the later portion of the vocalic segments.

Table 3.27 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T2 produced by the older group

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-1.9	8.6	13.0	14.2	18.1	24.0	29.7	26.0	3.4
At Point 7, Mean Diff is the biggest = 29.7									

Figure 3.21 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular.

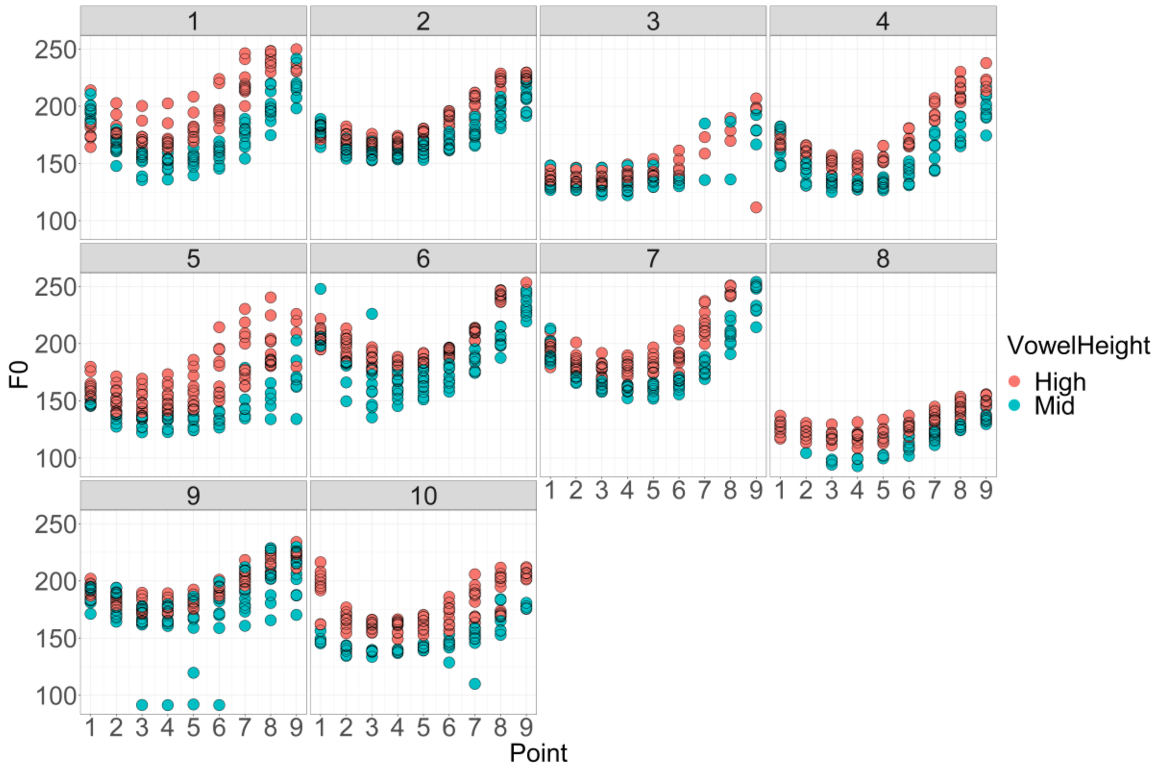


Figure 3.21 F0 data points for high vowels and mid vowels carrying T2 produced by each speaker in the older group

Figure 3.22 shows the GAMM non-linear smooths of the high and mid vowels carrying T2 produced by each participant. For all ten speakers, high vowels carrying T2 appear to show higher F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

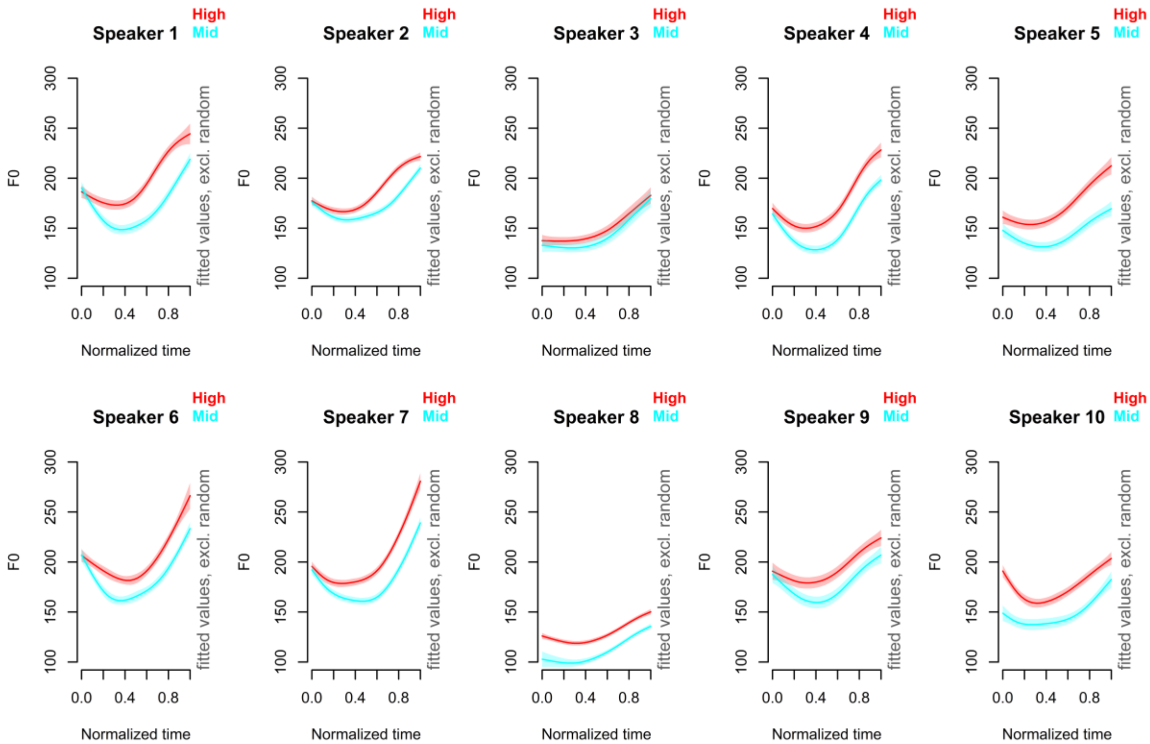


Figure 3.22 Non-linear smooths of high vowels and mid vowels carrying T2 produced by each speaker in the older group

Table 3.28 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T2 at timepoint 7, where high and mid vowels differ the most, for each speaker of the older group. All show positive values, which indicate higher F0 in high vowels compared to mid vowels. For all except for speaker 3, high vowels carrying T2 are higher than their mid counterparts by around 15-45 Hz. The mean difference across the ten speakers is 29.7 Hz.

Table 3.28 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T2 at measurement point seven for each speaker of the older group

Speaker	1	2	3	4	5	6	7	8	9	10
Mean Diff	45.7	27.9	5.6	37.5	42.1	24.7	35.0	15.7	16.4	34.5
Mean Diff for all the 10 speakers = 29.7										

3.2.3 Older group, T3

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T3 produced by the older group. GAMM does not show a significant difference between the two levels of vowel heights carrying T3. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive from around timepoint 2 to the end of the vocalic segment. Table 3.29 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the older group. The mean differences are positive at all timepoints, and they are very small throughout the vocalic segment. The biggest mean difference value is 5.5 Hz, which occurs at timepoint 8. The mean difference is similar throughout the middle of the vocalic segment.

Table 3.29 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T3 produced by the older group

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-1.0	1.9	2.2	4.2	4.1	4.8	5.0	5.5	4.2
At Point 8, Mean Diff is the biggest = 5.5									

Figure 3.23 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular.

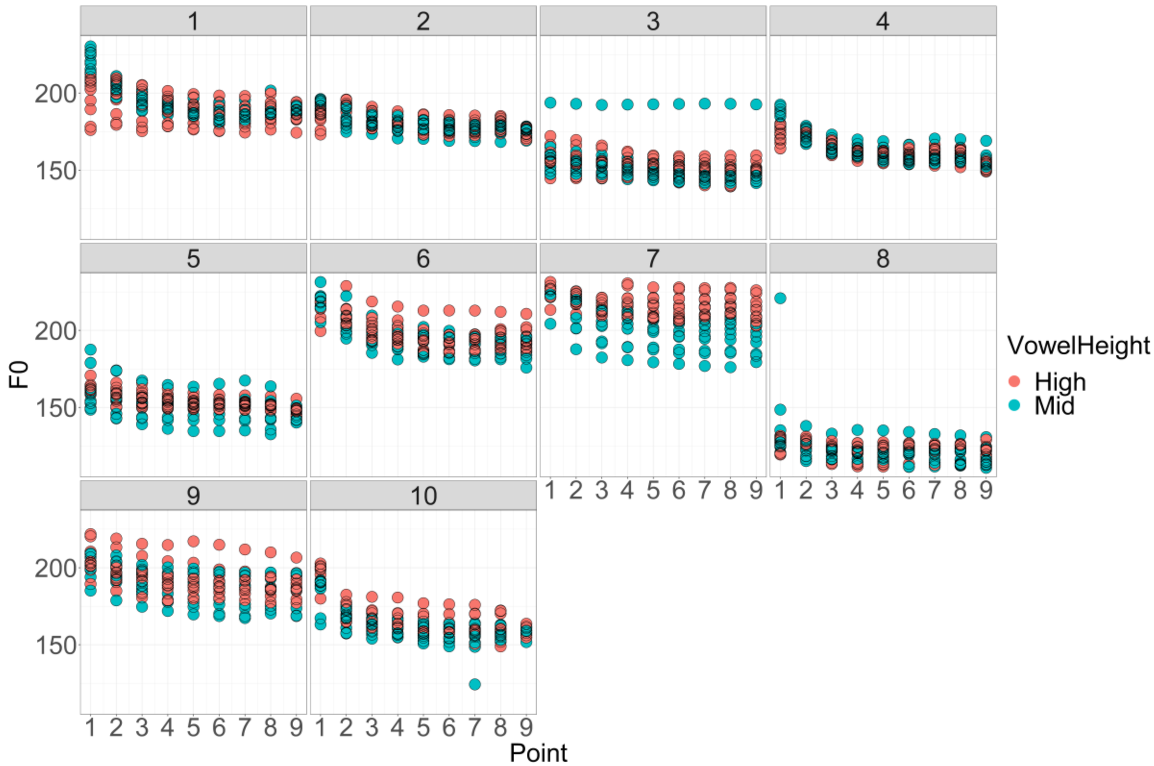


Figure 3.23 F0 data points for high vowels and mid vowels carrying T3 produced by each speaker in the older group

Figure 3.24 shows the GAMM non-linear smooths of the high and mid vowels carrying T3 produced by each participant. For six speakers (speakers 2, 5, 6, 7, 9, 10), high vowels carrying T3 appear to show higher F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

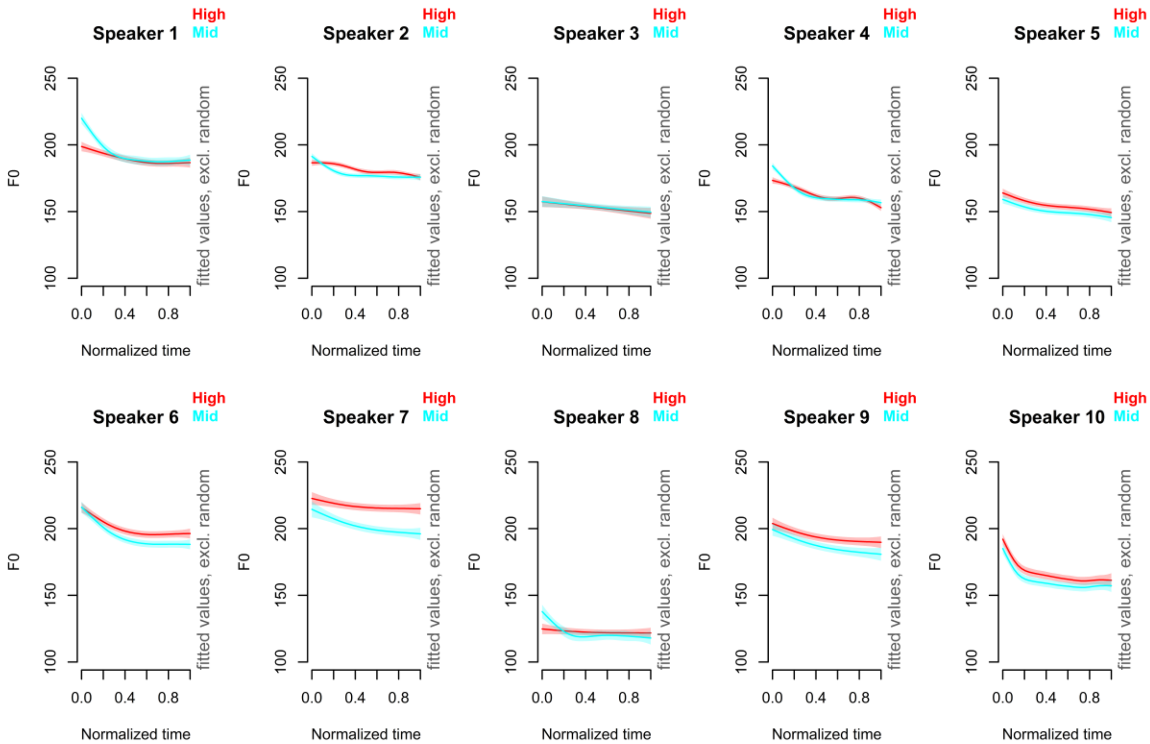


Figure 3.24 Non-linear smooths of high vowels and mid vowels carrying T3 produced by each speaker in the older group

Table 3.30 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T3 at timepoint 8, where high and mid vowels differ the most, for each of the six speakers whose high vowels appear to be higher in F0 than mid vowels. All six speakers show that the F0 values of high vowels are higher than mid vowels by about 3-20 Hz. The mean difference of the six speakers is 7.9 Hz. Table 3.31 shows that for the four speakers who appear to show no IF0, the mean difference between high and mid vowels (high – mid) is 0.7 Hz.

Table 3.30 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T3 at measurement point eight for each speaker of the older group who, on visual inspection, appears to show IF0 effect

Speaker	2	5	6	7	9	10
Mean Diff	3.6	4.4	9.1	18.7	7.6	4.0
Mean Diff for the six speakers = 7.9						

Table 3.31 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T3 at measurement point eight for each speaker of the older group who, on visual inspection, appears to show no IF0 effect

Speaker	1	3	4	8
Mean Diff	-1.0	-1.2	0.4	4.7
Mean Diff for the four speakers =0.7				

3.2.4 Older group, T4

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T4 produced by the older group. GAMM does not show a significant difference between the two levels of vowel heights carrying T4. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment. Table 3.32 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the older group. The mean differences are positive at all timepoints. The mean difference is the biggest at timepoint 6, with a value of 17.6 Hz. The mean difference is similar from timepoint 3 to the end of the vocalic segment.

Table 3.32 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T4 produced by the older group

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-8.6	0.5	12.0	8.5	10.5	17.6	12.6	12.4	6.7
At Point 6, Mean Diff is the biggest = 17.6									

Figure 3.25 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is less concentrated and regular than that of T1, T2 and T3. For speaker 10, many data points at the later portion of the vocalic segment are missing. Speakers 6, 7, 8, 10 sometimes creak when producing T4. Speaker 6 only creaks when producing mid vowels carrying T4. For the speakers who show creaky voice, they creak more in mid vowels than high vowels.

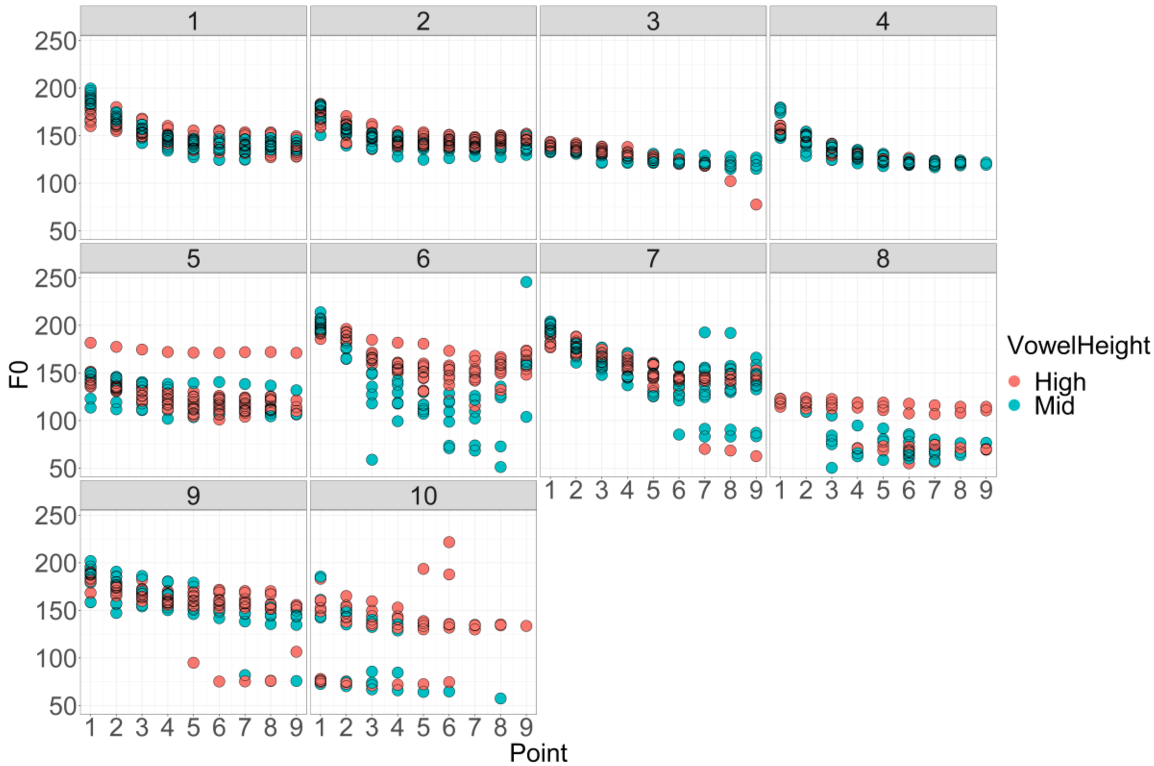


Figure 3.25 F0 data points for high vowels and mid vowels carrying T4 produced by each speaker in the older group

Figure 3.26 shows the GAMM non-linear smooths of the high and mid vowels carrying T4 produced by each participant. For seven speakers (speakers 1, 2, 5, 6, 8, 9, 10), high vowels carrying T4 appear to show higher F0 than mid vowels throughout most of the vocal segment, or in the later portion of the vocalic segment. For speaker 3, high vowels carrying T4 appear to show lower F0 than mid vowels in the later portion of the vocalic segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

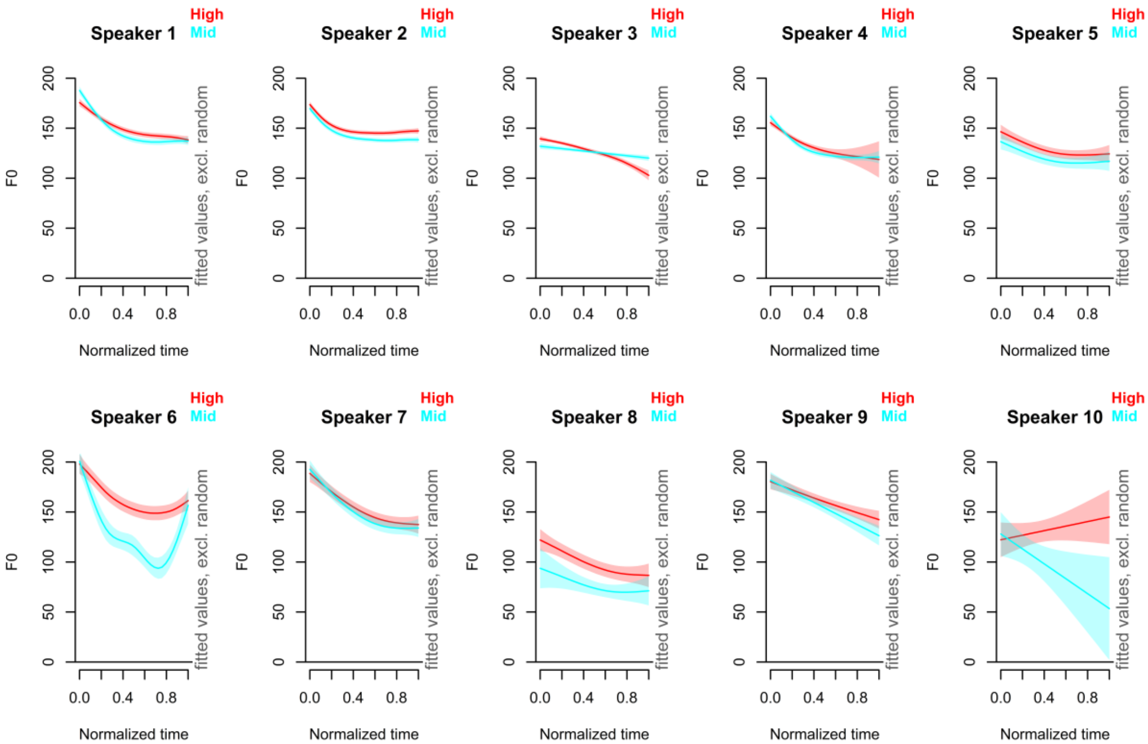


Figure 3.26 Non-linear smooths of high vowels and mid vowels carrying T4 produced by each speaker in the older group

Table 3.33 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T4 at timepoint 6, where high and mid vowels differ the most, for each of the seven speakers whose high vowels appear to be higher in F0 than mid vowels. Among the seven speakers, four of them (speakers 1, 2, 5, 9) show smaller differences of less than 10 Hz, while the other three (speakers 6, 8, 10) show bigger differences of 13-83 Hz. However, according to figure 3.25, speaker 10 has very few data points, so the result is not very accurate. The three speakers that show bigger IF0 show more creaky voice than the other speakers, according to figure 3.25. The mean difference of the seven speakers is 24.4 Hz. Table 3.34 shows that for the three speakers who appear to show no difference between high and mid vowels, the mean difference (high - mid) is 4.0 Hz.

Table 3.33 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T4 at measurement point six for each speaker of the older group who, on visual inspection, appears to show IF0 effect

Speaker	1	2	5	6	8	9	10
Mean Diff	7.1	7.6	6.9	49.0	13.5	3.6	83.0
Mean Diff for the seven speakers = 24.4							

Table 3.34 **Mean F0 difference between high vowels and mid vowels (high – mid) carrying T4 at measurement point six for each speaker of the older group who, on visual inspection, appears to show no IF0 effect**

Speaker	3	4	7
Mean Diff	-2.5	4.4	10.2
Mean Diff for the three speakers = 4.0			

3.2.5 Older group, T5

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T5 produced by the older group. GAMM does not show a significant difference between the two levels of vowel heights carrying T5. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment. Table 3.24 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the older group. The mean differences are positive at all timepoints. The biggest mean difference occurs at timepoint 6, with a value of 8.6 Hz. The mean difference is similar throughout the middle of the vocalic segment.

Table 3.35 **Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T5 produced by the older group**

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-5.4	5.3	7.0	7.6	8.2	8.6	5.8	4.1	-0.2
At Point 6, Mean Diff is the biggest = 8.6									

Figure 3.27 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular.

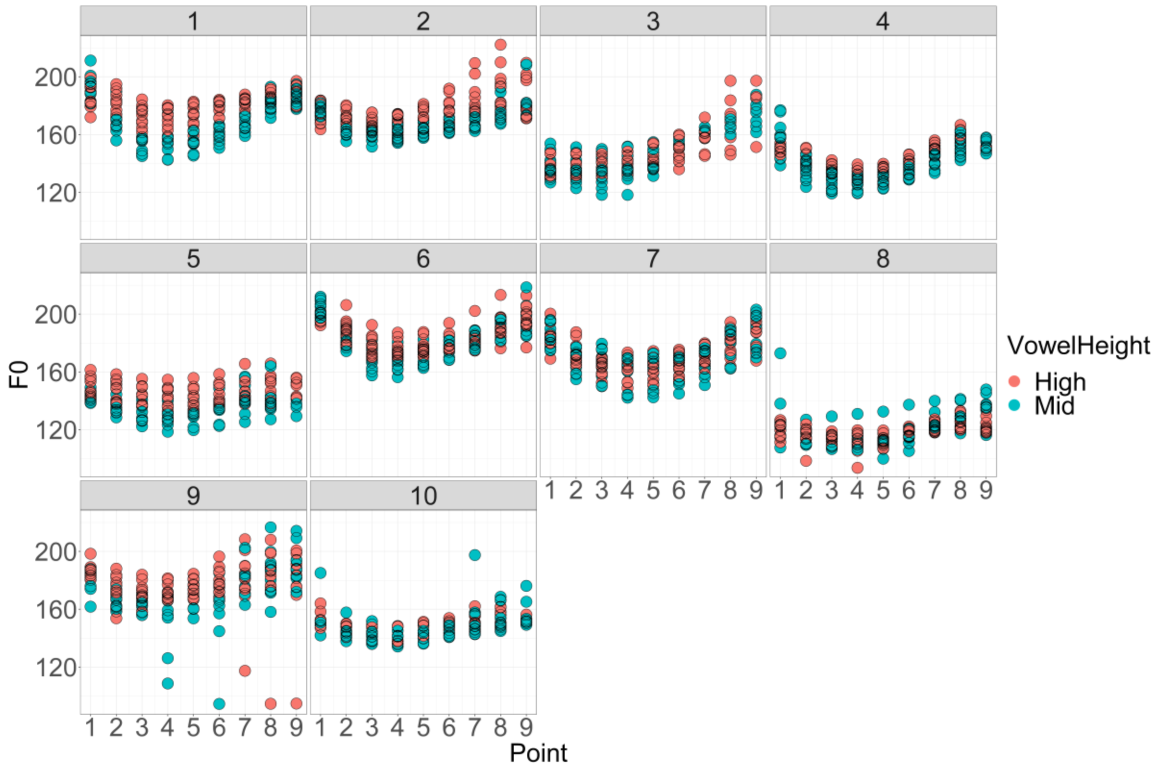


Figure 3.27 Non-linear smooths of high vowels and mid vowels carrying T5 produced by each speaker in the older group

Figure 3.28 shows the GAMM non-linear smooths of the high and mid vowels carrying T5 produced by each participant. For six speakers (speakers 1, 2, 4, 5, 6, 9), high vowels carrying T5 appear to show higher F0 than mid vowels throughout most of the vocal segment. For speaker 3 and speaker 8, high vowels carrying T5 appear to show lower F0 than mid vowels throughout most of the vocal segment This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

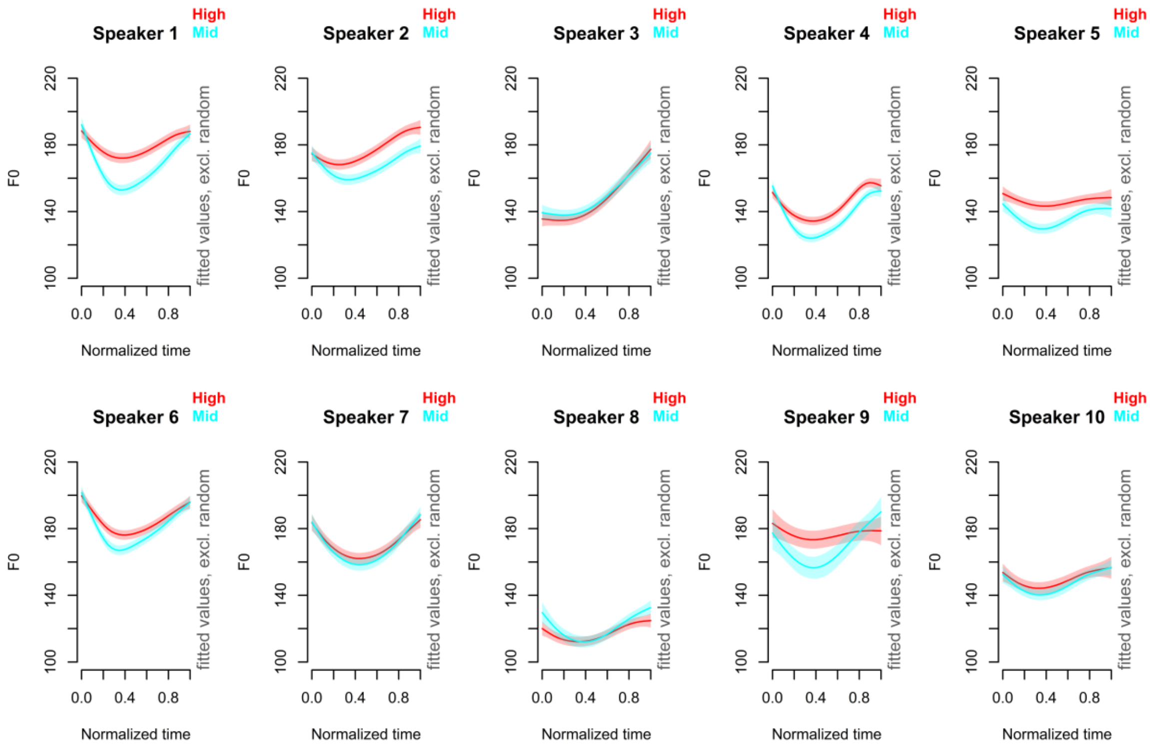


Figure 3.28 Non-linear smooths of high vowels and mid vowels carrying T5 produced by each speaker in the older group

Table 3.36 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T5 at timepoint 6, where high and mid vowels differ the most, for each of the six speakers whose high vowels appear to be higher in F0 than mid vowels. All of the six speakers show that the F0 values of high vowels are higher than mid vowels by about 5-20 Hz. The mean difference of the six speakers is 12.4 Hz. Table 3.37 shows the four speakers who show no IF0 all show differences of less than 5 Hz, and the mean difference is -0.2 Hz.

Table 3.36 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T5 at measurement point six for each speaker of the older group who, on visual inspection, appears to show IF0 effect

Speaker	1	2	4	5	6	9
Mean Diff	16.8	12.4	8.9	10.0	4.6	22.2
Mean Diff for the six speakers = 12.5						

Table 3.37 **Mean F0 difference between high vowels and mid vowels (high – mid) carrying T5 at measurement point six for each speaker of the older group who, on visual inspection, appears to show no IF0 effect**

Speaker	3	7	8	10
Mean Diff	-9.8	3.2	0.8	4.9
Mean Diff for the four speakers = -0.2				

3.2.6 Older group, T6

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T6 produced by the older group. GAMM does not show a significant difference between the two levels of vowel heights carrying T6. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment. Table 3.38 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the older group. The mean differences are positive at all timepoints, and they are very small (< 5 Hz) throughout the vocalic segment, which appears to be undetectable for humans. The biggest mean difference value is 4.7 Hz, which occurs at timepoint 6. The mean difference is similar throughout the middle of the vocalic segment.

Table 3.38 **Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T6 produced by the older group**

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-5.4	3.2	3.7	3.6	3.8	4.7	4.1	4.5	3.2
At Point 6, Mean Diff is the biggest = 4.7									

Figure 3.29 shows the F0 data points for each of the ten speakers in the older group. Datapoint dispersion is relatively concentrated and regular.

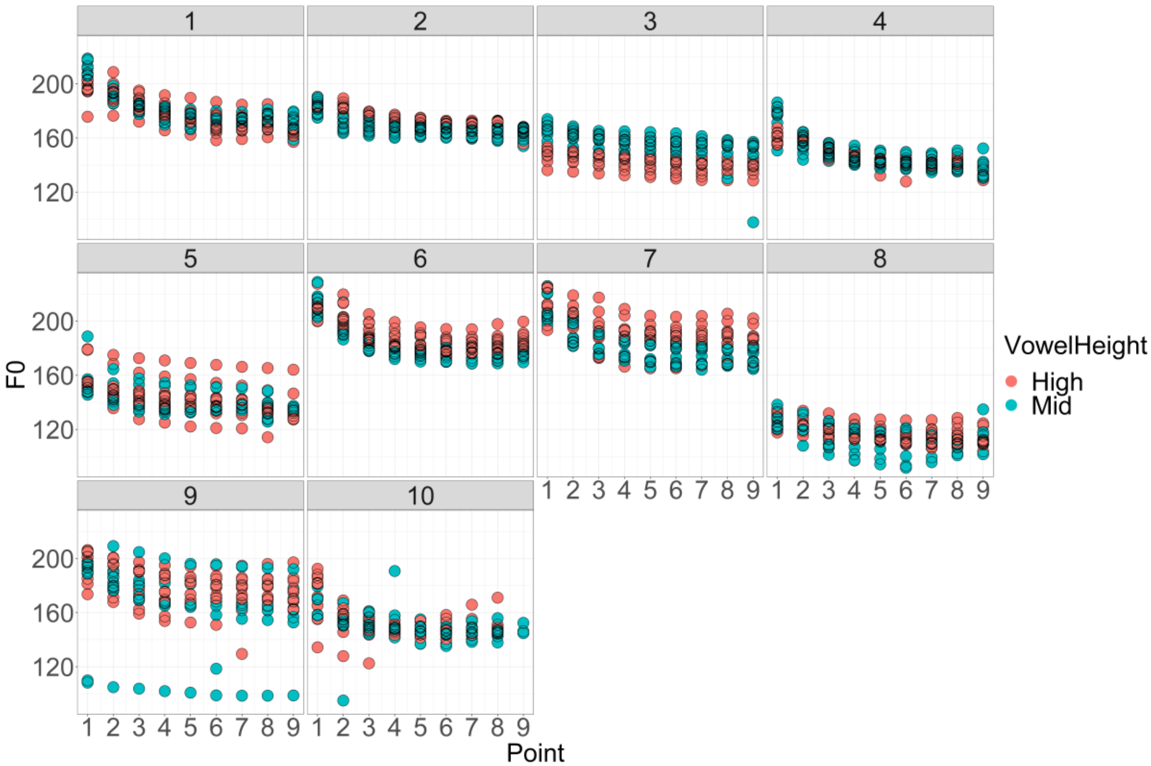


Figure 3.29 F0 data points for high vowels and mid vowels carrying T6 produced by each speaker in the older group

Figure 3.30 shows the GAMM non-linear smooths of the high and mid vowels carrying T6 produced by each participant. For five speakers (speakers 2, 6, 7, 8, 9), high vowels carrying T6 appear to show higher F0 than mid vowels throughout most of the vocal segment. For speaker 3, high vowels carrying T6 appear to show lower F0 than mid vowels throughout the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

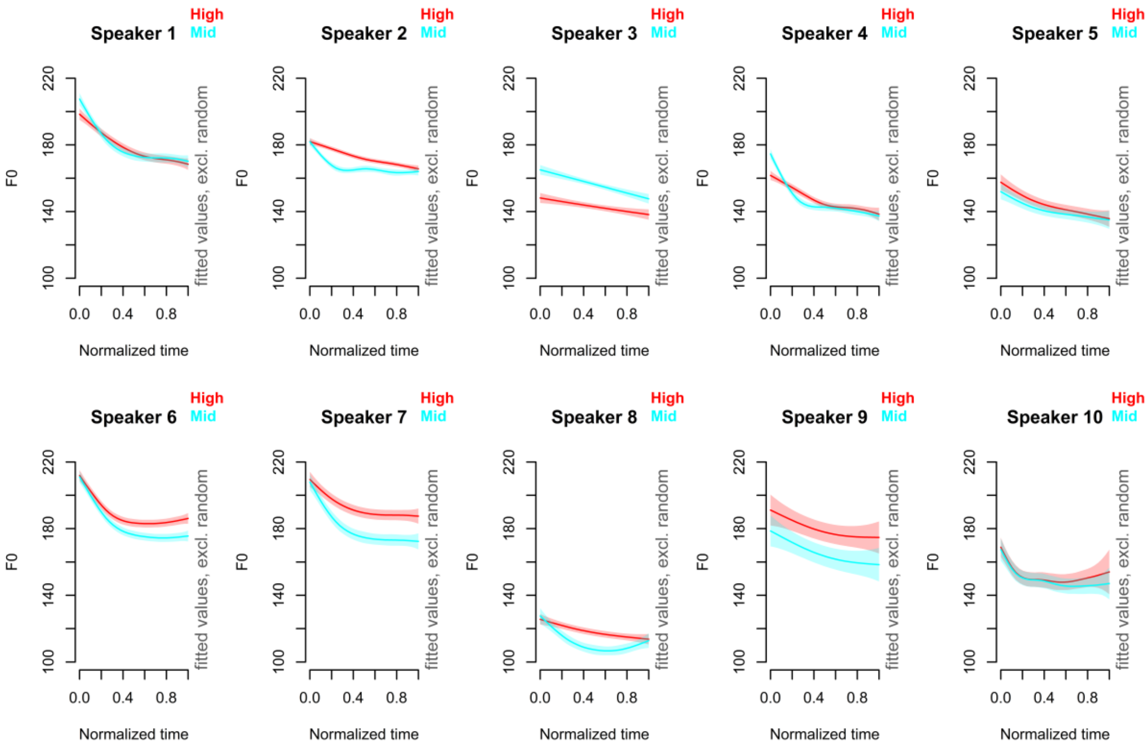


Figure 3.30 Non-linear smooths of high vowels and mid vowels carrying T6 produced by each speaker in the older group

Table 3.39 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T6 at timepoint 6, where high and mid vowels differ the most, for each of the five speakers whose high vowels appear to be higher in F0 than mid vowels. All five speakers show that the F0 values of high vowels are higher than mid vowels by about 5-20 Hz. The mean difference of the nine speakers is 11.5 Hz. Table 3.40 shows that for the five speakers who appear to show no IF0, four of them show differences of less than five Hz, and the mean difference is -1.3 Hz.

Table 3.39 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T6 at measurement point six for each speaker of the older group who, on visual inspection, appears to show IF0 effect

Speaker	2	6	7	8	9
Mean Diff	5.3	6.9	14.2	9.4	21.6
Mean Diff for the five speakers = 11.5					

Table 3.40 **Mean F0 difference between high vowels and mid vowels (high – mid) carrying T6 at measurement point six for each speaker of the older group who, on visual inspection, appears to show no IF0 effect**

Speaker	1	3	4	5	10
Mean Diff	-0.5	-14.6	1.5	2.6	4.5
Mean Diff for the five speakers = -1.3					

3.2.7 Younger group, T1

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T1 produced by the younger group. GAMM does not show a significant difference between the two levels of vowel heights carrying T1. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment. Table 3.28 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the younger group. The mean differences are positive at all timepoints, and they are very small (< 5 Hz) throughout the vocalic segment. The biggest mean difference value is 4.5 Hz, which occurs at timepoint 5. The mean difference is similar throughout the middle of the vocalic segment.

Table 3.41 **Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T1 produced by the younger group**

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-2.8	1.7	3.6	3.8	4.5	2.9	3.3	3.6	4.1
At Point 5, Mean Diff is the biggest = 4.5									

Figure 3.31 shows the F0 data points for each of the ten speakers in the younger group. Speakers 15 and 16 show less concentrated and regular data dispersion.

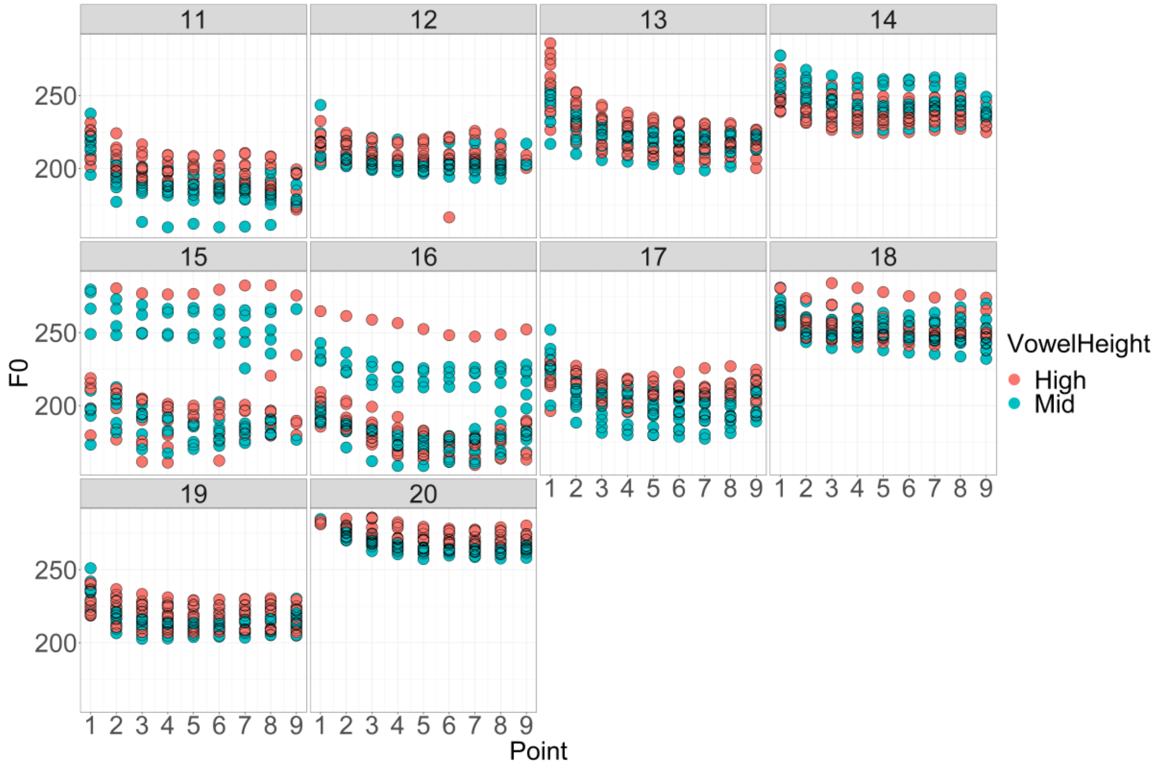


Figure 3.31 F0 data points for high vowels and mid vowels carrying T1 produced by each speaker in the older group

Figure 3.32 shows the GAMM non-linear smooths of the high and mid vowels carrying T1 produced by each participant. For seven speakers (speakers 11, 12, 13, 17, 18, 19, 20), high vowels carrying T1 appear to show higher F0 than mid vowels throughout most of the vocal segment. For speakers 14, 15, and 16, high vowels carrying T1 appear to show lower F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

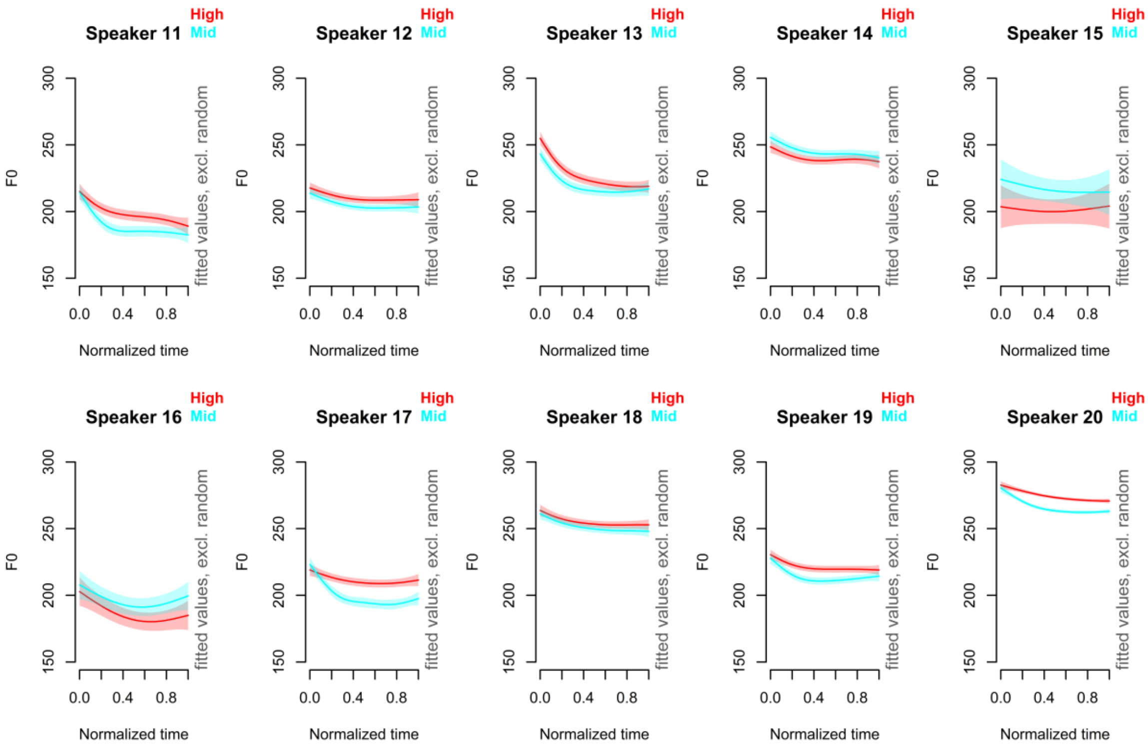


Figure 3.32 Non-linear smooths of high vowels and mid vowels carrying T1 produced by each speaker in the younger group

Table 3.42 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T1 at timepoint 5, where high and mid vowels differ the most, for each of the seven speakers whose high vowels appear to be higher in F0 than mid vowels. All seven speakers show that the F0 values of high vowels are higher than mid vowels by about 2-15 Hz. The mean difference of the seven speakers is 8.5 Hz. Table 3.43 shows that for the three speakers who appear to show no IF0, the mean difference (high - mid) is -8.9 Hz.

Table 3.42 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T1 at measurement point five for each speaker of the younger group who, on visual inspection, appears to show IF0 effect

Speaker	11	12	13	17	18	19	20
Mean Diff	10.4	7.1	6.2	15.0	2.5	7.9	10.1
Mean Diff for the seven speakers = 8.5							

Table 3.43 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T1 at measurement point five for each speaker of the younger group who, on visual inspection, appears to show no IF0 effect

Speaker	14	15	16
Mean Diff	-4.0	-12.5	-10.1
Mean Diff for the three speakers = -8.9			

3.2.8 Younger group, T2

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T2 produced by the younger group. GAMM shows a significant difference between the two levels of vowel heights carrying T2 at the middle and later portion of the vocalic segment. Additionally, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment. Table 3.30 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the younger group. The mean differences are positive at all timepoints. The biggest mean difference value is 22.1 Hz, which occurs at timepoint 8. The mean difference increases from the beginning to the end of the vocalic segment.

Table 3.44 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T2 produced by the younger group

Point	2	3	4	5	6	7	8
Mean Diff	10.0	12.6	14.2	13.7	15.5	18.4	22.1
At Point 8, Mean Diff is the biggest = 22.1							

Figure 3.33 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is relatively concentrated and regular. Speakers 15 and 16's data dispersion is less concentrated than the other speakers.

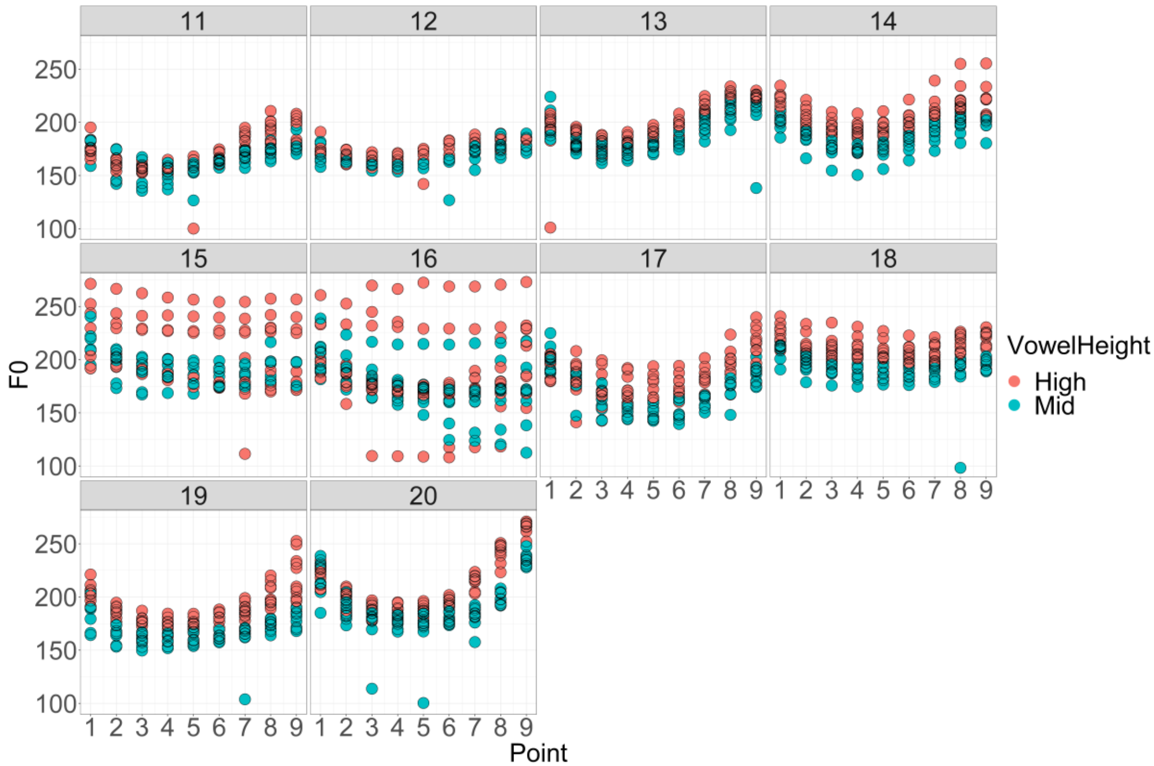


Figure 3.33 F0 data points for high vowels and mid vowels carrying T2 produced by each speaker in the younger group

Figure 3.34 shows the GAMM non-linear smooths of the high and mid vowels carrying T2 produced by each participant. For all ten speakers, high vowels carrying T2 appear to show higher F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

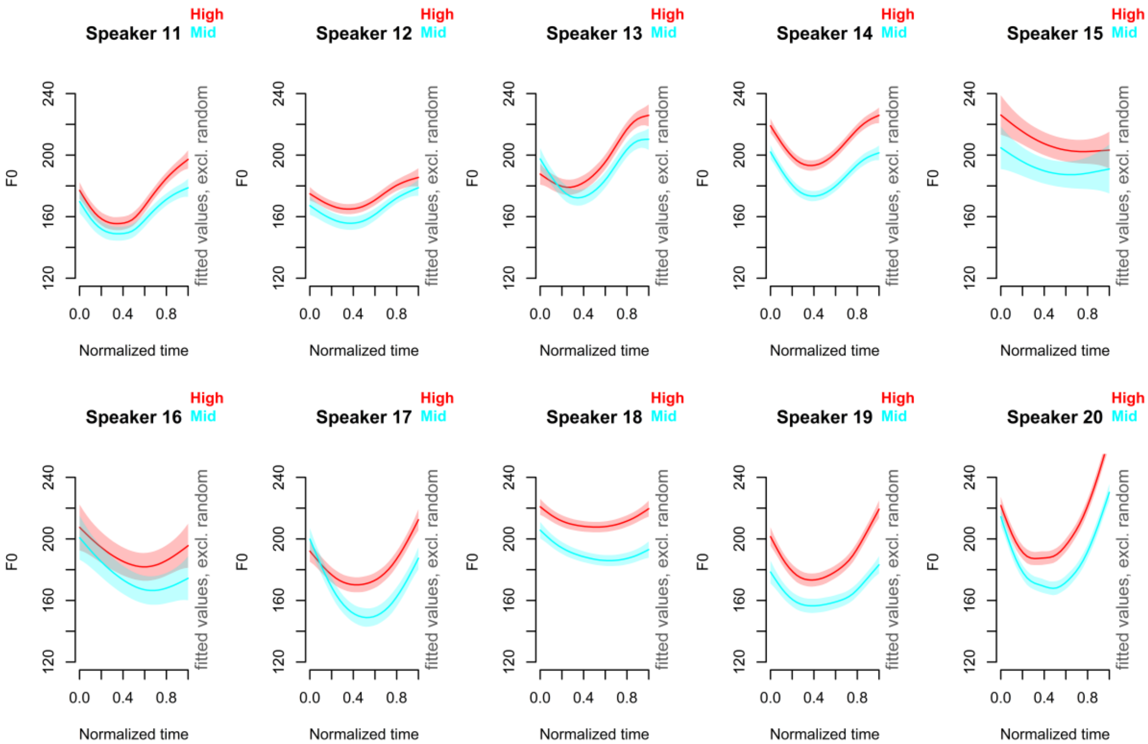


Figure 3.34 Non-linear smooths of high vowels and mid vowels carrying T2 produced by each speaker in the younger group

Table 3.45 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T2 at timepoint 8, where high and mid vowels differ the most, for each speaker of the younger group. All the speakers show positive values, which indicate higher F0 in high vowels compared to mid vowels. The speakers show difference values of 7-44 Hz.

Table 3.45 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T2 at measurement point eight for each speaker of the younger group

Speaker	11	12	13	14	15	16	17	18	19	20
Mean Diff	17.6	7.0	10.9	23.2	11.4	17.9	23.1	32.6	27.7	43.4
Mean Diff for all the 10 speakers = 22.1										

3.2.9 Younger group, T3

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T3 produced by the younger group. GAMM does not show a significant difference between the two levels of vowel heights carrying T3. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the later portion of the vocalic segment. Table 3.46 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the younger group. The mean differences are positive at all timepoints except for timepoint 6. However, the values of the differences are very small (<2 Hz), which appears to be undetectable for humans. The biggest mean difference value is 1.2 Hz, which occurs at timepoint 8. The mean difference is similar throughout the vocalic segment.

Table 3.46 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T3 produced by the younger group

Point	2	3	4	5	6	7	8
Mean Diff	0.8	0.9	0.2	0.3	-0.6	0.4	1.2
At Point 8, Mean Diff is the biggest = 1.2							

Figure 3.35 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is relatively concentrated and regular. Speakers 15 and 16's data dispersion is less concentrated than the other speakers.

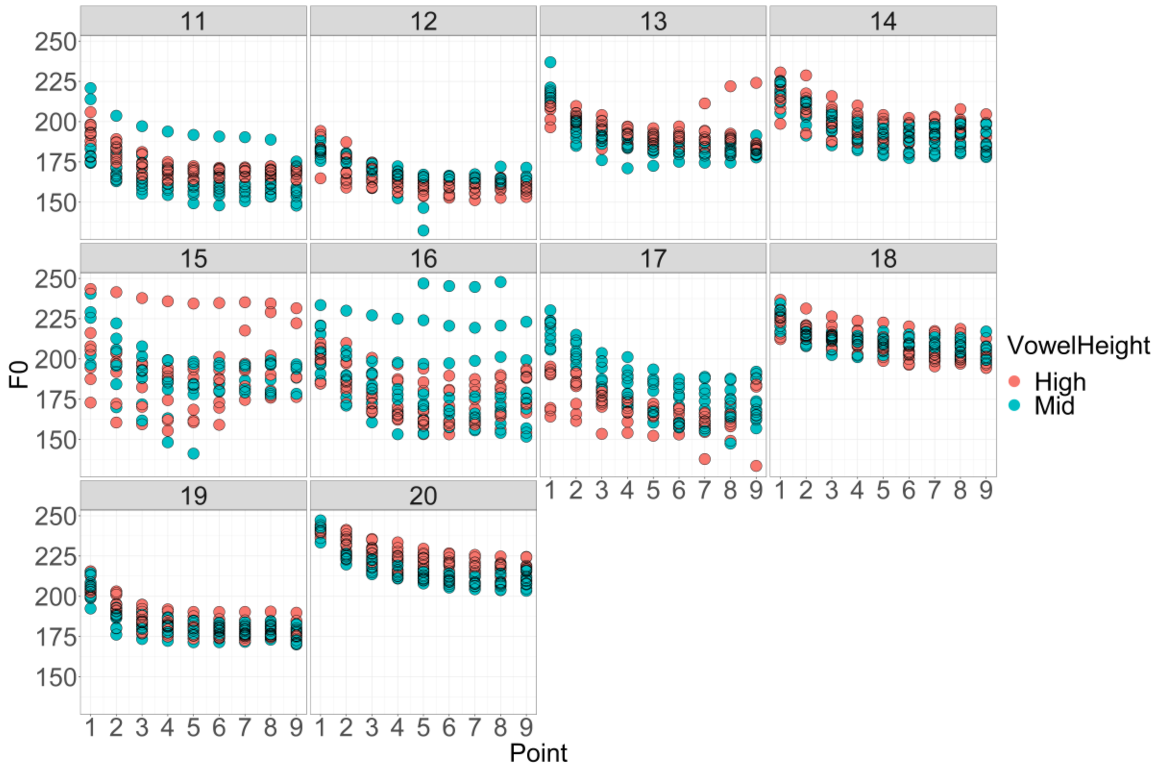


Figure 3.35 **F0 data points for high vowels and mid vowels carrying T3 produced by each speaker in the younger group**

Figure 3.36 shows the GAMM non-linear smooths of the high and mid vowels carrying T3 produced by each participant. For five speakers (speakers 11, 13, 14, 19, 20), high vowels carrying T3 appear to show higher F0 than mid vowels throughout most of the vocal segment. For speakers 12, 16, 17 and 18, high vowels carrying T3 appear to show lower F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

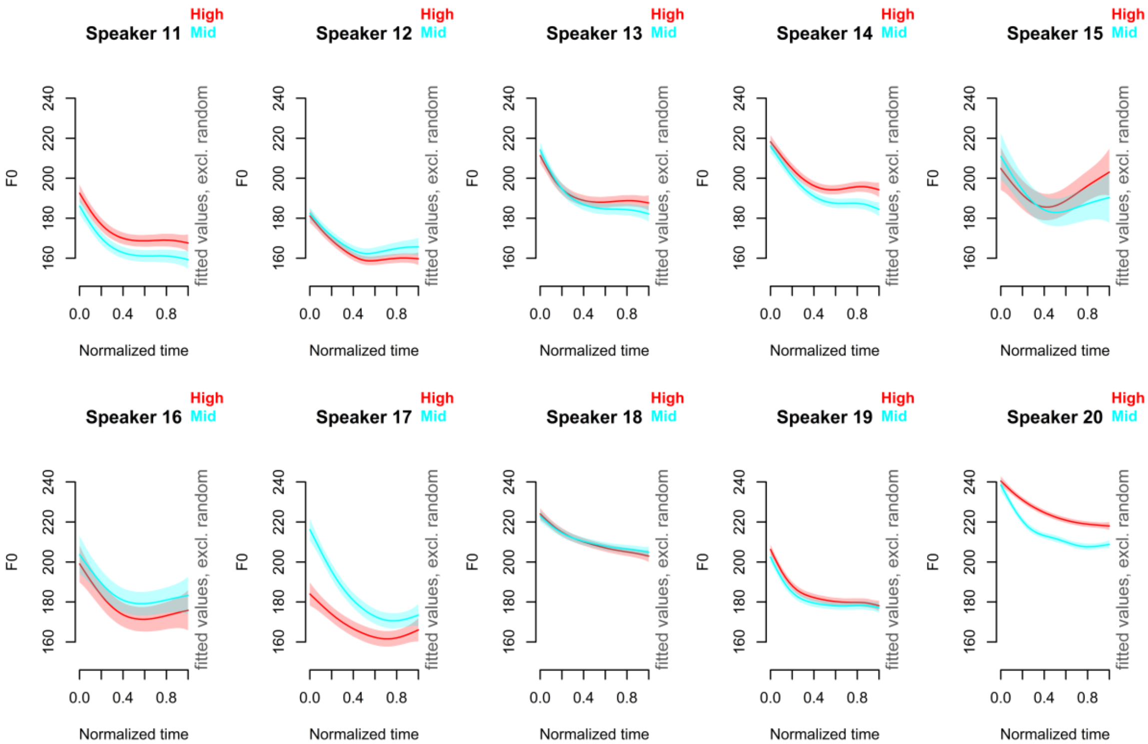


Figure 3.36 Non-linear smooths of high vowels and mid vowels carrying T3 produced by each speaker in the younger group

Table 3.47 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T3 at timepoint 8, where high and mid vowels differ the most, for each of the five speakers whose high vowels appear to be higher in F0 than mid vowels. All five speakers show that the F0 values of high vowels are higher than mid vowels by about 0.4 -12 Hz. The mean difference of the nine speakers is 6.6 Hz. Table 3.48 shows that for the five speakers who appear to show no IF0, the mean difference (high - mid) is -3.2 Hz.

Table 3.47 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T3 at measurement point eight for each speaker of the younger group who, on visual inspection, appears to show IF0 effect

Speaker	11	13	14	19	20
Mean Diff	7.5	5.7	8.4	0.4	11.2
Mean Diff for the five speakers = 6.6					

Table 3.48 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T3 at measurement point eight for each speaker of the younger group who, on visual inspection, appears to show no IF0 effect

Speaker	12	15	16	17	18
Mean Diff	-5.2	8.7	-9.0	-9.7	-0.7
Mean Diff for the five speakers = -3.2					

3.2.10 Younger group, T4

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T4 produced by the younger group. GAMM does not show a significant difference between the two levels of vowel heights carrying T4. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive in the later portion of the vocalic segment. Table 3.49 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the younger group. The mean differences are positive in the later portion of the vocalic segment, which include timepoint 6, 7, 8, and 9. The biggest mean difference value is 9.9 Hz, which occurs at timepoint 7. The mean difference is similar throughout the later portion of the vocalic segment.

Table 3.49 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T4 produced by the younger group

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-11.0	-5.2	-1.8	-1.6	-3.0	5.9	9.9	5.7	6.5
At Point 7, Mean Diff is the biggest = 9.9									

Figure 3.37 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is much less relatively concentrated and regular than the other tones of the younger group. It is also less concentrated and regular than T4 produced by the older group. Most speakers creak a lot when producing this tone. The creaky voice causes many missing data points, especially for speakers 11, 12, 17 and 19. For speakers 17 and 19, all the data points of mid vowels in the later portion of the vocalic segment are missing, so they are excluded in the analysis of T4. For the eight speakers remaining, it appears that they creak more in mid vowels than high vowels.

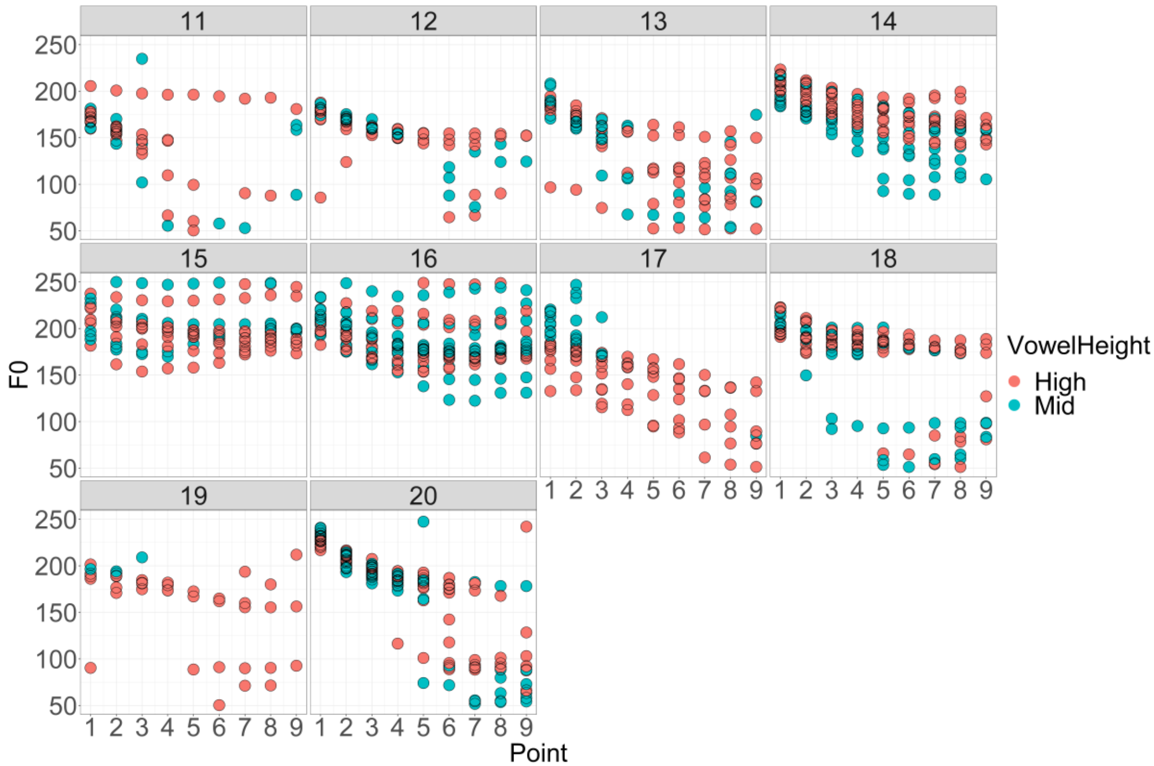


Figure 3.37 F0 data points for high vowels and mid vowels carrying T4 produced by each speaker in the younger group

Figure 3.38 shows the GAMM non-linear smooths of the high and mid vowels carrying T4 produced by each participant. Since all the data points of the mid vowel in the later portion of the vocalic segment are missing for speakers 17 and 19, the GAMM plots of these two speakers are not taken into consideration. Speakers 11, 13, 14, 18 and 20 show that high vowels carrying T4 are higher in F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C. Comparison of figure 3.37 and 3.38 shows that the speakers who appear to show IF0 tend to show more creaky voice.

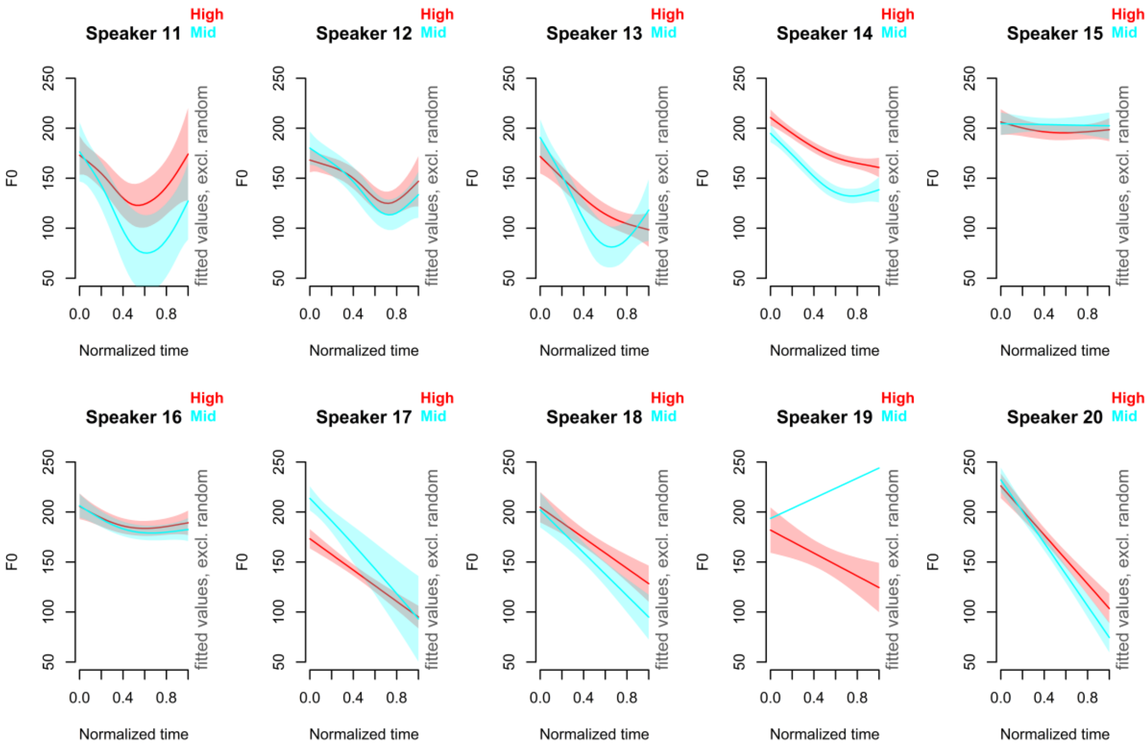


Figure 3.38 Non-linear smooths of high vowels and mid vowels carrying T4 produced by each speaker in the younger group

Speakers 17 and 19 are excluded due to missing data points. Table 3.50 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T4 at timepoint 7, where high and mid vowels differ the most, for each of the five speakers whose high vowels appear to be higher in F0 than mid vowels. Among these five speakers, speakers 11, 13, 14 and 20 show great differences of over 20 Hz, while speaker 18 shows a smaller difference of only 2.9 Hz. However, it can be seen from Figure 3.37 that speaker 11 has only three valid data points at timepoint 7, so the result is not very representative. The mean difference of the five speakers is 36.6 Hz. Table 3.51 shows the three speakers who appear to show no IF0. Speaker 15 show that mid vowels are higher than high vowels by 1.4 Hz. The other two speakers all show higher high vowels than mid vowels. The mean difference (high - mid) for these speakers is 7.2 Hz.

Table 3.50 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T4 at measurement point seven for each speaker of the younger group who, on visual inspection, appears to show IF0 effect

Speaker	11	13	14	18	20
Mean Diff	88.2	21.2	36.9	2.9	33.8
	Mean Diff for the five speakers = 36.6				

Table 3.51 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T4 at measurement point seven for each speaker of the younger group who, on visual inspection, appears to show no IF0 effect

Speaker	12	15	16
Mean Diff	15.0	-1.4	7.9
Mean Diff for the three speakers = 7.2			

3.2.11 Younger group, T5

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T5 produced by the older group. GAMM does not show a significant difference between the two levels of vowel heights carrying T5. However, it can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is positive throughout the vocalic segment. Table 3.52 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the younger group. The mean differences are positive at all timepoints, and they are very small throughout the vocalic segment. The biggest mean difference value is 7.2 Hz, which occurs at timepoint 7. The mean difference is similar throughout the vocalic segment.

Table 3.52 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T5 produced by the younger group

Point	1	2	3	4	5	6	7	8	9
Mean Diff	-4.7	3.9	4.8	3.6	2.8	6.2	7.2	4.7	2.9
At Point 7, Mean Diff is the biggest = 7.2									

Figure 3.39 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is relatively concentrated and regular.

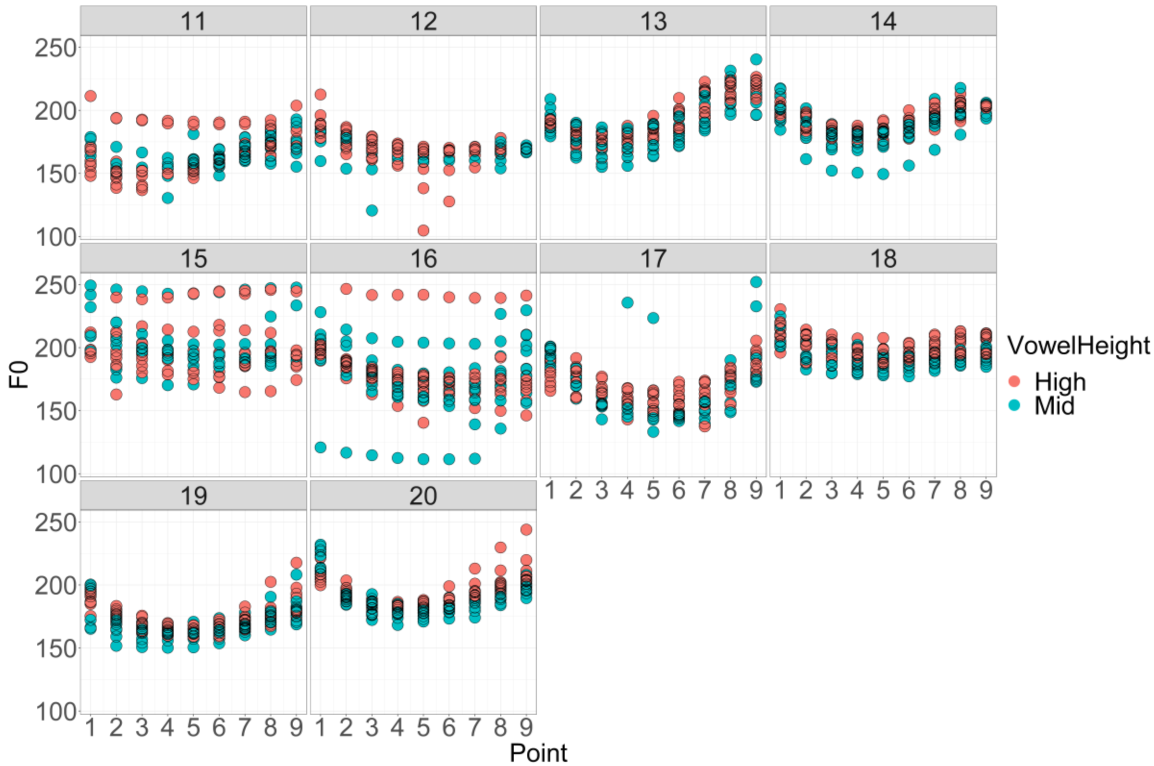


Figure 3.39 F0 data points for high vowels and mid vowels carrying T5 produced by each speaker in the younger group

Figure 3.40 shows the GAMM non-linear smooths of the high and mid vowels carrying T5 produced by each participant. For five speakers (speakers 13, 14, 18, 19, 20), high vowels carrying T5 appear to show higher F0 than mid vowels throughout most of the vocal segment. For speaker 12, high vowels carrying T5 appear to show higher F0 than mid vowels in the first half of the vocalic segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

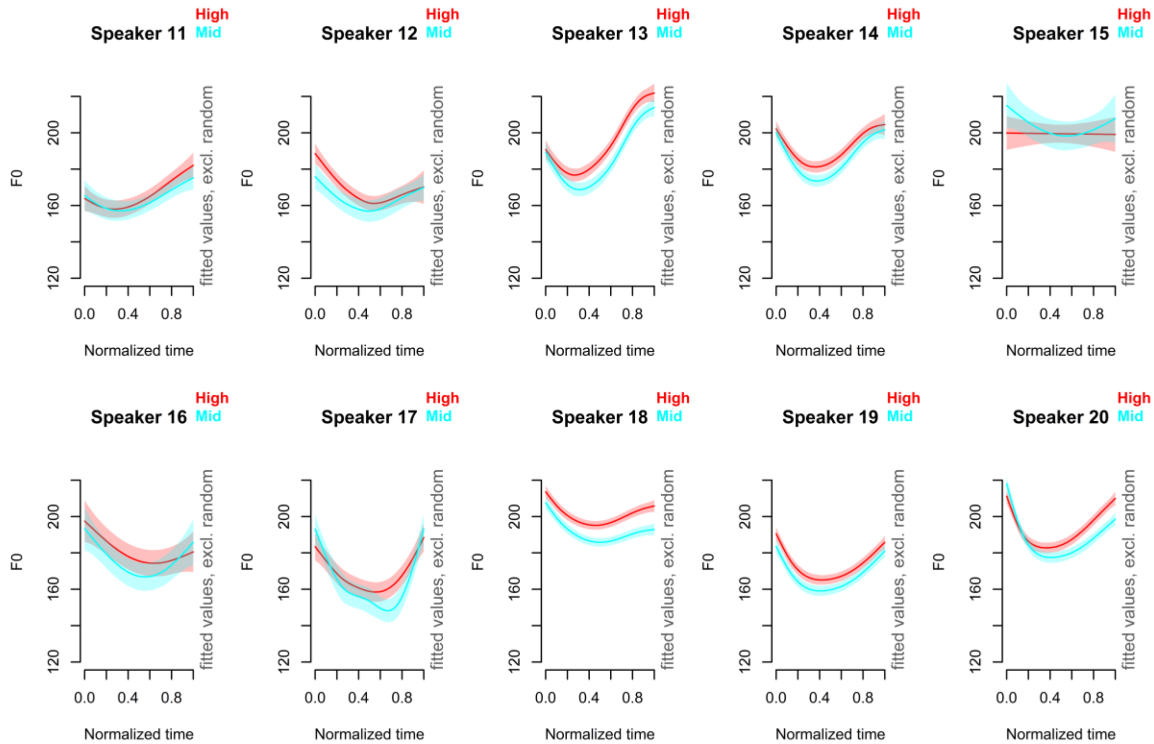


Figure 3.40 Non-linear smooths of high vowels and mid vowels carrying T5 produced by each speaker in the younger group

Table 3.37 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T5 at timepoint 7, where high and mid vowels differ the most, for each speaker of the younger group. All ten speakers show positive values, which indicate higher F0 in high vowels compared to mid vowels. Among these speakers, speaker 12 and speaker 15 show positive differences of around 2 Hz, which is very small. The other speakers show positive differences of between 5-15 Hz.

Table 3.53 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T5 at timepoint 7, where high and mid vowels differ the most, for each of the five speakers whose high vowels appear to be higher in F0 than mid vowels. All five speakers show that the F0 values of high vowels are higher than mid vowels by about 5-15 Hz. The mean difference of the five speakers is 9.2 Hz. Table 3.54 shows the five speakers who appear to show no IF0. For these speakers, high vowels are higher than mid vowels by 2-13 Hz. The mean difference (high - mid) is 6.7 Hz.

Table 3.53 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T5 at measurement point seven for each speaker of the younger group who, on visual inspection, appears to show IF0 effect

Speaker	13	14	18	19	20
Mean Diff	14.5	5.7	10.4	5.1	10.5
Mean Diff for the five speakers = 9.2					

Table 3.54 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T5 at measurement point seven for each speaker of the younger group who, on visual inspection, appears to show no IF0 effect

Speaker	11	12	15	16	17
Mean Diff	5.3	2.1	2.4	11.1	12.5
Mean Diff for the five speakers = 6.7					

3.2.12 Younger group, T6

Figure 3.15 and 3.17 included the non-linear smooths and difference smooths for T6 produced by the younger group. GAMM does not show a significant difference between the two levels of vowel heights carrying T6. Different from the other tones, It can be seen from the GAMM figures that the estimated difference in F0 (high – mid) is negative throughout the vocalic segment. Table 3.55 shows the mean F0 difference between the two levels of vowel height (high - mid) for each F0 measurement point (2-8) for all speakers of the younger group. The mean differences are positive at only timepoint 8, and the absolute values of mean differences are very small throughout the vocalic segment. The biggest positive mean difference value is 0.3 Hz, which occurs at timepoint 8.

Table 3.55 Mean F0 difference between high vowels and mid vowels (high - mid) for each F0 measurement point of T6 produced by the younger group

Point	2	3	4	5	6	7	8
Mean Diff	-0.4	-4.2	-3.8	-3.5	-2.1	-1.3	0.3
At Point 8, Mean Diff is the biggest = 0.3							

Figure 3.41 shows the F0 data points for each of the ten speakers in the younger group. Datapoint dispersion is relatively concentrated and regular. Speakers 15 and 16's datapoint dispersion is relatively less concentrated and regular.

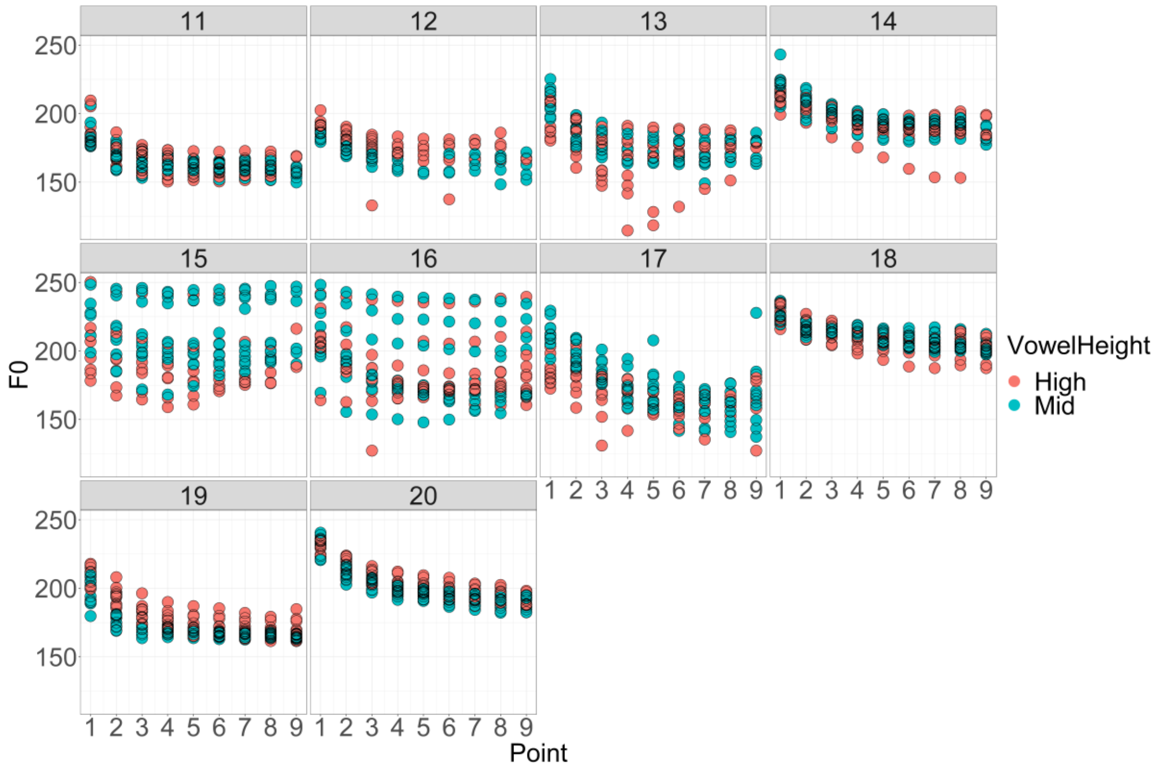


Figure 3.41 F0 data points for high vowels and mid vowels carrying T6 produced by each speaker in the younger group

Figure 3.42 shows the GAMM non-linear smooths of the high and mid vowels carrying T6 produced by each participant. For only three speakers (speakers 12, 19, 20), high vowels carrying T6 appear to show higher F0 than mid vowels throughout most of the vocal segment. For speakers 13, 14, 15, 17 and 18, high vowels carrying T6 appear to show lower F0 than mid vowels throughout most of the vocal segment. This finding is also supported by the GAMM difference smooths of high and mid vowels produced by each participant, which are shown in Appendix C.

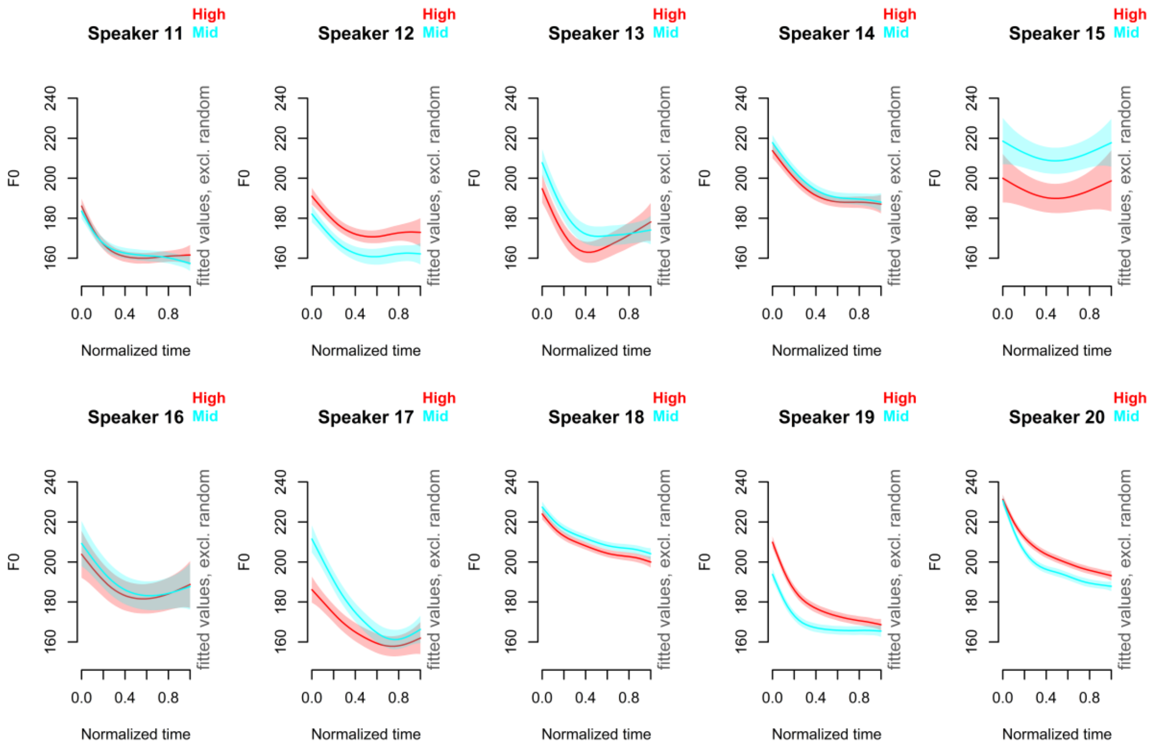


Figure 3.42 Non-linear smooths of high vowels and mid vowels carrying T6 produced by each speaker in the younger group

Table 3.56 shows the mean F0 difference between high vowels and mid vowels (high – mid) carrying T6 at timepoint 8, where high and mid vowels differ the most, for each of the three speakers whose high vowels appear to be higher in F0 than mid vowels. All three speakers show that the F0 values of high vowels are higher than mid vowels by about 4-12 Hz. The mean difference of the three speakers is 7.5 Hz. Table 3.57 shows the seven speakers who appear to show no IF0. The mean difference (high - mid) is -2.8 Hz.

Table 3.56 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T6 at measurement point eight for each speaker of the younger group who, on visual inspection, appears to show IF0 effect

Speaker	12	19	20
Mean Diff	11.9	4.0	6.6
Mean Diff for the three speakers = 7.5			

Table 3.57 Mean F0 difference between high vowels and mid vowels (high – mid) carrying T6 at measurement point eight for each speaker of the younger group who, on visual inspection, appears to show no IF0 effect who, on visual inspection, appears to show no IF0 effect

Speaker	11	13	14	15	16	17	18
Mean Diff	1.9	0.4	-2.9	-16.3	0.8	0.0	-3.6
Mean Diff for the seven speakers = -2.8							

3.2.13 Summary of analyses for the IF0 effects

Table 3.58 shows the GAMM estimates and *p* values for each tone pair by age group combination. Table 3.59 shows the mean F0 differences at the measurement point where the difference is the biggest for speakers who appear to show IF0 effect. Table 3.60 shows the mean F0 differences at the measurement point where the difference is the biggest for speakers who appear to show no IF0 effect. When comparing the older group to the younger group, both absolute values of the GAMM estimates and the values of the biggest mean F0 difference for speakers who show IF0 suggest that for all the tones except for T4, the older group shows a bigger difference between the two levels of vowel heights than the younger group, with high vowels showing higher F0 than mid vowels. Therefore, the results imply that although the IF0 effect is statistically non-significant for most tones, the older group appears to show a more substantial IF0 effect than the younger group. For T4, the younger group shows a bigger difference. However, there is a lack of datapoints for T4, especially for the younger group due to creaky voice and low F0. These imply that the mean F0 difference for T4 is not very accurate, especially for the younger group. T2 is the only tone that shows a significant or nearly significant difference in the GAMM outputs and the GAMM figures. The mean F0 difference at the measurement point where the difference is the biggest for speakers who show IF0 is also more substantial in T2 than that in the other tones for the older group. For the younger group, it is also bigger than the other tones except for T4. For the older group, both absolute values of the GAMM estimates and the values of the biggest mean F0 difference for speakers who show IF0 suggest that the F0 difference between the two levels of vowel heights (high – mid) from the biggest to the smallest is T2 > T4 > T5/T1 > T6 > T3. For the younger group, it is T4 > T2 > T5/T1 > T6 > T3.

Table 3.58 GAMM estimates and p values for each tone by age group combination (mid – high)

Older						
Tone	T1	T2	T3	T4	T5	T6
GAMM Estimate	-5.263	-19.400	-3.333	-11.248	-4.854	-3.398
p	0.748	0.067	0.780	0.263	0.601	0.747
Younger						
Tone	T1	T2	T3	T4	T5	T6
GAMM Estimate	-2.001	-15.786	-0.324	-1.976	-4.398	2.980
p	0.867	0.012	0.967	0.860	0.477	0.695

Table 3.59 Mean F0 difference at the measurement point where the difference is the biggest for speakers who, on visual inspection, appear to show IF0 effect

	Older (age 35-45)		Younger (age 18-28)	
	Point	Mean F0 Diff	Point	Mean F0 Diff
T1	5	12.4	5	8.5
T2	7	29.7	8	22.1
T3	8	7.9	8	6.6
T4	6	24.4	7	36.6
T5	6	12.5	7	9.2
T6	6	11.5	8	7.5

Table 3.60 Mean F0 difference at the measurement point where the difference is the biggest for speakers who, on visual inspection, appear to show no IF0 effect

	Older (age 35-45)		Younger (age 18-28)	
	Point	Mean F0 Diff	Point	Mean F0 Diff
T1	5	0.1	5	-8.9
T2	7	N/A	8	N/A
T3	8	0.7	8	-3.2
T4	6	4.0	7	7.2
T5	6	-0.2	7	6.7
T6	6	-1.3	8	-2.8

Although the general GAMM results do not show statistical significance between the two levels of vowel heights for most tone by age group combinations, individual analyses show that for all of the tones except for T6 produced by younger speakers, more than half or half of the speakers in both age groups appear to show IF0. Table 3.61 shows the speakers who show obviously higher F0 in high vowels than mid vowels in each tone by age group combination. For both groups, all ten speakers show obvious IF0 in T2, but only some speakers show obvious IF0 in the other tones. Only speaker 15 and 16 show no IF0 in all five tones except for T2. Only speakers 2, 6, 9, and 20 show IF0 in all six tones. The other speakers show IF0 in some tones, but show no IF0 in the other tones. There are more speakers who show obvious an IF0 effect in the older group compared to the younger group in every tone except for T1. For the older group, the percentage of speakers who show obvious IF0 from the greatest to the least is T2 > T4 > T1, T3, T5 > T6. For the younger group, it is T2 > T1 > T4 > T3, T5 > T6.

Table 3.61 Speakers that show obviously higher F0 in high vowels than mid vowels in each tone by age group combination

	Older (age 35-45)			Younger (age 18-28)		
	Speakers	Num.	%	Speakers	Num.	%
T1	2, 3, 5, 6, 7, 9	6 out of 10	60%	11, 12, 13, 17, 18, 19, 20	7 out of 10	70%
T2	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10 out of 10	100%	11, 12, 13, 14, 15, 16, 17, 18, 19, 20	10 out of 10	100%
T3	2, 5, 6, 7, 9, 10	6 out of 10	60%	11, 13, 14, 19, 20	5 out of 10	50%
T4	1, 2, 5, 6, 8, 9, 10	7 out of 10	70%	11, 13, 14, 18, 20	5 out of 8	62.5%
T5	1, 2, 4, 5, 6, 9	6 out of 10	60%	13, 14, 18, 19, 20	5 out of 10	50%
T6	2, 6, 7, 8, 9	5 out of 10	50%	12, 19, 20	3 out of 10	30%

Individual analyses show inter-speaker variation in the tones in which they show an obvious IF0 effect. Table 3.62 shows that in each tone, except for T2, some speakers show obvious IF0, while other speakers do not. Table 3.41 and 3.43 show that some speakers (speakers 2, 5, 6, 9, 10 in the older speakers; speaker 11, 19, 20 in the younger speakers) tend to show higher F0 in high vowels than mid vowels at the measurement point where two tones differ the most in all six tones, while other speakers do not. When looking at whether an individual tends to show obvious IF0 in all tones, there is not obvious within-speaker consistency. In other words, some speakers always show obvious IF0 in all six tones (speakers 6, 9, 20), some speakers (speakers 15, 16) do not show obvious IF0 in any tones except for T2, while other speakers show obvious IF0 in some of the six tones. For example, speaker 1 does not show obvious IF0 in T1, T3 and T6, which are level tones, but shows obvious IF0 in T2, T4, T5, which are rising and low-falling tones, Speaker 18 show obvious IF0 in T1, T2, T4, T5, but not in T3 and T6.

Table 3.62 Mean F0 difference between high vowels and mid vowels (high – mid) at the measurement point where the difference is the greatest for each tone by speaker combination

Older group										
Speaker	1	2	3	4	5	6	7	8	9	10
T1	-2.9	13.8	17.6	-8.2	5.8	6.3	-1.7	-28.0	13.4	12.0
T2	45.7	27.9	5.6	37.5	42.1	24.7	35.0	15.7	16.4	34.5
T3	-1.0	3.6	-1.2	0.4	4.4	9.1	18.7	4.7	7.6	4.0
T4	7.1	7.6	-2.5	4.4	6.9	49.0	10.2	13.5	3.6	83.0
T5	16.8	12.4	-9.8	8.9	10.0	4.6	3.2	0.8	22.2	4.9
T6	-0.5	5.3	-14.6	1.5	2.6	6.9	14.2	9.4	21.6	4.5
Younger group										
Speaker	11	12	13	14	15	16	17	18	19	20
T1	10.4	7.1	6.2	-4.0	-12.5	-10.1	15.0	2.5	7.9	10.1
T2	17.6	7.0	10.9	23.2	11.4	17.9	23.1	32.6	27.7	43.4
T3	7.5	-5.2	5.7	8.4	8.7	-9.0	-9.7	-0.7	0.4	11.2
T4	88.2	15.0	21.2	36.9	-1.4	7.9	Missing	2.9	Missing	33.8
T5	5.3	2.1	14.5	5.7	2.4	11.1	12.5	10.4	5.1	10.5
T6	1.9	11.9	0.4	-2.9	-16.3	0.8	0	3.6	4.0	6.6

To summarize, for both age group, although GAMM model outputs do not show statistical significance *($p < 0.05$)* between F0 values of high and mid vowels in all tones except for T2 produced by younger speakers, individual analysis shows that more than half or half of the speakers in both age groups appear to show IF0 in all tones except for T6 produced by younger speakers. Both general analysis and individual analyses show that if we exclude T4 due to creaky voice and lack of data points, for the other tones, the size of IF0 from the greatest to the smallest is: T2 > T5/T1 > T3, T6.

Chapter 4

Discussion and conclusion

This study investigates the intrinsic fundamental frequency (IF0) effect in the six Hong Kong Cantonese tones. The tension between exaggerating the IF0 effect to enhance vowel height contrasts and restraining the IF0 effect to keep easily confusable tone pairs contrastive is the main focus. The goals of the investigation are to 1) study the relationship between IF0 and tone merging and 2) test the mixed physiological-enhancement hypothesis (*MPEH*), which states that IF0 is physiologically determined, but may be enhanced or restrained by speakers to help with the perception of phonological contrasts. GAMM models on IF0 only showed significant IF0 effects in younger speakers' T2 and near-significant IF0 effects in older speakers' T2, but not in other tone-by-age group combinations. Additionally, the statistical power of the GAMM model was low and indicated great inter-speaker variation. Although overall statistical outputs provided very limited information, examinations of the mean F0 difference at the timepoint where F0 values of the two levels of vowel height differ the most, as well as analysis of individual data pointed to interesting phenomena. These findings will be addressed in the sections that follow.

4.1 Discussion of results

4.1.1 T4 and creaky voices

In Section 1.3, it was predicted that if the *MPEH* is correct, T4 would show the smallest IF0 effect. However, the results showed that for the speakers who showed IF0 in T4, IF0 of T4 was much greater than predicted, being the second greatest in the older group, and the greatest in the younger group. Individual analyses showed that there was great inter-speaker variation in the size of IF0 of T4 for both age groups. Even when we exclude the speakers with very few valid data points (speaker 10 and speaker 11), some speakers showed very great IF0 of over 30 Hz (speakers 6, 14, 20), while the others did not show IF0 (speakers 3, 4, 7, 12, 15, 16) or showed small values of less than 10 Hz (speakers 1, 2, 5, 9, 18). Visual inspections of data distributions of T4 showed considerable production of creaky voice, and people tended to creak more when producing mid vowels. For both age groups, the speakers who showed bigger IF0 in T4 were the ones who showed more creaky voice. These results suggest that the bigger-than-predicted IF0 in T4 is partially because people show more creaky voice with very low F0 when producing T4, especially when producing mid vowels carrying T4. This finding aligns with the finding of Yu and Lam (2014) that differences in voice quality play a role in perceiving T4 in HK Cantonese. What's more, visual inspection of the data distribution figures showed that younger speakers creaked more when producing T4 than older speakers. Results of the confusable tone pairs showed that some younger speakers appeared to merge T4 and T6, which aligns with previous research (Mok et al., 2013, Fung and Lee, 2019), while none of the older speakers merged these two tones. These results suggest that younger speakers may rely more on creaky voice to distinguish T4 and T6, as these two tones begin to merge in F0. This tendency may explain the great IF0 in T4 especially for the younger speakers, because if speakers rely more on creaky voice as a cue for T4 in production and perception, they no longer need to suppress IF0 to make T4 distinct from T6 in F0. What's more, since T4 is a low falling tone while T6 is a level tone, other than mean F0, speakers may also distinguish T4 from T6 by a tone shape difference. However, previous studies have not studied whether people distinguish T4 from other tones based on tone shape, so future work on acoustic cues for Cantonese tone perception may be needed to establish this. In the present study, the results of T4 were less accurate than those of the other tones, especially for the younger group, because Praat sometimes failed to track extremely low F0 in creaky voice; consequently, many data points were lost. However, there were still some speakers (speakers 1, 2, 5, 14) who did not show creaky voice in T4 according to visual inspection of the data distributions. According to Whalen & Levitt (1995) and Connell (2002), IF0 is very small or disappears in the low F0 range universally, which is likely due to physiological factors that constrain IF0. However, these four speakers all showed greater IF0 in T4, which is in the very low F0 range, than T3 and T6, which are in a higher F0 range. This finding suggests that the relatively greater size of IF0 in T4 in the speakers who did not show creaky voice challenges the physiological hypothesis that IF0 is a physiological by-product of vowel production that speakers cannot actively control. Instead, these speakers may suppress IF0 less in T4 and suppress IF0 more in T3 and T6, because T4 can be distinguished from T6 by creaky voice, while mean F0 is the only known cue for distinguishing T3 and T6.

4.1.2 Contour vs. level tones

It was also predicted that T1 would show the largest IF0. The overall results of the speakers who showed IF0 was that T1 showed smaller IF0 than T2 and T4, similar IF0 to T5, and greater IF0 than T3 and T6. The fact that T1 showed greater IF0 than T3 and T6 supports the *MPEH*, because T1 is in the very high range and is not confusable with the other tones. However, IF0 of T1 was smaller than T2 and similar to T5 for both the older group and the younger group. When we investigate this individually, for all speakers except speaker 12, IF0 of T1 appeared non-existent or much smaller than for T2. Even for speaker 12, IF0 of T1 and T2 were very similar, both being around 7 Hz. A possible explanation for the greater IF0 in T2 than T1 in almost all speakers is that as a level tone, speakers need to suppress IF0 in T1 more than in contour tones to make it distinct from the other two level tones. In fact, visual inspection of the tone curves of T2 and T5 produced by the speakers showed that two tones differ from each other at the later portion of the vocalic segment, with T2 ending at higher F0. This aligns with previous findings that they are distinguished from each other based on the magnitude of F0 change at the later portion of the vocalic segment (Khouw & Ciocca, 2007; Ciocca et al., 2002). However, visual inspection of the tone curves of the high and mid vowels carrying T2 and the high and mid vowels carrying T5 showed that IF0 is more of an overall additive effect of high vowels throughout the vocalic segment. Therefore, if people can exaggerate IF0 to enhance vowel contrasts, this exaggeration of IF0 on the rising tones will not interfere with listeners' perceptions of tonal distinctions. So, speakers may suppress IF0 less when producing the rising tones than T1, which makes IF0 of T2 bigger than T1. Another possible explanation for greater IF0 in T2 than T1 is that IF0 of contour tones may be bigger than that of level tones for physiological reasons. However, this explanation is called into question because T5 is also a contour tone in the mid F0 range, yet IF0 of T5 is overall similar to T1. Individual analyses also showed that eight of the 20 speakers showed greater IF0 in T1 than T5, while the other 12 speakers showed greater IF0 in T5 than T1. These findings imply that contour or rising tones do not intrinsically show bigger IF0 than level tones. Rather, speakers may control the size of IF0 actively to enhance phonological contrasts. T2 and T5 are both rising tones, but IF0 of T2 was overall greater than that in T5. Individual analyses also showed that for nearly all speakers, except speakers 9 and 13, IF0 of T2 was much greater than T5. There are two possible explanations for this. The first is that the magnitude of change in F0 is smaller in T5 than

T2, which makes T5 flatter than T2. Therefore, producing T5 may require more suppression of IF0 to make it distinct from the level tones in the similar F0 range (T3 and T6). The second explanation is that F0 of T2 is higher than T5 in the later portion of the vocalic segment, so IF0 of T2 may be higher than T5 because IF0 may be higher for physiological reasons. This possible explanation is not supported by the previous literature, as the findings of Whalen & Levitt (1995) and Connell (2002) suggest only that IF0 is very small or disappears at the low F0 range. No studies have investigated whether IF0 gradually decreases as F0 decreases from the high F0 range to the mid F0 range. Since the physiological basis of IF0 is still not well understood, this hypothesis needs to be tested by future studies.

4.1.3 Tone merger in age groups and effect on IF0

GAMM models showed a significant F0 difference in each confusable tone pair of both the younger group and the older group, which indicated that most speakers of HK Cantonese still distinguish the confusable tone pairs in production. Individual analyses also showed that fewer than half of the speakers in both age groups appeared to merge one or more than one of the three confusable tone pairs in production. A higher percent of younger speakers showed tone merging in each tone pair than older speakers. These results suggest that the three confusable tone pairs in HK Cantonese are at the beginning stage of tone merging, which aligns with previous findings (Fung & Lee, 2019; Mok et al., 2013). If speakers could suppress IF0 to enhance the F0 contrasts of the tones, it was predicted that they would show smaller IF0 in T3 and T6 than T1, T2 and T5, because these are level tones and they are easily confusable with each other. This prediction was shown in the IF0 results of both the older group and the younger group. What's more, the younger group, which is more advanced in the merging process of T3 and T6, showed smaller IF0 than the older group in both tones. These results support the hypothesis that speakers can actively control IF0 to enhance phonological contrasts. Another possibility is that IF0 of younger speakers is intrinsically smaller than older speakers due to physiological constraints, because younger speakers overall showed smaller IF0 in all the tones except for T4. Whether this explanation is correct requires future study with more participants and better variable control.

4.1.4 Inter-speaker variability

For both age groups, although GAMM model outputs did not show statistical significance *(p < 0.05)* between F0 values of high and mid vowels except for T2 produced by younger speakers, individual analyses showed that for all tones except for T6 produced by younger speakers, at least half of the speakers in both age groups appeared to show IF0. For both groups, all 10 speakers showed obvious IF0 in T2, but only some speakers did so in the other tones. Speakers 15 and 16 showed no IF0 in any tones but T2. Only speakers 2, 6, 9, and 20 showed IF0 in all six tones. The other speakers showed IF0 in some, but not all, tones. These results imply that there is not a clear boundary between speakers who show IF0 in all tones and speakers who do not show IF0 in any tones – most speakers are in the middle, showing IF0 in some of the tones. This finding supports the *MPEH*, because if IF0 is purely physiological and people cannot actively control it, we should see IF0 in all speakers, and all speakers should show IF0 in certain tones while showing no IF0 in the others (for example, all speakers may show IF0 in T1 because it is a high tone, but show no IF0 in T4 because it is a low tone).

General results for each age group showed that younger speakers, who are more advanced in the merging process, showed smaller IF0 than older speakers; T3 and T6, which are level tones that are easily confused with each other showed smaller IF0 than the other tones. However, individual speaker analyses yielded inconsistencies. Some speakers who showed tone confusion or showed smaller F0 differences between the two tones of each confusable tone pair tended to show smaller IF0 or no IF0 in the relevant tones, which aligns with the general group results. However, some speakers who did not confuse the confusable tone pairs and showed relatively bigger F0 differences between the confusable tone pairs showed a very small IF0, which is inconsistent with the general group results. For example, speaker 3 is the only older speaker who appeared to show tone merging in T2 vs T5 and T3 vs T6. This speaker appeared to show no IF0 in T3, T4, T5 and T6. Her mid vowels even appeared to show higher F0 than high vowels in T4, T5 and T6. This aligns with the finding that speakers who merge tones tend to show smaller or no IF0 in the relevant tones. However, she appeared to show IF0 in T1 and T2, which are in the higher range. This may be due to the fact that the F0 differences between the tones in the mid-low F0 ranges are relatively smaller, so she suppressed IF0 of T3, T4, T5 and T6 more. Speaker 8 did not confuse any of the confusable tone pairs, but showed smaller differences between T2 and T5 as well as T3 and T6 compared to most other older speakers. Also, Speaker 8 did not appear to show IF0 in T1, T3, T5 but did show IF0 in T2, T4, T6. For T5, her mid vowels even appeared to show higher F0 than high vowels. It is hard to explain her IF0 pattern using *MPEH*. Speaker 4 did not confuse any tone pairs and showed relatively big F0 differences between confusable tone pairs. However, she did not appear to show IF0 in T1, T3, T4 or T6, which is hard to explain based on *MPEH* as well.

For the younger speakers, speakers 15 and 16 showed tone merging in all three confusable tone pairs, and they are the only two speakers who showed no IF0 in any tones but T2. For T2, these speakers showed relatively smaller IF0 compared to the other speakers. Speaker 18 appeared to show merging in T3 vs T6, and showed no IF0 in both T3 and T6. Speakers 12 and 14 also appeared to show merging in T3 vs T6. Speaker 12 showed no IF0 in T3, and speaker 14 showed no IF0 in T6. These results align with the overall result that speakers who show tone merging showed smaller or no IF0 because of a need to suppress IF0 more. What's more, speaker 15's mid vowels even appeared to show higher F0 than high vowels in T1, T4, T5 and T6. Speaker 16's mid vowels even appeared to show higher F0 than high vowels in T1 and T3. This "reversed IF0" is also shown in some other speakers, especially in T6 of the younger group. Speakers 13, 14, 15, 17, and 18 all appeared to show "reversed IF0" in T6. Furthermore, speakers 14, 15 and 18 showed tone merging in T3 and T6. Thus, there may be some relationship between the "reversed IF0" and tone merging. One possible explanation is that although some speakers confuse T3 and T6 in production, these two tones are still at the beginning of the merging process, so speakers can still perceive the difference between them and are trying to make them different from each other in production. Since T6 is the lower of these two tones, the speakers may produce lower high vowels than mid vowels to make F0 of the high vowels carrying T6 more distinct from T3. Similarly, speakers 12, 16, 17 and 18 appeared to show lower F0 in high vowels than mid vowels as well, because this can make F0 of the low vowels carrying T3 more distinct from T6. Future studies are required to confirm this hypothesis. The "reversed IF0" also occurs in a few speakers in all the tones except for T2. Speakers 14, 15, and 16 even showed "reversed IF0" in T1, which is not confusable with the other tones. Since these speakers all confused T3 and T6 and speaker 14 and 15 also showed "reversed IF0 in T6, the "reversed IF0" in T1 may be related to the "reversed

IF0” in the other level tones related to tone merging. In other words, for the speakers who showed “reversed IF0” in T6 and T3, once they are used to producing the “reversed IF0” in a level tone, they may produce the “reversed IF0” in the other level tones including T1 as well. Similarly, some speakers showed reversed or no IF0 in T2 and T5, in which suppressing IF0 does not help with perception of tonal contrasts. These findings imply that when speakers suppress IF0 of the confusable level tones to enhance tonal contrasts, they may be used to the suppression and suppress IF0 in the other tones as well. In other words, for many speakers, there is likely to be a general suppression of IF0 effect across many of the six tones. However, since none of the speakers suppressed IF0 of T2, whether the “general suppression” effect exists requires future studies.

4.2 Conclusion

To conclude, the findings of the current study support the MPEH to some extent. However, there are some results that cannot be explained by the MPEH, which require future research on tone merging, the physiological basis of the MPEH, and the relationship between tone merging and change in IF0. In contrast to previous studies on IF0 (Whalen et al., 1995; Connell, 2002) which assumed IF0 to be a simple additive or subtractive effect on the vowels and only measured it on the vowel mid points, the current study measured it on nine points throughout the vocalic segment and analyzed the IF0 effect with GAMM models, which takes the change of F0 over time into consideration and allows visualization the F0 curves. The results showed that IF0 is an overall additive or subtractive effect throughout the vocalic segment, but it appears to be bigger in the later portion of the vocalic segment, or even near the end of the vocalic segment. This finding suggests that other than vowel midpoint, future studies may measure IF0 at other portions of the vocalic segment as well. However, the current study measured only nine data points per token. To better fit into the GAMM models, it would be better for future studies to measure more data points throughout the vocalic segment. Previous studies on tone merging used a multivariate analysis of variance (MANOVA) model (Hay et al., 2006; Fung & Lee, 2019) or predictive discriminant analysis (Tabachnick & Fidell, 2018; Mok et al., 2013). Different from previous studies, the current study used GAMM models to analyze whether two tones are different from each other in production. The advantage is that the change of F0 over time is taken into consideration and visualized, but this methodology also has a disadvantage. In the individual analyses, because there are very limited data points for each tone by speaker combination to fit into the GAMM models, the statistical power of the models is very low. Therefore, visual inspection of the GAMM figures must be used to decide whether two tones are confused or not in production. What's more, being acoustically significant different in production does not mean people are able to tell the difference in perception. Similarly, being acoustically similar in production does not mean people cannot tell the difference in perception. Therefore, other than analyzing the acoustic differences in production, future studies on tone merging may include a perception experiment in which participants judge whether two confusable tones sound different or not in AX tasks (Fung & Lee, 2019; Mok et al., 2013). Future studies may also recruit more participants and include more information in the language background questionnaires to obtain more data and investigate whether language experiences and contact play a role in IF0, as well as the relationship between IF0 and tone merging. In the current study, many data points were lost due to bad F0 tracking at the low pitch range, especially when the speakers showed creaky voice. Future studies may try to solve this technical constraint. Finally, this study included only one syllable structure in the stimuli, which is /s/ + vowel. More syllable structures could be included in future research to get more generalizable conclusions.

References

Audacity Team (2008). Audacity v. 2.4.2. Computer program, URL http://audacity.sourceforge.net.

Atkinson, J. E. (1972). Correlation analysis of the physiological features controlling fundamental voice frequency. *Journal of the Acoustical Society of America, 63*, 211-222.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120-144. https://doi.org/10.1016/j.wocn.2017.09.005

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Berlin/Boston De Gruyter Mouton.

Bauer, R. S., Cheung, K. H., & Cheung, P. M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change 15*:211-225.

Chao, Yuen Ren. (1930). A system of tone-letters. *Le Maître Phonétique* 45:24-27.

Chao, Yuen Ren. (1947). *Cantonese primer*. New York: Greenwood Press.

Chen, W. R., Whalen, D. H., & Tiede, M. K. (2021). A dual mechanism for intrinsic f0. *Journal of Phonetics*, *87*, 101063.

Ciocca, V., Whitehill, T. L., & Ng, S. S. (2002). Contour tone production by Cantonese speakers with cerebral palsy. *Journal of Medical Speech-Language Pathology, 10*(4), 243-248.

Connell, B. (2002). Tone languages and the universality of intrinsic F0: evidence from Africa. *Journal of Phonetics, 30*(1), 101-129.

Crystal, D. (1987). *The Cambridge encyclopedia of language* (Vol. 2). Cambridge: Cambridge University Press.

Diehl, R. (1991). The role of phonetics within the study of language. *Phonetica, 48*, 120-134.

Diehl, R. L., & Kluender, K. R. (1989a). On the objects of speech perception. *Ecological Psychology, 1,* 121-144.

Diehl, R. L., & Kluender, K. R. (1989b). Reply to commentators. *Ecological Psychology, 1*, 195-225.

Dyhr, N. (1990). The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. *Phonetica, 47*, 141-154.

Erickson, D., Honda, K., Hirai, H., & Beckman, M. E. (1995). The production of low tones in English intonation. *Journal of Phonetics*, *23*(1-2), 179-188.

Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica, 47*, 99-140.

Fok, C. Y. Y. (1974). *A perceptual study of tones in Cantonese (Occasional Papers and Monographs No. 18)*. Hong Kong: University of Hong Kong, Centre of Asian Studies.

Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *The Journal of the Acoustical Society of America, 146*(5), EL424-EL430. https://doi.org/10.1121/1.5133661

Gandour, J. & Weinberg, B. (1980). On the relationship between vowel height and fundamental frequency: evidence from esophageal speech. *Phonetica, 37*, 344-354.

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese*. Journal of Chinese Linguistics 9*:20-36.

Hallé, P. A. (1994) Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese, *Language and Speech, 37*, 103-123.

Hirose, H. & Gay, T. J. (1972). The activity of the intrinsic laryngeal muscles in voicing control: an electromyographic study. *Phonetica, 25*, 140-164.

Honda, K. (1983). Relation between pitch control and vowel articulation. *Haskins Laboratories Status Report on Speech Research, SR 73*, 269-282.

Honda, K., Hirai, H., & Dang, J. (1994). A physiological model of speech production and the implication of tongue-larynx interaction. In *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP 94),* (pp. 157-178)., Yokohama, Japan.

Hoole, P., Honda, K., Murano, E., Fuchs, S. & Pape, D. (2006). Go with the flow: Between automaticity and enhancement in control of segmental F0, In *Proceedings of the 7th International Seminar on Speech Production*, H.C Yehia, D. Demolin & R. Laboissière (Eds.), UFMG Belo Horizonte, pp. 357-360.

Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua, 134*, 62-74.

Kei, J., Smyth, V., So, L. K. H., Lau, C. C., & Capell, K. (2002). Assessing the accuracy of production of Cantonese lexical tones: A comparison between perceptual judgement and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing 7*:25-38.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics 35*:104-117.

Kingston, J. (1992). The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech, 35*, 99-113.

Kingston, J. (2007). Segmental influences on F0: Automatic or controlled? In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes. Vol. 2: Experimental Studies in Word and Sentence Prosody* (pp. 171-210). Berlin: Mouton de Gruyter.

Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language, 70*, 419-454.

Meyer, E. A. (1896-7). Zur Tonbewegung des Vokals im gesprochenen und gesungenen einzelwort. *Phonetische Studien* (Beiblatt zu der Zeitschrift Die Neuren Sprachen), *10*, 1-21.

Mok, P. P. K., Zuo, D., & Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. Language Variation and Change, 25(3), 341-370. https://doi.org/10.1017/S0954394513000161

Pettorino, M. (1987). Intrinsic pitch of vowels: an experimental study on Italian. In *Proceedings 11th International Congress of Phonetic Sciences, 1*, 138-141. Tallinn: Academy of Sciences of the Estonian SSR.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reinholt Petersen, N. (1986). Perceptual compensation for segmentally conditioned fundamental frequency perturbation. *Phonetica, 43*, 31-42.

Roubeau, B., Chevrie-Muller, C., & Saint Guily, J. L. (1972). *Electromyographic activity of strap and cricothyroid muscles in pitch change. Acta Otolaryngolical, 117*, 459-464.

Sapir, S. (1989). The intrinsic pitch of vowels: theoretical, physiological and clinical considerations. *Journal of Voice, 3*, 44-51.

Siddins, J., & Harrshibaington, J. (2015). Does vowel intrinsic F0 affect lexical tone?. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: the University of Glasgow.

So, L. K., & Varley, R. (1991). Cantonese Lexical Comprehension Test. *Hong Kong: Department of Speech and Hearing Sciences, University of Hong Kong*.

Sonderegger, M., McAuliffe, M., and Bang, H.-Y. (2017, June 12). Segmental influences on F0: a large-scale study of crosslinguistic and interspeaker variability. *Phonetics and Phonology in Europe 2017*, Cologne, Germany. http://people.linguistics.mcgill.ca/~morgan/wsc2017_slides.pdf

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086-1100.

Tabachnick, B. G., & Fidell, L. S. (2018). *Using multivariate statistics* (7th ed.). Pearson.

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America, 69,* 1465-1475.

Van Hoof, S., & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics, 39*(2), 168-177. https://doi.org/10.1016/j.wocn.2011.02.007

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica, 33*, 368-392.

Vance, T. J. (1977). Tonal distinctions in Cantonese*. Phonetica 34*:93-107.

Varley, R. & So, L. K. H. (1995). Age effects in tonal comprehension in Cantonese. *Journal of Chinese Linguistics 23*:76-97.

Whalen, D. H. & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics, 23*(3), 349-366.

Whalen, D. H., Levitt, A. G., Hsiao, P.-L., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants. *Journal of the Acoustical Society of America, 97*(4), 2533-2539.

Whitehill, T. L., Ciocca, V., & Chow, T. Y. (2000). Acoustic analysis of lexical tone contrast in dysarthria. *Journal of Medical Speech-Language Pathology, 8,* 337–344.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

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.

Wood, S.N. (2017). Generalized Additive Models: An Introduction with R, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Yiu, C. Y. (2009). A preliminary study on the change of rising tones in Hong Kong Cantonese: An experimental study (in Chinese). *Language and Linguistics 10*:269-291.

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, *136*(3), 1320-1333.

Zee, E. (1999). Chinese (Hong Kong Cantonese). *Handbook of the International Phonetic Association* (pp. 58-60). Cambridge: Cambridge University.

Appendix A

Statistical Outputs

1. Confusable tone pairs

Older group, T2 and T5

	Estimate	SE	*t*	*p*	r^2
(Intercept)	169.152	4.910	34.450	< 0.001	0.829
T5	-12.462	0.745	-16.720	< 0.001	0.829

Older group, T3 and T6

	Estimate	SE	t	p	r2
(Intercept)	172.664	5.416	31.880	< 0.001	0.877
T6	-12.632	0.511	-24.740	< 0.001	0.877

Older group, T4 and T6

	Estimate	SE	t	p	r2
(Intercept)	135.241	4.921	27.480	< 0.001	0.760
T6	25.260	0.802	31.480	< 0.001	0.760

Younger group, T2 and T5

	Estimate	SE	t	p	r2
(Intercept)	186.198	3.224	57.750	< 0.001	0.551
T5	-4.888	0.938	-5.210	< 0.001	0.551

Younger group, T3 and T6

	Estimate	SE	t	p	r2
(Intercept)	189.436	3.663	51.721	< 0.001	0.657
T6	-3.743	0.683	-5.482	< 0.001	0.657

Younger group, T4 and T6

	Estimate	SE	t	p	r2
(Intercept)	159.895	4.156	38.470	< 0.001	0.527
T6	26.869	1.415	18.980	< 0.001	0.527

2. High vowel and low vowel of each tone

Older group, T1

	Estimate	SE	t	p	r2
(Intercept)	226.699	11.852	19.128	< 0.001	0.856
VowelHeightMid	-5.263	16.407	-0.321	0.748	0.856

Older group, T2

	Estimate	SE	t	p	r2
(Intercept)	178.764	7.712	23.180	< 0.001	0.873
VowelHeightMid	-19.400	10.598	-1.830	0.067	0.873

Older group, T3

	Estimate	SE	t	p	r2
(Intercept)	174.455	8.810	19.803	< 0.001	0.914
VowelHeightMid	-3.333	11.929	-0.279	0.780	0.914

Older group, T4

	Estimate	SE	t	p	r2
(Intercept)	141.290	6.487	21.780	< 0.001	0.684
VowelHeightMid	-11.248	10.045	-1.120	0.263	0.684

Older group, T5

	Estimate	SE	t	p	r2
(Intercept)	159.857	6.934	23.053	< 0.001	0.851
VowelHeightMid	-4.854	9.277	-0.523	0.601	0.851

Older group, T6

	Estimate	SE	t	p	r2
(Intercept)	161.975	7.820	20.712	< 0.001	0.848
VowelHeightMid	-3.398	10.531	-0.323	0.747	0.848

Younger group, T1

	Estimate	SE	t	p	r2
(Intercept)	222.766	8.763	25.422	< 0.001	0.730
VowelHeightMid	-2.001	11.900	-0.168	0.867	0.730

Younger group, T2

	Estimate	SE	t	p	r2
(Intercept)	194.560	4.869	39.958	< 0.001	0.572
VowelHeightMid	-15.786	6.275	-2.516	0.012	0.572

Younger group, T3

	Estimate	SE	t	p	r2
(Intercept)	189.101	6.045	31.284	< 0.001	0.756
VowelHeightMid	-0.324	7.945	-0.041	0.967	0.756

Younger group, T4

	Estimate	SE	t	p	r2
(Intercept)	162.925	8.015	20.328	<0.001	0.556
VowelHeightMid	-1.976	11.175	-0177	0.860	0.556

Younger group, T5

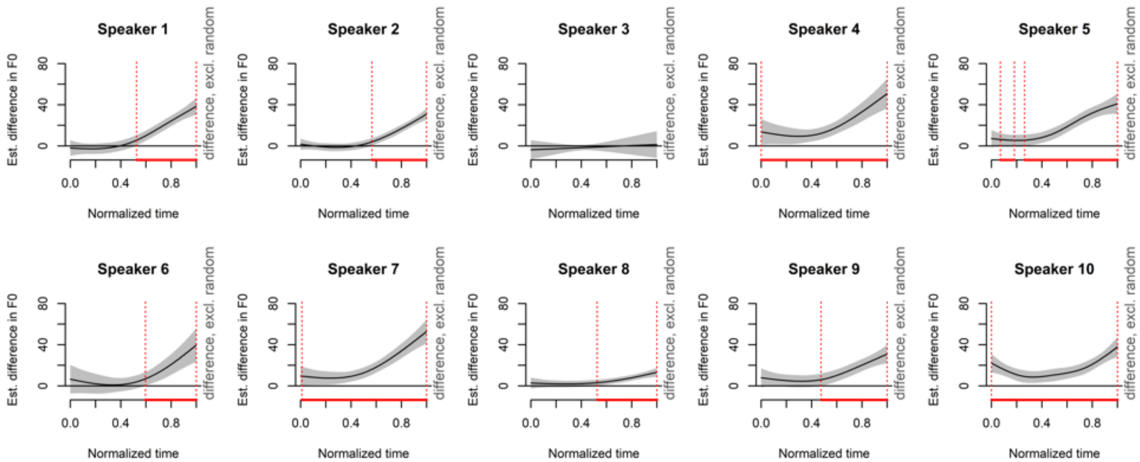
	Estimate	SE	t	p	r2
(Intercept)	184.327	4.383	42.053	< 0.001	0.571
VowelHeightMid	-4.398	6.189	-0.711	0.477	0.571

Younger group, T6

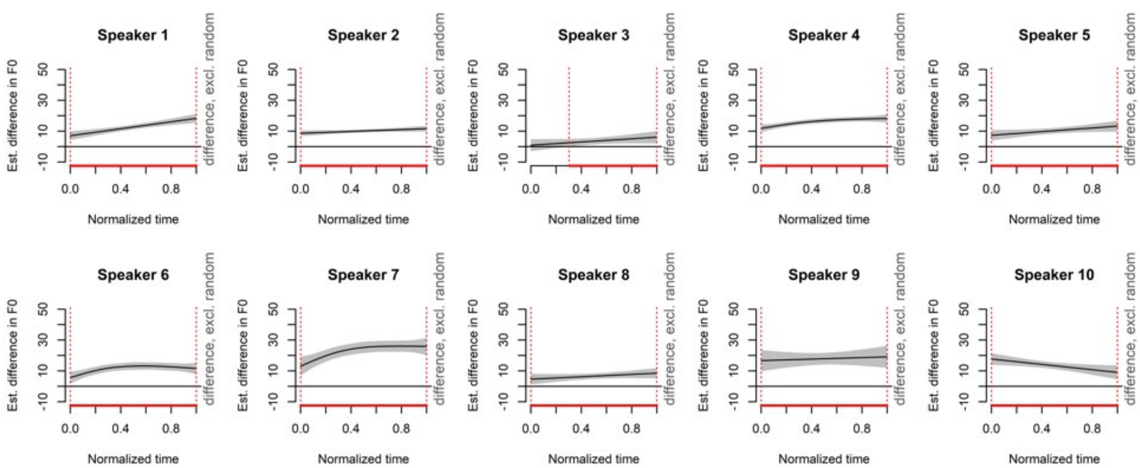
	Estimate	SE	t	p	r2
(Intercept)	184.500	5.060	36.460	< 0.001	0.652
VowelHeightMid	2.980	7.598	0.392	0.695	0.652

Appendix B

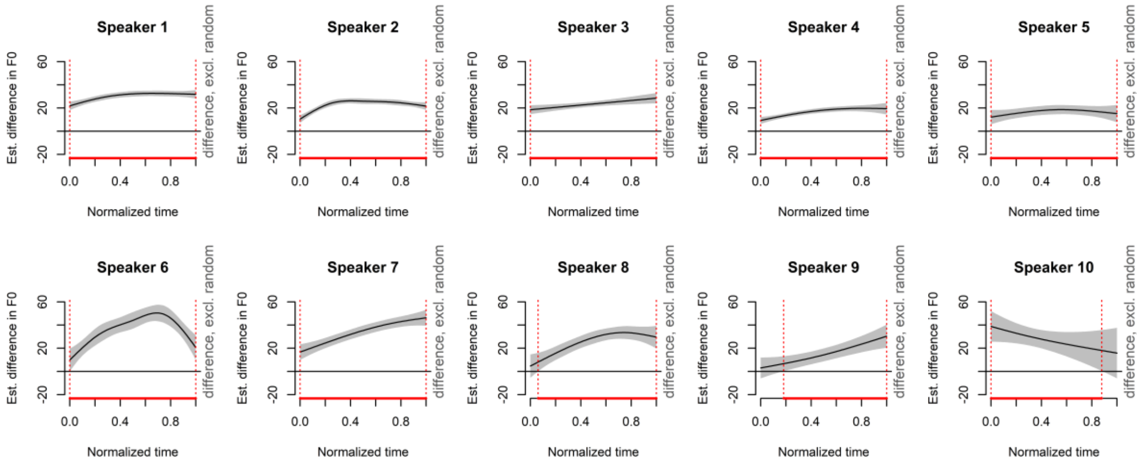
Figures of GAMM difference smooths of confusable tone pairs



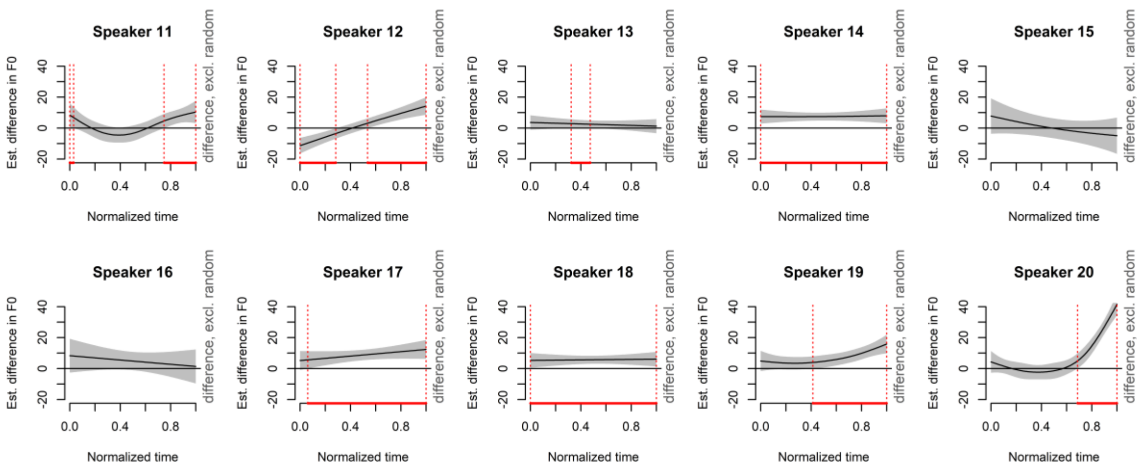
Difference smooths of T2 and T5 (T2 – T5) produced by each participant of the older group



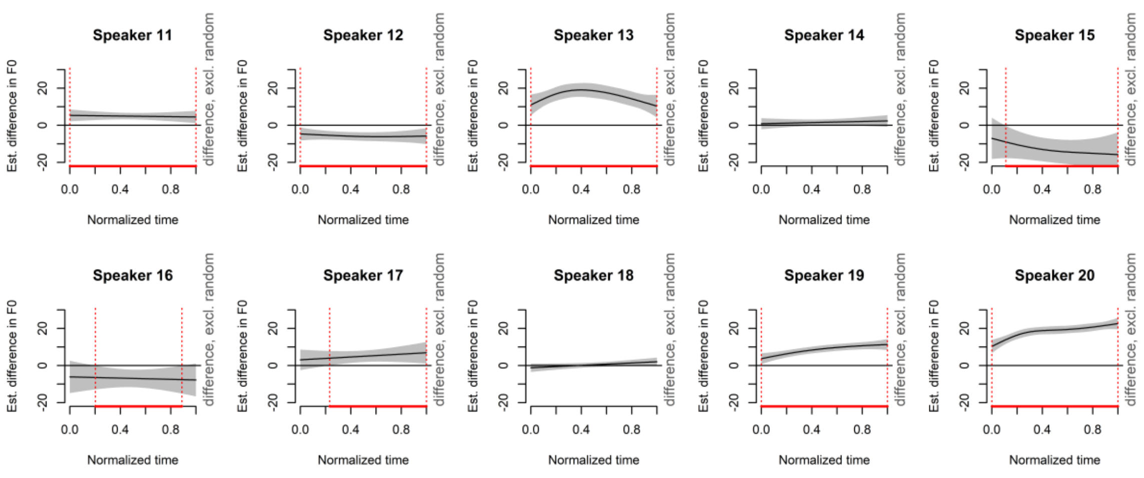
Difference smooths of T3 and T6 (T3 – T6) produced by each participant of the older group



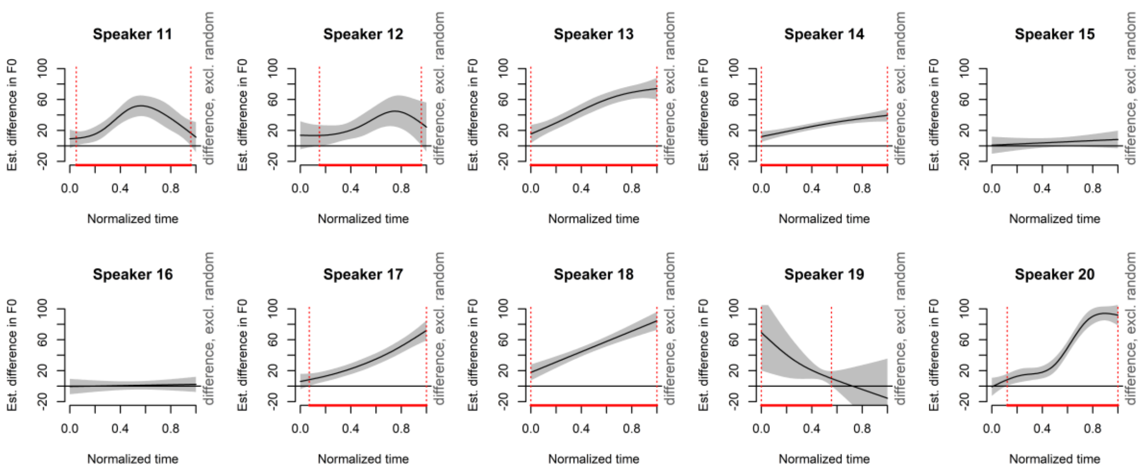
Difference smooths of T4 and T6 (T6 – T4) produced by each participant of the older group



Difference smooths of T2 and T5 (T2 – T5) produced by each participant of the younger group



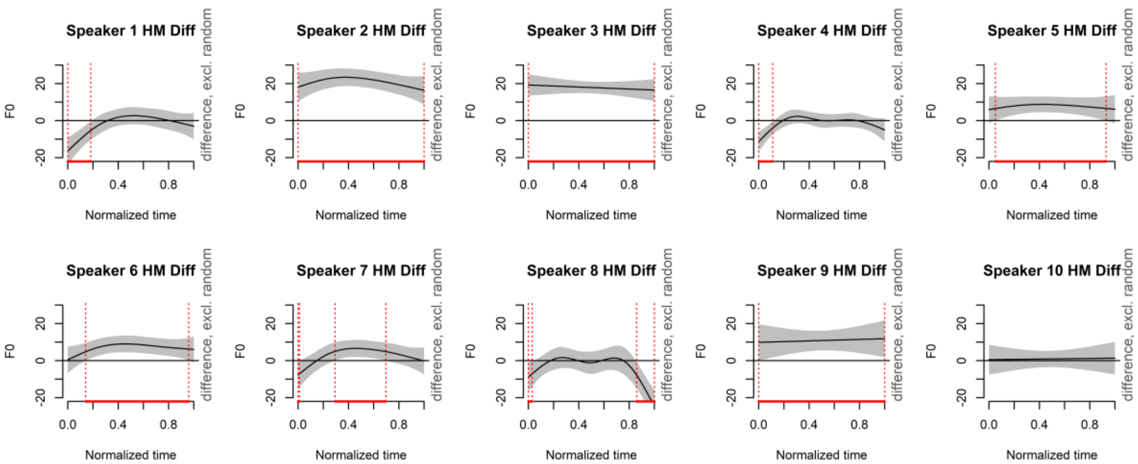
Difference smooths of T3 and T6 (T3 – T6) produced by each participant of the younger group



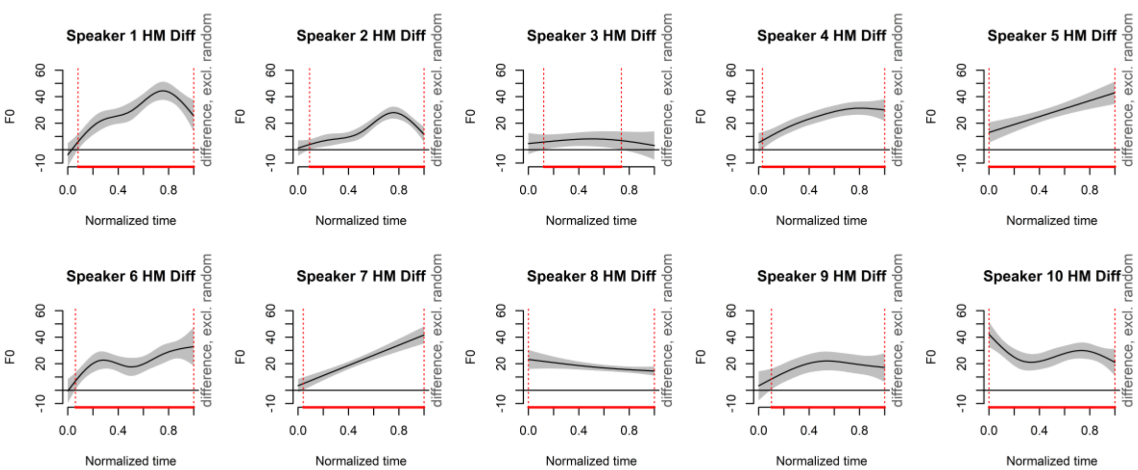
Difference smooths of T4 and T6 (T6 – T4) produced by each participant of the younger group

Appendix C

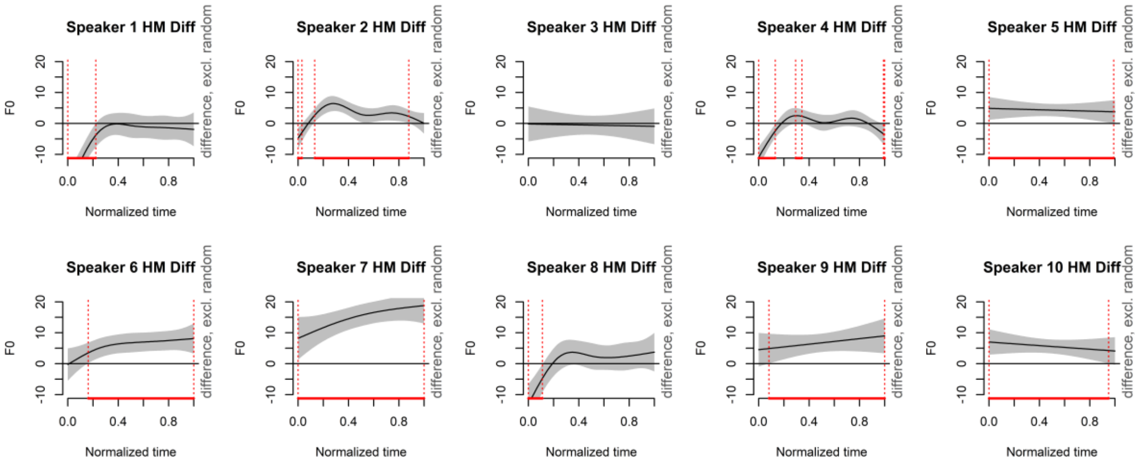
Figures of GAMM difference smooths of high vowels and mid vowels carrying each tone



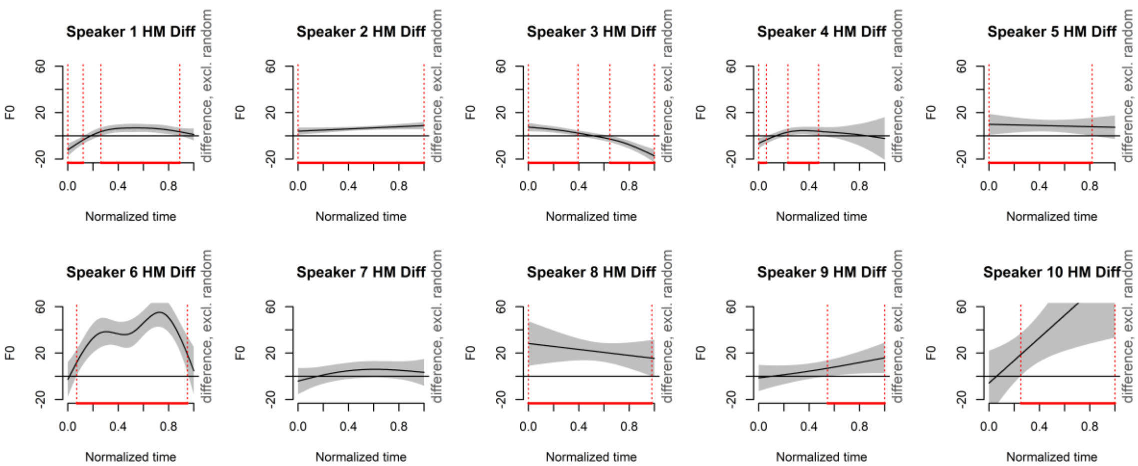
Difference smooths of high vowels and mid vowels (high – mid) carrying T1 produced by each participant of the older group



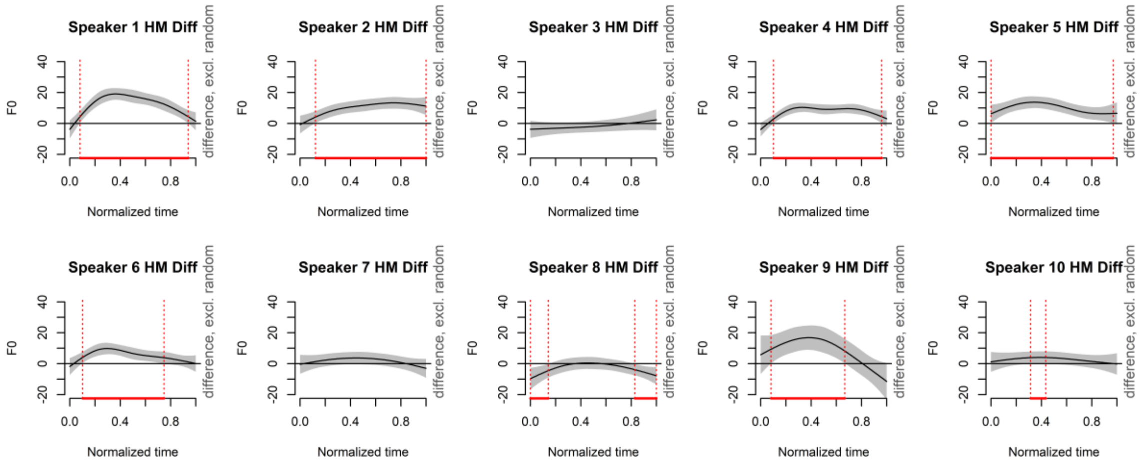
Difference smooths of high vowels and mid vowels (high – mid) carrying T2 produced by each participant of the older group



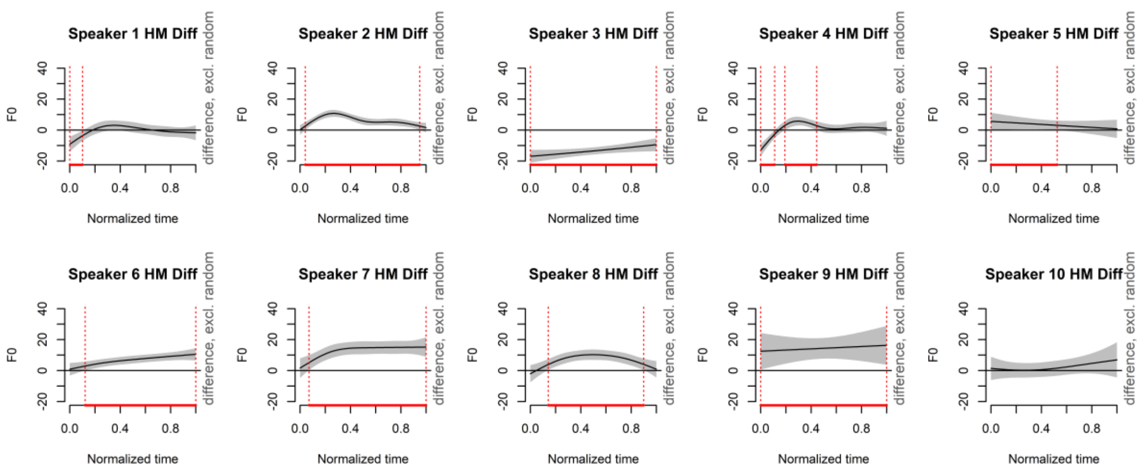
Difference smooths of high vowels and mid vowels (high – mid) carrying T3 produced by each participant of the older group



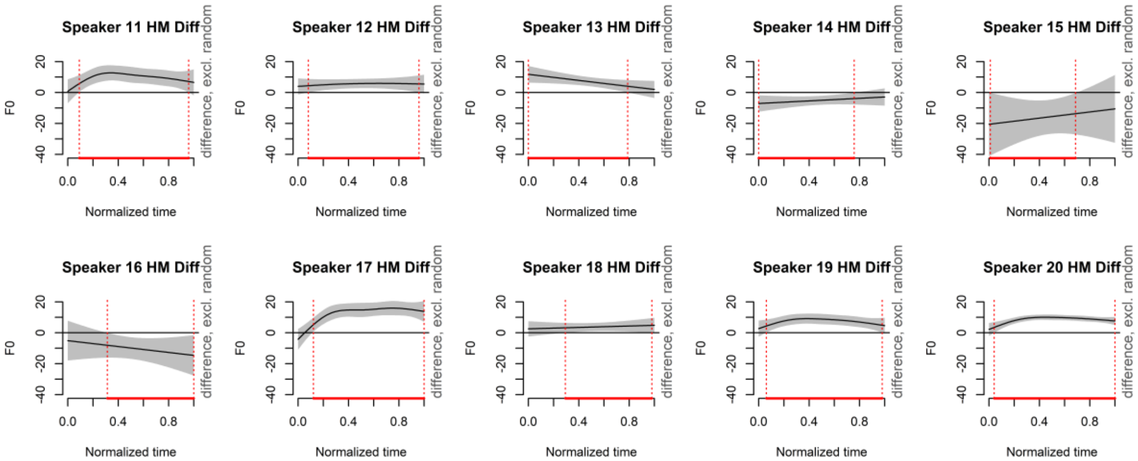
Difference smooths of high vowels and mid vowels (high – mid) carrying T4 produced by each participant of the older group



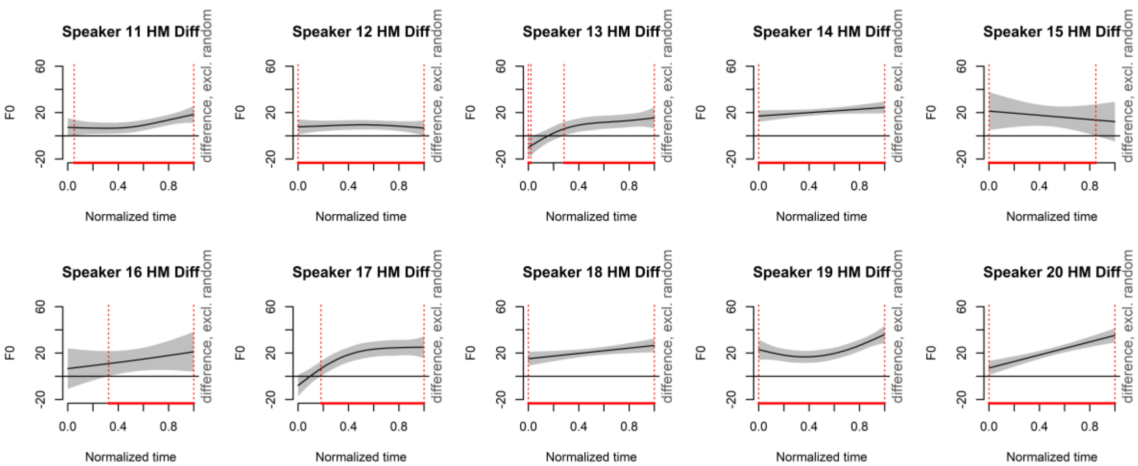
Difference smooths of high vowels and mid vowels (high – mid) carrying T5 produced by each participant of the older group



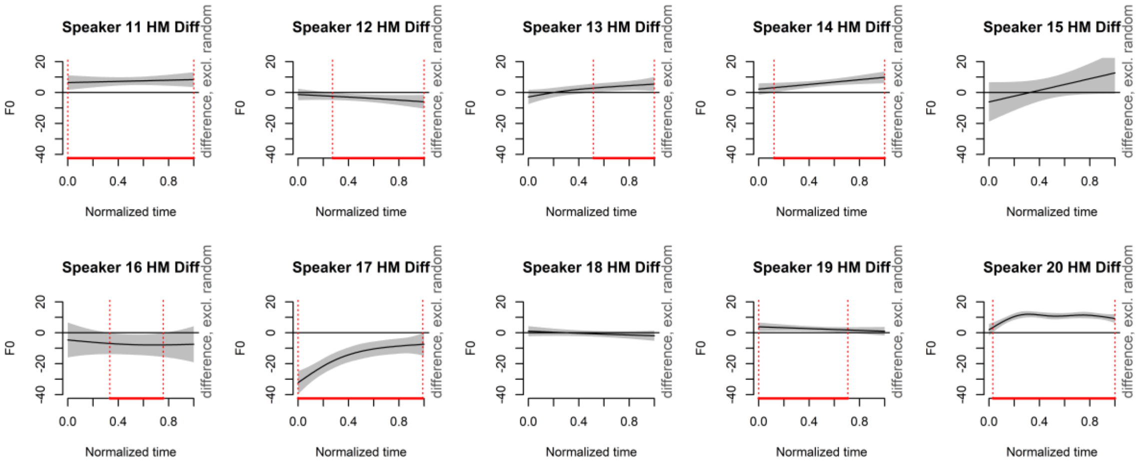
Difference smooths of high vowels and mid vowels (high – mid) carrying T6 produced by each participant of the older group



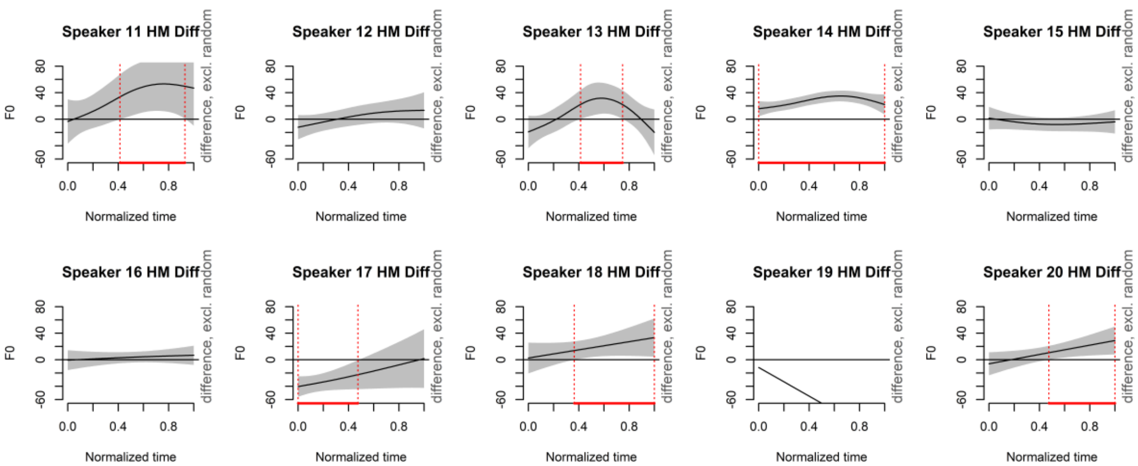
Difference smooths of high vowels and mid vowels (high – mid) carrying T1 produced by each participant of the younger group



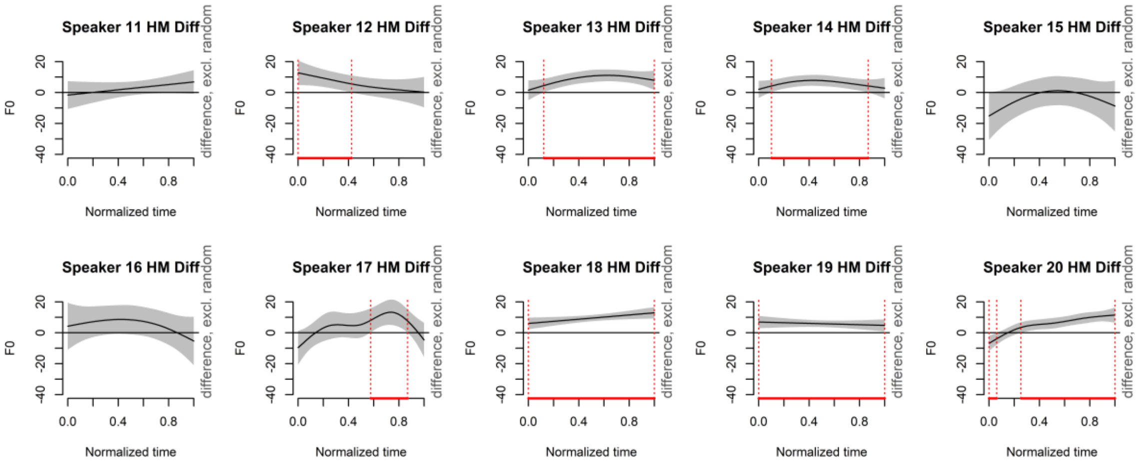
Difference smooths of high vowels and mid vowels (high – mid) carrying T2 produced by each participant of the younger group



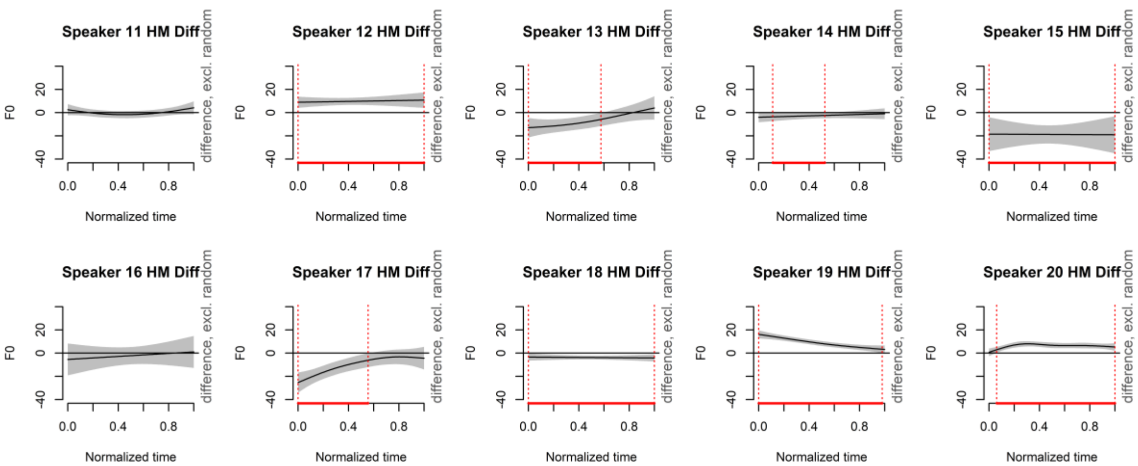
Difference smooths of high vowels and mid vowels (high – mid) carrying T3 produced by each participant of the younger group



Difference smooths of high vowels and mid vowels (high – mid) carrying T4 produced by each participant of the younger group



Difference smooths of high vowels and mid vowels (high – mid) carrying T5 produced by each participant of the younger group



Difference smooths of high vowels and mid vowels (high – mid) carrying T6 produced by each participant of the younger group

Appendix D

Supplementary Data File

Description:

The accompanying Excel spreadsheet shows the F0 values extracted using Praat from the audio recordings. In the column headings: “Subject” represents the subject ID; “Group” represents the age group each subject belongs to (old or young); “Tone” represents the tone of each stimulus word (1-6); “Vowel Height” represents the level of vowel height (high or mid) of each stimulus word; “Repetition” represents the number of repetition (1-10) of each trial; “Point” represents the time point from which each F0 value was extracted (1-9); “F0” shows the F0 value of each data point, which was measured in Hz.

Filename:

Data_Bingqing_MasterThesis.csv