

CONSONANTAL EFFECTS ON F0 IN TONAL LANGUAGES

By

Qian Luo

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Linguistics—Doctor of Philosophy

2018

## ABSTRACT

### CONSONANTAL EFFECTS ON F0 IN TONAL LANGUAGES

By

Qian Luo

Consonant types can influence F0 values of the adjacent vowels (this F0 perturbation effect is henceforth referred as C-F0). C-F0 may be enlarged to maximize perceptual distinctiveness and thus reinforce aspiration contrast. In tonal languages, the effect may also be inhibited to cue tone contrast by constraining F0 variability by the demands of the lexical tone system. This dissertation investigates how C-F0 can be related to tone contrast and aspiration contrast by asking the following questions: (1) Is C-F0 conditioned by lexical tones? (2) Is C-F0 conditioned by F0 difference between tones? (3) Is there a cue trading relation between F0 and VOT for aspiration contrast?

Fifteen Mandarin speakers and fifteen Cantonese speakers participated in the production experiments. All stimuli followed a CV template. The Mandarin stimuli had four Mandarin tones: M-T1(55), M-T2(35), M-T3(214) and M-T4(51). The Cantonese stimuli covered six Cantonese tones: C-T1(55), C-T2(25), C-T3(33), C-T4(21), C-T5(23) and C-T6(22). Initial consonants were aspirated obstruents, unaspirated obstruents and sonorants. F0 values following sonorants were the baseline for evaluating C-F0.

The major results are as follow: for Question 1, the major findings are that tones can influence the vowel duration that C-F0 can extend and the difference between F0 following different consonants in Mandarin and Cantonese. A consistent direction was found: F0 following aspirated stops was higher than F0 following unaspirated stops and following sonorants. The trajectory of F0 following aspirated stops started as the second highest and converged with the

lowest baseline F0 following sonorants. The results indicate a robust aspiration raising and a weaker voiceless unaspirated raising effect in both Mandarin and Cantonese.

For Question 2, the pattern of the vowel duration that C-F0 can extend was found to follow the pattern of the F0 difference between the target tone and its closest tone in the tone inventory. However, the difference between F0 following different consonants within the first 10ms did not follow the patterns of F0 difference between tones in Mandarin or Cantonese. Finally, the cross-linguistic comparison provides support to the hypothesis that higher degree of tone competition may restrict C-F0.

For Question 3, the results show that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue. However, the findings do not provide evidence for a cue trading relation between VOT and onset F0 in Mandarin or Cantonese.

This dissertation has offered a new angle that few previous studies have dealt with before: traditionally, the enhancement of voicing contrast or the physical properties of producing voicing were proposed as the major trigger that give rise to C-F0. This dissertation has introduced a new contrast, i.e. the contrast of lexical tones, to explore the question. Furthermore, the major findings for all three questions confirm a hybrid account for explaining C-F0 in tonal languages: the tone enhancement account and the consonant automatic account.

Copyright by  
QIAN LUO  
2018

This dissertation is dedicated to my baby girls:  
Goda and Ella

## ACKNOWLEDGEMENTS

I would like to thank my advisors, Yen-Hwei Lin and Karthik Durvasula, for showing me the beauty of phonology and phonetics and leading me into a desire of learning and knowledge. They are not only my academic advisors, but also my family during my journey in the US. I would also thank my committee members, Suzanne Wagner, Alan Beretta and Anne Violin-Wigent, for their support and valuable inputs for my dissertation.

My big thanks go to James Kirby, from whom I learned the beauty of statistical models for linguistics. I would also like to extend my gratitude to the following great linguists, who have fulfilled my curiosity and inquisitive appetite: Yiya Chen, John Coleman, Christian DiCanio, Larry Hyman, Bob Ladd, Alan Yu and Jie Zhang.

I would also like to thank my family and friends, who have always been supporting me: my parents (罗双跃 and 崔海燕), Guang, and Giedrius. A ‘hello’ to my baby girls, Goda and Ella, if they will read this dissertation in the future. The world is awesome. May you always be curious to ‘explore it, poke at it, question it, and turn it inside out’.

# TABLE OF CONTENTS

LIST OF TABLES.....	x
LIST OF FIGURES.....	xiv
1. Introduction.....	1
1.1 An introduction to consonantal effects on F0 (C-F0).....	1
1.1.1 Consistent consonantal effects on F0.....	2
1.1.2 Inconsistent consonantal effects on F0.....	4
1.1.3 Effects in non-tonal languages vs. tonal languages.....	5
1.2 The debate between the enhancement account and the automatic account.....	6
1.2.1 The enhancement account.....	8
1.2.1.1 The feature enhancement hypothesis.....	8
1.2.1.2 The probabilistic enhancement hypothesis.....	9
1.2.2 The automatic account.....	10
1.2.2.1 Articulation maneuvers.....	10
1.2.2.2 Aerodynamic conditions.....	11
1.3 Research questions.....	13
1.3.1 Question 1: Is C-F0 conditioned by tones?.....	13
1.3.2 Question 2: Is C-F0 conditioned by F0 difference between tones?.....	16
1.3.3 Question 3: cue trading?.....	18
1.4 An overview of the dissertation.....	21
2. Language background: Mandarin and Cantonese.....	23
2.1 Introduction.....	23
2.2 Mandarin.....	24
2.2.1 Mandarin consonants.....	24
2.2.2 Mandarin tones.....	26
2.3 Cantonese.....	30
2.3.1 Cantonese consonants.....	30
2.3.2 Cantonese tones.....	31
3. Methods.....	35
3.1 Introduction.....	35
3.2 Mandarin group.....	35
3.2.1 Participants of native Mandarin.....	35
3.2.2 Mandarin stimuli.....	36
3.2.3 Procedure for the Mandarin experiment.....	37
3.3 Cantonese group.....	37
3.3.1 Participants of Cantonese.....	37
3.3.2 Cantonese stimuli.....	38
3.3.3 Procedure for the Cantonese experiment.....	39
3.4 Measurements.....	40

3.4.1	Segmentation labeling and measurements of durations.....	40
3.4.2	Measurements of F0.....	42
3.5	Statistic analysis.....	43
4.	Results for Question 1: C-F0 conditioned by tones? .....	44
4.1	Introduction.....	44
4.2	C-F0 conditioned by tones in Mandarin?.....	45
4.2.1	Overall Mandarin tone trajectories.....	45
4.2.2	C-F0 in Mandarin tones.....	48
4.2.2.1	General results.....	48
4.2.2.2	C-F0 in M-T1(55).....	50
4.2.2.3	C-F0 in M-T2(35).....	53
4.2.2.4	C-F0 in M-T3(214).....	54
4.2.2.5	C-F0 in M-T4(51).....	57
4.2.3	Summary of results-Q1 for Mandarin.....	59
4.3	C-F0 conditioned by tones in Cantonese?.....	60
4.3.1	Overall Cantonese tone trajectories.....	60
4.3.2	C-F0 in Cantonese tones.....	63
4.3.2.1	General results.....	63
4.3.2.2	C-F0 in C-T1(55).....	65
4.3.2.3	C-F0 in C-T2(25) and C-T5(23).....	67
4.3.2.4	C-F0 in C-T3(33), C-T6(22) and C-T4(21).....	70
4.3.3	Summary of results-Q1 for Cantonese.....	73
4.4	Overall summary for results-Q1.....	74
5.	Results for Question 2: C-F0 conditioned by tone distance? .....	76
5.1	Introduction.....	76
5.2	Intra-linguistic comparison in Mandarin.....	78
5.2.1	Tone distance in Mandarin.....	78
5.2.1.1	Smallest TONE DIFF in Mandarin.....	78
5.2.1.2	Overall TONE DIFF in Mandarin.....	80
5.2.2	The magnitudes of C-F0 in Mandarin tones.....	81
5.2.3	C-F0 and Mandarin tone distance.....	83
5.2.4	Summary of intra-linguistic comparison in Mandarin.....	85
5.3	Intra-linguistic comparison in Cantonese.....	86
5.3.1	Tone distance in Cantonese.....	86
5.3.1.1	Smallest TONE DIFF in Cantonese.....	86
5.3.1.2	Overall TONE DIFF in Cantonese.....	88
5.3.2	C-F0 difference in Cantonese tones.....	89
5.3.3	C-F0 and Cantonese tone distance.....	91
5.3.4	Summary of intra-linguistic comparison in Cantonese.....	92
5.4	Cross-linguistic comparison.....	93
5.5	Overall summary for results-Q2.....	95
6.	Results for Question 3: cue trading between VOT and F0? .....	97
6.1	Introduction.....	97



6.2 Cue trading between VOT and F0 in Mandarin? .....	98
6.2.1 VOT and F0 as cues for the aspiration contrast in Mandarin tone contexts.....	98
6.2.1.1 VOT as a strong cue for the aspiration contrast in Mandarin.....	98
6.2.1.2 Onset F0 as a weak cue for the aspiration contrast in Mandarin.....	100
6.2.2 VOT and F0 for M-T1(55).....	103
6.2.3 VOT and F0 for M-T2(35).....	104
6.2.4 VOT and F0 for M-T3(214).....	105
6.2.5 VOT and F0 for M-T4(51).....	106
6.2.6 Summary of results-Q3 for Mandarin.....	107
6.3 Cue trading between VOT and F0 in Cantonese?.....	108
6.3.1 VOT and F0 as cues for the aspiration contrast in Cantonese tone contexts.....	108
6.3.1.1 VOT as a strong cue for the aspiration contrast in Cantonese.....	108
6.3.1.2 Onset F0 as a weak cue for the aspiration contrast in Cantonese.....	110
6.3.2 VOT and F0 for C-T1(55).....	111
6.3.3 VOT and F0 for C-T2(25).....	112
6.3.4 VOT and F0 for C-T3(33).....	113
6.3.5 Summary of results-Q3 for Cantonese.....	114
6.4 Overall summary of results-Q3.....	115
7. General discussion and conclusion.....	117
7.1 Summary of major findings and discussion.....	117
7.1.1 Discussion: C-F0 can be conditioned by lexical tones.....	118
7.1.1.1 C-F0 is generally stronger in tones with high onset pitch.....	118
7.1.1.2 C-F0 DIRECTION is generally consistent across tones.....	119
7.1.2 Discussion: C-F0 can be restricted by competition of tones.....	121
7.1.3 Discussion: C-F0 and cue trading for consonant laryngeal contrast.....	123
7.2 Implications for the enhancement account and the automatic account for C-F0.....	126
7.3 Conclusion.....	130
APPENDICES.....	132
APPENDIX A Participant information.....	133
APPENDIX B Mandarin stimuli.....	134
APPENDIX C Cantonese stimuli.....	135
BIBLIOGRAPHY.....	136

## LIST OF TABLES

Table 1. Example languages where the voicing effect has been observed (adapted from Kingston & Diehl (1994)).....	2
Table 2. Aspiration and sonorant effects across languages.....	4
Table 3. The Consonant Inventory of Mandarin. For each column, the segments on the left of the dash lines are voiceless aspirated, and the ones on the right are voiceless unaspirated or voiced.....	25
Table 4. The Tone Inventory of Mandarin.....	26
Table 5. The Consonant Inventory of Cantonese. For each column, the segments on the left of the dash lines are voiceless aspirated, and the ones on the right are voiceless unaspirated or voiced.....	31
Table 6. The Tone Inventory of Cantonese.....	32
Table 7. Mean F0 values(Hz) and the standard deviation at every 10% of the vowel. The lower portion of the table includes results of ANOVAs at each time point where the dependent variable was the mean F0, and the within-subjects independent variable was TONE (4 tones).....	47
Table 8. T-tests for Tone Pair M-T1(55) vs. M-T4(51) and M-T2(35) vs. M-T3(214) at every 10% of the vowel.....	48
Table 9. ANOVA where the independent variables are CONSONANT, TONE and VOWEL and the dependent variable is normalized F0 values in Mandarin. T = Tone; C= Consonant; V = Vowel.....	50
Table 10. In the context of M-T1(55), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded. T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant.....	52
Table 11. In the context of M-T2(35), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values.....	54
Table 12. In the context of M-T3(214), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded. T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant.....	56

Table 13. In the context of M-T4(51), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded. T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant.....	58
Table 14. Summary of Results-Q1 for Mandarin.....	59
Table 15. Mean F0 values(Hz) and the standard deviation at every 10% of the vowel. The lower portion of the table includes results of ANOVAs at each time point where the dependent variable was the mean F0, and the within-subjects independent variable was TONE (6 tones).....	62
Table 16. T-tests for Tone Pair C-T2(25) vs. C-T5(23), C-T3(33) vs. C-T5(23), C-T3(33) vs. C-T6(22) and C-T5(23) vs. C-T6(22) at every 10% of the vowel.....	63
Table 17. ANOVA tests for effects from the independent variables CONSONANT, TONE and VOWEL on the dependent variable normalized F0 values in Cantonese. T = Tone; C= Consonant; V = Vowel.....	64
Table 18. In the context of C-T1(55), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded. T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant.....	66
Table 19. In the context of C-T2(25), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded. T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant.....	69
Table 20. In the context of C-T5(23), t-tests at every 5ms time point for normalized F0 after aspirated obstruents and sonorants. Significant results in t-tests are shaded. TH = Voiceless Aspirated Stop; N = Sonorant.....	70
Table 21. In the context of C-T3(33), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded. T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant.....	72
Table 22. In the context of C-T6(22), t-tests at every 5ms time point for normalized F0 after unaspirated obstruents and sonorants. Significant results in t-tests are shaded. T = Voiceless Unaspirated Stop; N = Sonorant.....	72
Table 23. In the context of C-T4(21), t-tests at every 5ms time point for normalized F0 after aspirated obstruents and sonorants. Significant results in t-tests are shaded. TH = Voiceless	

Aspirated	Stop;	N	=
Sonorant.....			73
Table 24. Summary of Results-Q1 for Cantonese.....			73
Table 25. T-tests comparing the Overall TONE DIFF of each Mandarin tone (within the first 10ms).....			81
Table 26. Summary of C-F0 LENGTH results in Question1 for Mandarin. The values of maximum C-F0 LENGTH is the timepoint after which C-F0 becomes non-significant.....			83
Table 27. T-tests comparing C-F0 DIFF(TH-N) in each Mandarin tone (within the first 10ms).....			83
Table 28. Linear Regression tests for the relation between C-F0 DIFF(TH-N) and Smallest TONE DIFF in Mandarin (within the first 10ms).....			84
Table 29. T-tests comparing the Overall TONE DIFF of each Cantonese tone (within the first 10ms).....			89
Table 30. Summary of C-F0 LENGTH results in Question1 for Cantonese. The values of maximum C-F0 LENGTH is the timepoint after which C-F0 becomes non-significant.....			90
Table 31. T-tests comparing C-F0 DIFF(TH-N) in each Mandarin tone (within the first 10ms).....			91
Table 32. Linear Regression tests for the relation between C-F0 DIFF(TH-N) and Smallest TONE DIFF in Cantonese (within the first 10ms).....			92
Table 33. Summary of C-F0 LENGTH results in Question1 for Cantonese and Mandarin. The values of maximum C-F0 LENGTH is the timepoint after which C-F0 becomes non-significant.....			94
Table 34. ANOVA test for average VOT values of aspirated and unaspirated obstruents across Mandarin tones.....			99
Table 35. T-tests comparing normalized VOT of ASPIRATED obstruents in different tone contexts.....			100
Table 36. ANOVA test for average F0 values (the first 10ms) after aspirated and unaspirated obstruents across Mandarin tones.....			101
Table 37. T-tests comparing average F0 values (the first 10ms) after aspirated and unaspirated obstruents in each Mandarin tone.....			102
Table 38. A summary of the linear regression results for the relation between VOT and F0 in four Mandarin tones.....			108
Table 39. ANOVA test for average VOT values of aspirated and unaspirated obstruents across Cantonese tones.....			110

Table 40. ANOVA test for average F0 values (the first 10ms) after aspirated and unaspirated obstruents across Cantonese.....	111
Table 41. T-tests comparing average F0 values (the first 10ms) after aspirated and unaspirated obstruents in each Cantonese tone.....	111
Table 42. A summary of the linear regression results for the relation between VOT and F0 in three Cantonese tones.....	115
Table 43. A summary of predictions by the enhancement account and the automatic account respectively and the major findings.....	130
Table 44. Participant information .....	133
Table 45. Mandarin stimuli.....	134
Table 46. Cantonese stimuli.....	135

## LIST OF FIGURES

Figure 1. Average F0 contours of syllable /ta/ versus /t <sup>h</sup> a/ in four tones. The right boundary of each panel represents syllable offset. (from Xu & Xu (2003)).....	14
Figure 2. Mean F0 (in cents) for syllables beginning with aspirated and unaspirated bilabial stops and having C-T1(55), C-T2(35) or C-T4(21). Error bars indicate standard error of the mean. (from Francis et al. (2006)).....	15
Figure 3. Acoustic tone space at onset and offglide in Cantonese, Thai and Mandarin in semitones. Tone spaces are illustrated as plots of F0 offglide (mean F0 at the tenth equidistant timepoint) x F0 onset (mean F0 at the second equidistant timepoint k1). L=Low Tone; LR=Low-Rising Tone; LF=Low-Falling Tone; M=Mid Tone; MR=Mid-Rising Tone; H=High Tone; FR=Falling-Rising Tone. (from Alexander (2010: 122-123)).....	17
Figure 4. Relation between F0 and VOT coefficients in production (a) and perception (b) (from Shultz et al. (2012)).....	19
Figure 5. Average F0 values in Mandarin Tone contours within normalized vowels. The 11 <sup>th</sup> end point is excluded due to the large amount of undefined measurement from Praat. The data were produced by 15 Mandarin female speakers.....	28
Figure 6. Average F0 values in Cantonese Tone contours within normalized vowels. The 11 <sup>th</sup> end point is excluded due to the large amount of undefined measurement from Praat. The data were produced by 15 Cantonese female speakers.....	33
Figure 7. Segmentation for p <sup>h</sup> a1(55) in Mandarin. ‘c’ = closure; ‘r’ = release; ‘a’=the vowel /a/.....	41
Figure 8. Segmentation for pa1(55) in Mandarin. ‘c’ = closure; ‘r’ = release; ‘a’=the vowel /a/.....	41
Figure 9. segmentation for ma1(55) in Mandarin. ‘m’ = nasal /m/; ‘a’=the vowel /a/.....	42
Figure 10. Four Mandarin tone contours within vowels in normalized duration.....	46
Figure 11. Normalized F0 trajectories within the first 50ms of the vowels in four Mandarin tones. Shading shows 1 SD around the mean.....	49
Figure 12. Normalized F0 within the first 35ms of the vowel in M-T1(55). Shading shows 1 SD around the mean value.....	51
Figure 13. Normalized F0 within the first 35ms of the vowel in M-T2(35). Shading shows 1 SD around the mean.....	53

Figure 14. Normalized F0 trajectories within the first 50ms of the vowel in M-T3(214). Shading shows 1 SD around the mean.....	55
Figure 15. Normalized F0 trajectories within the first 50ms of the vowels in M-T4(51). Shading shows 1 SD around the mean.....	57
Figure 16. Six Cantonese tone contours within normalized vowels.....	61
Figure 17. Normalized F0 trajectories within the first 50ms of the vowels in six Cantonese tones. Shading shows 1 SD around the mean.....	64
Figure 18. Normalized F0 trajectories within the first 35ms of the vowels in C-T1(55). Shading shows 1 SD around the mean.....	65
Figure 19. Normalized F0 trajectories within the first 35ms of the vowel in C-T2(25). Shading shows 1 SD around the mean.....	67
Figure 20. Normalized F0 trajectories within the first 35ms of the vowel in C-T5(23). Shading shows 1 SD around the mean.....	68
Figure 21. Normalized F0 trajectories within the first 35ms of the vowels in C-T3(33) and C-T6(22). Shading shows 1 SD around the mean.....	70
Figure 22. Normalized F0 trajectories within the first 35ms of the vowels in C-T4(21). Shading shows 1 SD around the mean.....	71
Figure 23. Scaled F0 difference in six Mandarin tone pairs (within the first 10ms). Each tick on the x-axis separates the names of the tones. E.g. ‘MT1_MT2’ represents the tone pair M-T1(55) vs. M-T2(35).....	79
Figure 24. Overall TONE DIFF: the average values of all scaled F0 differences between a target tone and all other tones in Mandarin (within the first 10ms). E.g. Overall TONE DIFF (M-T1) = (TONE DIFF[M-T1 vs. M-T2) +TONE DIFF(M-T1 vs. M-T3)+TONE DIFF(M-T1 vs. M-T4)) ÷3.....	81
Figure 25. Average C-F0 DIFF in each Mandarin tone (within the first 10ms). The red boxes represent the difference between F0 values after unaspirated obstruents and after sonorants. The blue boxes represent the difference between F0 values after aspirated obstruents and after sonorants.....	82
Figure 26. Relation between C-F0 DIFF(TH-N) for M-T1(55) and M-T4(51) and Smallest TONE DIFF between M-T1(55) and M-T4(51).....	84
Figure 27. Relation between C-F0 DIFF(TH-N) for M-T2(35) and M-T3(214) and Smallest TONE DIFF between M-T2(35) and M-T3(214).....	84

Figure 28. Scaled F0 difference in ten Cantonese tone pairs (within the first 10ms). Each tick on the x-axis separates the names of the tones. E.g. ‘CT1_CT2’ represents the tone pair C-T1(55) vs. C-T2(25).....	87
Figure 29. Overall TONE DIFF: the average values of all scaled F0 differences between a target tone and all other tones in Cantonese (within the first 10ms). E.g. Overall TONE DIFF (C-T1) = (TONE DIFF[C-T1 vs. C-T2) +TONE DIFF(C-T1 vs. C-T3)+TONE DIFF(C-T1 vs. C-T4)) ÷3.....	88
Figure 30. Average C-F0 DIFF in each Cantonese tone (within the first 10ms). The red boxes represent the difference between F0 values after unaspirated obstruents and after sonorants. The blue boxes represent the difference between F0 values after aspirated obstruents and after sonorants. C-T4(21) only has two consonant categories, unaspirated obstruents and sonorants, but no aspirated obstruent).....	90
Figure 31. Relation between C-F0 DIFF(TH-N) for all four Cantonese tones and Smallest TONE DIFF of the tones.....	92
Figure 32. Average C-F0 DIFF in each C-T1(55) and M-T1(55) (within the first 10ms).....	94
Figure 33. Normalized VOT of aspirated and unaspirated obstruents across Mandarin tones.....	99
Figure 34. F0 values(within the first 10ms) after aspirated and unaspirated obstruents across Mandarin tones.....	101
Figure 35. The relation between normalized VOT and F0 (within the first 10ms) for M-T1(55).....	103
Figure 36. The relation between normalized VOT and F0 (within the first 10ms) for M-T2(35).....	104
Figure 37. The relation between normalized VOT and F0 (within the first 10ms) for M-T3(214).....	105
Figure 38. The relation between normalized VOT and F0 (within the first 10ms) for M-T4(51).....	106
Figure 39. Normalized VOT of aspirated and unaspirated obstruents across Cantonese tones.....	109
Figure 40. F0 values (within the first 10ms) after aspirated and unaspirated obstruents across Cantonese tones.....	110
Figure 41. The relation between normalized VOT and F0 (within the first 10ms) for C-T1(55).....	112



Figure 42. The relation between normalized VOT and F0 (within the first 10ms) for C-T2(25).....	113
Figure 43. The relation between normalized VOT and F0 (within the first 10ms) for C-T3(33).....	114

# **1. Introduction**

## **1.1 An introduction to consonantal effects on F0 (C-F0)**

Consonants may influence F0 of adjacent vowels (the consonantal effect on F0 is henceforth called C-F0). Voiced obstruents are known to consistently lower F0 of the adjacent vowels across languages, while voiceless obstruents usually raise F0. This effect has been argued to lead to tonogenesis, a process in which laryngeal articulation of adjacent consonants introduces tones into a previously toneless language (Kingston, 2011). On the other hand, there is no consensus regarding the effects of aspiration on F0: Some studies show that aspiration lowers pitch (Gandour, 1974; Jeel, 1975; Xu & Xu, 2003; Francis et al., 2006), while others suggest that it raises pitch (Ewan, 1976; Zee, 1980; Lai et al., 2009). This dissertation is interested in investigating aspiration C-F0 that has not reached an agreement.

It is also intriguing to compare consonantal effects in tonal languages and non-tonal languages. It has been observed that influences on pitch perturbations extend to 100ms post-onset in non-tonal languages like English, while the duration of the effect is much shorter in tonal languages, in which C-F0 extends approximately 30-50ms into the vowels (Francis et al., 2006). The restricted magnitudes of C-F0 in tonal languages indicate that lexical tones can influence C-F0, but it is unclear what aspects of tones or tone inventories are making a difference. An important goal of this dissertation asks is how consonant aspiration and lexical tones influence C-F0. Another purpose is to investigate whether C-F0 is originated from the enhancement of consonant or tone features, or is an unintended by-product arising from physical factors.

To be more specific, this dissertation investigates how C-F0 can be related to tone and consonant voicing by asking the following questions: (1) Is C-F0 conditioned by lexical tones?

(2) Is C-F0 conditioned by F0 difference between tones? (3) Is there a cue trading relation between F0 and VOT for aspiration contrast?

The first question will reveal whether there is consistent C-F0 in tonal languages that have aspiration contrast in obstruents. The second and the third questions show how tone contrast and aspiration contrast is related to C-F0. Elaboration of the research questions will be given in Section 1.3.

The following sub-sections 1.1.1 – 1.1.3 provide the background review for how the two main factors, consonants and tones, have been previously observed to influence C-F0.

### 1.1.1 Consistent consonantal effects on F0

Voiced obstruents usually depress F0, and voiceless ones usually raise F0. This is a robust observation found in many languages, especially those that have contrastive [voice] (Kingston & Diehl, 1994; Tang, 2008). A typological survey is given below to show the consistency of the voicing effect.

Effects Consonant	Lowering	Raising
Voiced	Spanish, French, Italian, Portuguese, Hindi, Thai, German, Swedish, English	
Voiceless		Spanish, French, Italian, Portuguese, Hindi, Thai, German, Swedish, English

Table 1. Example languages where the voicing effect has been observed (adapted from Kingston & Diehl (1994))

There are at least two proposed sources for the voicing effects on F0: an automatic account and an enhancement account. The automatic account proposes that this effect is due to physiological and aerodynamic factors (Hombert, Ohala, & Ewan, 1979; Whalen & Levitt, 1995; Connell, 2002). Lowered F0 after voiced stops is believed to be related to vertical tension of the

vocal cords, which can be caused by articulations designed to enlarge the supraglottal cavity and facilitate voicing, such as expansion of the pharyngeal cavity or lowering of the larynx (Ewan, 1976; Hombert et al., 1979). The raising effects of voiceless obstruents, on the other hand, can be seen as a side-effect of the elevated transglottal airflow through the open glottis. In contrast, the enhancement account does not consider the consonantal effects as unintended side effects of automatic aerodynamic or articulation differences, but an outcome of controlled articulation for enhancing the contrastive feature [voice] (Kingston & Diehl, 1994; Kingston, 2007). The enhancement account is supported by the observation that consistent differences in pitch perturbations after phonologically voiced and voiceless consonants are not affected by different phonetic realizations. For example, although English has allophonically aspirated voiceless consonants and unaspirated voiceless consonants elsewhere, e.g. [k<sup>h</sup>i] vs. [ski], the pitch perturbation does not differ next to these two allophonically different contexts. The enhancement account attributes such a lack of main difference to how both aspirated and unaspirated allophones are [-voice] underlyingly, and thus there is no need to enhance any phonological contrast through pitch perturbation differences. However, this argument is vulnerable to the criticism that the influence on pitch perturbations from phonetic aspiration is unclear. The lack of pitch perturbation differences after aspirated and unaspirated consonants may be attributed to the fact that the phonetic aspiration effect on F0 is weak or unstable in general, even in languages in which the phonemic status of [spread glottis] is not disputed. For example, voiceless unaspirated obstruents and voiceless aspirated obstruents are phonemically distinct in Hindi, but some studies show that the pitch perturbation differences between these two types of voiceless consonants, are much smaller than the differences between voiced consonant and voiceless consonants (Dutta, 2007). More understanding of the less known consonantal effects may thus

help us to further explore the source of the voicing effects. The following section reviews the aspiration effect and the sonorant effect on pitch.

### 1.1.2 Inconsistent consonantal effects on F0

Unlike the voicing effect, there is greater disagreement in the literature on consonantal effects on F0 by aspiration and sonorancy. As shown in Table 2, some studies show that aspiration lowers pitch, while others suggest that it raises pitch. The inconsistency even appears among studies on the same language, e.g. aspiration in Thai was reported to lower pitch in (Gandour, 1974), but was reported to raise pitch in (Ewan, 1976). Likewise, sonorants are pitch depressors in some studies, but neutral in other studies as shown in Table 2.

Effects Consonant	Lowering	Raising	Neutral
Aspiration	Mandarin (Xu & Xu, 2003), Cantonese (Francis et al., 2006), Danish (Jeel, 1975), Thai (Gandour, 1974)	Mandarin (Chen, 2011), Cantonese (Zee, 1980), Taiwanese (Lai et al., 2009), Thai and Japanese (Ewan, 1976)	Danish (Fischer-Jørgensen, 1968)
Sonorant	Burmese <sup>1</sup> (Maddieson, 1984), Danish (Jeel, 1975), Hindi (J. Ohala, 1980)		Thai (Gandour, 1974), Tibetan (Kjellin, 1977), Bade (Tang, 2008)

Table 2. Aspiration and sonorant effects across languages

In an automatic account, the inconsistency of the consonantal effects can be attributed to unstable articulation or aerodynamic properties when producing aspirated consonants and sonorants, while voicing is more stable. It has also been found that the aspiration effect on pitch varies among speakers (Zee, 1980; Lai et al., 2009), which suggests considerable individual differences for the inconsistent consonantal effects. In an enhancement account, the asymmetric pattern of consistent and inconsistent effects is probably due to different demands for

---

<sup>1</sup> Voiced sonorants lower F0 in Burmese, while voiceless sonorants raise F0 (Maddieson, 1984).

enhancement of the features [voice], [spread glottis] and [sonorant]. For example, it is much more robust for [voice], regardless of phonetic realization, to induce the vowel lengthening effect, but in languages that are specified with [spread glottis], such effect is less robust or inconsistent. For example, Maddieson & Gandour (1976) found vowels were longer before aspirated than unaspirated coda consonants in Hindi, whereas Ohala & Ohala (1972) reported that the aspiration effect in Hindi was not consistent across different places of articulation, and Lampp & Reklis (2004) found no vowel lengthening effect of coda aspiration in Hindi. However, this is an unlikely explanation for inconsistency within the same language, since the phonological representations should be the same. Mielke (2005) also observed inconsistency of sonorants in other phonological processes, such as vowel lengthening, total assimilation and so on, which suggests that the inconsistent behaviors are results of inconsistent phonetic properties.

### **1.1.3 Effects in non-tonal languages vs. tonal languages**

It has been observed that influences on pitch perturbations extend to 100ms post-onset in non-tonal languages like English, while the duration of the effect can be 50-70ms shorter in tonal languages (Francis et al., 2006). In addition to duration of pitch perturbation, tonal languages and non-tonal languages also differ in their interactions with prosodic contexts. For non-tonal languages, there is evidence of reliable consonantal effects in different prosodic contexts: Kingston (2007) found that in English, consonantal effects on F0 in unaccented syllables and in accented syllables do not differ significantly. However, there are also contradictory findings that show that the consonant perturbation pattern in non-tonal languages is influenced by prosodic contexts, such as pitch accent and intonational contours: Jun (1996: 99-105) observed that C-F0 was smaller when the target CV syllable was put in the middle of an Accentual Phrase in French and Korean than at the beginning of an Accentual Phrase, The effect was also smaller when the

CV target was in a post-nuclear pitch accented position in English than in a nuclear pitch accented one.

On the other hand, some tonal languages are reported to have interactions between consonantal effects and lexical tonal categories. Xu & Xu (2003) found the lowering effect of aspiration in Mandarin was significant only in the mid rising tone category and the low tone category, whereas the lowering effect was not crucial in other tonal categories like the high level tone or the high rising tone. Perkins (2014) also showed that consonantal effects on F0 were only significant in the high level tone and the high rising tone in Thai, but not in other tonal categories. Nevertheless, there is no generalized account for what tonal category is most closely associated with consonantal pitch perturbation and why it is so. As part of the research question (2), the current dissertation asks whether lexical tonal contexts play a crucial role in conditioning consonantal effects on F0 in tonal languages.

## **1.2 The debate between the enhancement account and the automatic account**

There are at least two possible interpretations for C-F0: the enhancement account and the automatic account. Different terminologies have been proposed for these two accounts: Kingston & Diehl (1994) proposed ‘controlled’ articulation as the source of speakers’ adaptively planned enhancement of the perceptual distinctiveness of contrastive laryngeal feature, and ‘automatic’ implementation, on the other hand, for depicting the unintended biomechanical source (also see (Kingston, 2007; Kluender, Diehl, & Wright, 1988; Van Summers, 1987)). Hyman (1976, 2013) described the automatic process as ‘universal phonetics’ and the speaker-controlled one as ‘language-specific phonetics’. The two processes are sequentially related in a hypothetical diachronic analysis: the universal phonetic property, highlighting the ‘perhaps unavoidable’ automatic characteristics, later ‘takes on a language-specific form’, which may then be

phonologized into a structured category. Chen (2011) teased apart these theories by the ‘phonologically’ controlled view and the ‘phonetic’ view, distinguishing the trigger of maximizing phonological contrastive features and pure physiological and aerodynamic factors.

Two kinds of enhancement accounts will be reviewed in this sub-section: the feature enhancement hypothesis regards C-F0 as a consequence of controlled phonetic implementation for the teleological goal of reinforcing phonological contrastive features. It also captures the language-specific characteristics, as phonetic knowledge is argued to be intimately connected to the phonology of a specific language, through which ‘phonological strings are transformed into articulations or are recognized in the acoustic signal’ (Kingston & Diehl, 1994). However, this feature enhancement account does not predict trading relations between F0 and the physical correlates to the relevant phonological specification, while the probabilistic enhancement hypothesis (Kirby, 2010; 2013, p. 232) does. The probabilistic enhancement hypothesis claims that phonetic cues, such as VOT and F0, can be seen as probabilistic functions of contrast precision by listeners and the degrees of cue informativeness. The contrast precision is based on the statistical distribution of acoustic-phonetic cues to the contrast, and thus does not directly get categorically enhanced by the values of cues. A negative correlation between the weights of cues can be expected, when the contrast is maintained in a trading relationship between cues, i.e. the less weight one cue gets the contrast in question, the more weight the other competing cue will get.

The automatic account, on the other hand, denies a teleological explanation for C-F0. Instead, C-F0 is an unintended by-product arising from articulatory factors (such as larynx activities and cricothyroid activities) and aerodynamic factors (such as the degree of transglottal air flow, transglottal pressure, and subglottal pressure).



The following sub-sections present the debate between the two competing accounts in more detail. Section 1.2.1 introduces the feature enhancement hypothesis and the probabilistic enhancement hypothesis. Section 1.2.2 reviews how the automatic view account for C-F0 through articulatory and aerodynamic factors.

## **1.2.1 The enhancement account**

### **1.2.1.1 The feature enhancement hypothesis**

The feature enhancement hypothesis proposes that speakers adaptively control articulations to achieve the necessary restrictiveness that allows speakers to maximize perceptual distinctiveness, and thus reinforces a phonological contrast (Diehl, 2008; Diehl & Kluender, 1989; Kingston, 2007; Kingston & Diehl, 1994). A similar model of planned articulation by Keyser & Stevens (2001, 2006) also proposes that controlled motoric instructions can contribute to enhancing the saliency of the distinctive features.

To support the claim that the source of C-F0 originates from a phonological specification rather than phonetic attributes, the primary evidence comes from reports of F0 differences being independent of phonetic realization of a distinctive feature. For example, F0 is reported to be uniformly lower in vowels next to initial as well as intervocalic [+voice] stops in English, independent of the absence of voicing during the closure in the initial context or its presence in the intervocalic context (Caisse, 1982; Kingston, 1985; Kingston & Diehl, 1994). A second type of evidence is the great variability of C-F0 in neutralized contexts. In English, F0 is observed to be either higher or lower following the neutralized aspirated stop after [s] than F0 following the aspirated allophone of a voiceless stop occurring alone (Caisse, 1982; Kingston & Diehl, 1994; Ohde, 1984). For example, the vowel F0 after [sp] in ‘spy’ is more variable than F0 after [p<sup>h</sup>] in ‘pie’. The variability in the neutralized context, i.e. either high or low F0 after the neutralized

consonant, is predicted by the feature enhancement hypothesis, for the [voice] contrast has been neutralized after [s] and can no longer exert any control over F0.

A major weakness of the feature enhancement hypothesis is that it does not provide an explanation for the trading relations among acoustic cues for the phonological contrast<sup>2</sup> (Kirby 2010: 15). Following Repp (1982), the trading process can be defined as a change in one acoustic dimension, which would result in a change of the phonetic percept, that can be offset by an opposing change in another acoustic dimension. The trading relation is left unexplained by the feature enhancement theory, for it is based on the invariant phonetic implementation of non-primary acoustic cues independent of changes of the primary cues, which does not predict the dynamic offset process among cues that contribute to the perceptibility of the contrast. The following subsection introduces the probabilistic enhancement hypothesis that has examined the trading relations among cues.

#### **1.2.1.2 The probabilistic enhancement hypothesis**

The probabilistic model proposed by Kirby (2010; 2013) hypothesizes that when the precision of a contrast along one acoustic dimension is reduced, other dimensions may be enhanced to compensate. Such a model of trading relations claims that precision is based on the statistical distribution of acoustic-phonetic cues to the phonological contrast, and thus the contrast does not directly get categorically enhanced by the values of cues, but through the degree of precision influenced by the number of cues competing over some acoustic-phonetic space.

An example of such competing relations among cues comes from the study by Shultz et al. (2012) on cue weighting of consonant voicing in American English. The production data

---

<sup>2</sup> Kingston et al. (2008) proposed that the various acoustic correlates of voicing combined into an integral “intermediate perceptual property”, which is consistent with the trading relationship by Repp (1982).

shows that VOT coefficients<sup>3</sup> are inversely correlated with onset F0 coefficients, indicating that the speakers have a stronger weighting for VOT when the weighting of F0 is weak, and vice versa. However, the perception data from the same study do not show such a significant trade-off as in production. That is, VOT coefficients are not significantly inversely correlated to F0 coefficients. Similar to Shultz et al. (2012), Kirby & Ladd (2015) also found a significant inverse correlation between VOT and F0 for Italian voiceless stops, but no significant correlation for French voiceless stops. For voiced stops in both languages, in contrast, post-release onset F0 is higher following longer voicing lead, i.e. a positive correlation was found.

### **1.2.2 The automatic account**

This section provides a review for an automatic view of C-F0, which proposes that the difference of F0 is fundamentally a consequence of biomechanical effects of articulation maneuvers or/and aerodynamic factors. We focus on the physical properties of voicing and aspiration in this review.

#### **1.2.2.1 Articulation maneuvers**

Several articulatory mechanisms have been proposed to explain the general patterns of lower F0 after voiced than after voiceless:

Larynx activity: during the production of voiced stops, the larynx lowers to facilitate the initial state of voicing by expanding the oral cavity. A larger tube may thus lower the fundamental frequency after voiced stops, whereas the larynx does not lower after voiceless stops and thus F0 is not depressed (Collier, 1974; Ewan & Krones, 1974; Hombert et al., 1979; Ohala, 1970; Simada & Hirose, 1971). This account predicts that C-F0 is fundamentally a lowering effect from voiced consonants rather than a raising effect.

---

<sup>3</sup> A larger coefficient indicates a stronger weighting of a variable.

Cricothyroid (CT) activity and the state of vocal folds: CT is elevated to inhibit voicing in postvocalic voiceless stops by increasing horizontal vocal fold tension. It's believed that a longer tension gives rise to a bigger increase of F0 in following vowels after voiceless stops than a shorter tension in voiced stops does (Dixit & MacNeilage, 1980; Hirose, Lee, & Ushijima, 1974; Hirose et al., 1973; Hirose & Ushijima, 1978; Löfqvist et al., 1989; Löfqvist et al., 1984). Contrary to the larynx lowering account, CT activities being the primary factor means that C-F0 is a raising effect rather than a lowering effect.

For comparing F0 after voiceless aspirated stops and that after voiceless unaspirated stops, the following mechanisms have been proposed: during the voiceless interval of obstruents, the vocal folds are stiffened for the purpose of inhibiting vocal-fold vibration (Hanson & Stevens, 2002; Stevens, 2000). The stiffened state of vocal fold can carry over into the following vowel and elevate F0 at the vowel onset for both voiceless aspirated consonants and voiceless unaspirated consonants. However, Xu & Xu (2003) argued that there was no evidence of different muscular activities around the glottis after aspirated as opposed to unaspirated stops when the following voicing period initiate (Hombert, 1978). Alternatively, C-F0 for aspirated and unaspirated stops is more due to aerodynamic factors such as subglottal pressure and transglottal airflow.

#### **1.2.2.2 Aerodynamic conditions**

The following aerodynamic factors were brought up in the previous literature:

Transglottal airflow: an open glottis in voiceless stops may elevate transglottal air flow as a side effect, which also raise F0 afterwards (Kohler, 1984). This account assumes that C-F0 is primarily a raising effect from voiceless consonants, rather than lowering effects from voiced consonants.

Similarly, for languages that have an aspiration contrast, two sets of aerodynamic factors have been brought up for explaining higher or lower F0 after aspirated stops than unaspirated ones:

Transglottal pressure: transglottal pressure is commonly assumed to be much higher produced with aspiration than with release of the unaspirated stops, and such transglottal pressure may induce positively proportionally varied F0. Thus, an aspirated stop gives rise to a higher onset F0 than an unaspirated stop does (Hombert, 1975). However, see the arguments by Xu & Xu (2003) that transglottal pressure is not necessarily higher in aspirated consonants than in unaspirated consonants.

Subglottal pressure: Xu & Xu (2003) provided an account for lower F0 after aspirated stops based on aerodynamic conditions: a constant subglottal pressure is produced during the closure for all stops (Ladefoged, 1972; Löfqvist, 1975; Ohala & Ohala, 1972; Slis, 1970). Another aerodynamic factor comes later to shape the left-over subglottal pressure that survives before the vowel: the transglottal airflow, which is generally assumed to be much faster in the aspiration of aspirated stops than in the release of unaspirated stops. Different from Hombert's (1975) assumption that transglottal pressure is higher after the aspiration of an aspirated stop than after the release of an unaspirated stop, Xu & Xu (2003) regarded transglottal pressure as a consequences of the earlier subglottal pressure being influenced by the later transglottal airflow. At the aspiration of an aspirated stop, a high rate of airflow gives rise to a tremendous decrease of the earlier generated subglottal pressure, whereas a low rate of airflow induces a slow and gradual decrease of the constant subglottal pressure. By the time of vowel onset, the left-over subglottal pressure remains generally lower after an aspirated stop than an unaspirated one (Ohala, 1974, 1976, 1978; Ohala & Ohala, 1972), which result in lower transglottal pressure

after an aspirated stop than an unaspirated one. Since F0 is believed to be proportionally correlated with transglottal pressure, F0 is therefore lower after an aspirated stop than an unaspirated one.

### **1.3 Research questions**

This dissertation is interested in C-F0 in tonal languages that have obstruents distinguished by aspiration. The findings will firstly provide an answer to the controversy on whether aspiration has a lowering or raising effect on F0. Besides, this dissertation also asks the following three questions, and the background for each question is given in section 1.3.1 – 1.3.3:

- (1) Questions 1: Is C-F0 conditioned by lexical tones?
- (2) Questions 2: Is C-F0 conditioned by F0 difference between tones?
- (3) Questions 3: Is there a cue trading relation between F0 and VOT for aspiration contrast?

Another goal of this dissertation is to find out whether it is the enhancement account or the automatic account that can account for C-F0 in tonal languages. These two accounts will have different predictions for the above three questions, as clarified in the following sub-sections.

#### **1.3.1 Question 1: Is C-F0 conditioned by tones?**

Previous studies have shown that prosodic contexts may condition C-F0 (Pitch-accent: (Jun, 1996; Kingston, 2007); Focus: (Chen, 2011); Intonation: (Kirby & Ladd, 2016)). Lexical tones have also been reported as an influencing factor for the magnitude and extent of C-F0 (Chen, 2011; Erickson & Abramson, 2013; Francis et al., 2006; Lai et al., 2009; Xu & Xu, 2003). For example, Xu & Xu's (2003) study on Mandarin found that onset F0 was significantly lower after aspirated stops than after unaspirated stops only in the mid rising M-T2(35)<sup>4</sup> and the falling-

---

<sup>4</sup> In the pitch number system of Chinese, the tone code '5' indicates the highest pitch of the pitch range and '1' the lowest (Chao, 1930; 1968:26). In this paper, lexical tones are transcribed with Chao tone numbers. For example,

rising M-T3(214), whereas such statistical significance was not found in other tonal categories like the high level M-T1(55) or the high-falling M-T4(51). However, Xu & Xu (2003) only analyzed the mean values within the 105ms from the syllable onsets, so it is unclear what the differences are when individual points at the tone contour are compared: Figure 1 shows that the first point of the F0 trajectory is higher after the unaspirated /t/ than after the aspirated /t<sup>h</sup>/ in all tone contexts.

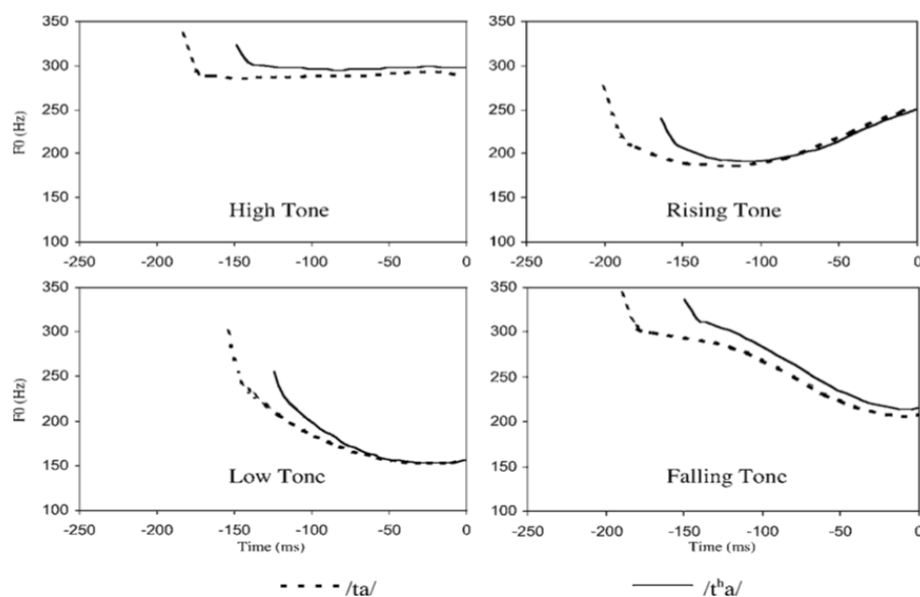


Figure 1. Average F0 contours of syllable /ta/ versus /t<sup>h</sup>a/ in four tones. The right boundary of each panel represents syllable offset. (from Xu & Xu (2003))

Francis et al.'s (2006) work, on the other hand, has studied C-F0 in three Cantonese tone contexts, i.e. high-level C-T1(55), mid-rising C-T2(35) and low falling C-T4(21) (figure 2). Their results showed that at contemporaneous points across all three Cantonese tones, the only significant differences due to aspiration were found at 0ms, where the unaspirated tokens were higher than the aspirated ones. The only other case in which the aspirated and unaspirated series

---

T1(55) or Tone 1(55) is the transcription for the high level tone, and *ba1*(55) is the transcription for the syllable /ba/ with a high level tone T1(55). If it's a tone in Mandarin, it's specified as M-T(tone number)(pitch number). For Cantonese tone, it's marked as C-T(tone number)(pitch number).

differed in frequency at the same time was in the low falling C-T4(21) at 100ms post voicing onset, where F0 was significantly higher after aspirated stops than after unaspirated.

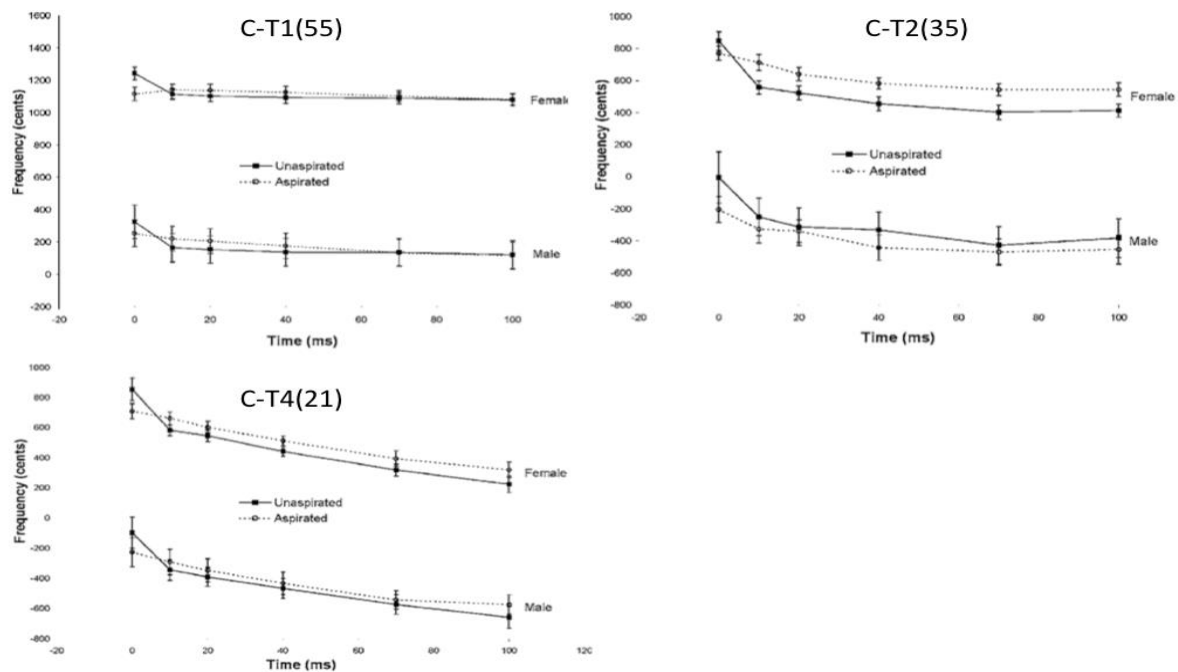


Figure 2. Mean F0 (in cents) for syllables beginning with aspirated and unaspirated bilabial stops and having C-T1(55), C-T2(35) or C-T4(21). Error bars indicate standard error of the mean. (from Francis et al. (2006))

Comparing the results from Xu & Xu (2003) and Francis et al. (2006), the patterns of C-F0 are similar between Mandarin and Cantonese at the beginning point of the F0 contour in the high-level contexts (i.e. M-T1(55) and C-T1(55)) and the rising contexts (i.e. M-T2(35) and C-T2(35)). However, it is unclear how to compare C-F0 beyond the first point of the F0 contour in these two studies, given that different variables (e.g. the gender effect in Francis et al. but not in Xu & Xu) were examined, and F0 values were not measured at the same points. Moreover, no baseline was used in these two studies, so it is unclear whether the C-F0 direction, i.e. lowering or raising, is the same in these two languages. This current dissertation aims at conducting a more controlled experiment to study the effect of tones on C-F0 in Mandarin and Cantonese.



### 1.3.2 Question 2: Is C-F0 conditioned by F0 difference between tones?

The Theory of Adaptive Dispersion (TAD) (Liljencrants & Lindblom, 1972; Lindblom, 1986) propose that perceptual contrast plays a role in shaping the phonetic structure of vowel inventories, where vowels of a given language are dispersed in an acoustic space to be considered sufficiently contrastive. In a given inventory, vowels are predicted to be widely dispersed across the vowel space to achieve sufficient contrast. To elaborate the notion of sufficient contrast, Lindblom (1986) had the algorithm enumerate the best subset of systems for each system size (n) and assumed that sufficient contrast was invariant across languages and system sizes, following which assumption, phonetic values of vowels were predicted to exhibit more variation for small-sized systems than for large ones.

Alexander (2010) has tested TAD on tone systems, especially on how the size of a tonal inventory affects acoustic tone space and tone dispersion within the tone space. She defined tone space as the F0 difference in semitones between each language's highest and lowest tones, and tone dispersion as the F0 distance between two tones shared by the target languages. F0 values were measured across the vowel at onset, midpoint, and offglide.<sup>5</sup> TAD predicts that languages with larger tone inventories have larger tone space at onset. Against this prediction, Alexander found that Cantonese (six tones) had a smaller tone space in the onset position than the onset tone space of Thai (five tones), which in turn has smaller tone space than Mandarin (four tones) does (See Figure 3).

---

<sup>5</sup> For the interest of this dissertation, only the results of onset F0 in (Alexander, 2010) are reviewed in details here.

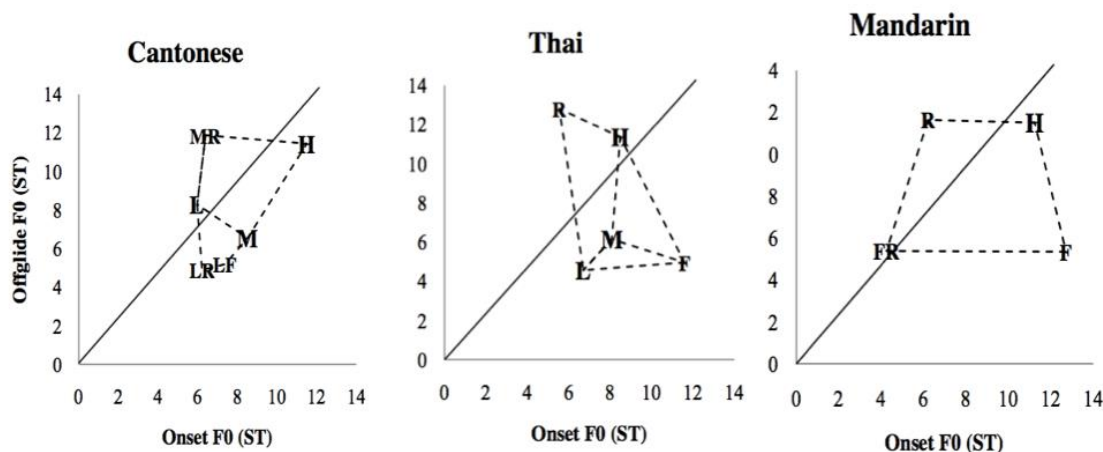


Figure 3. Acoustic tone space at onset and offglide in Cantonese, Thai and Mandarin in semitones. Tone spaces are illustrated as plots of F0 offglide (mean F0 at the tenth equidistant timepoint) x F0 onset (mean F0 at the second equidistant timepoint k1). L=Low Tone; LR=Low-Rising Tone; LF=Low-Falling Tone; M=Mid Tone; MR=Mid-Rising Tone; H=High Tone; FR=Falling-Rising Tone. (from Alexander (2010: 122-123))

To compare the degree of tone dispersion across languages, Alexander measured the mean F0 of a given tone relative to the mean F0 of a tonal baseline, a comparison tone that is phonetically similar in the languages. Thus, the pairs of high-level and mid-rising tones in Cantonese and Mandarin (i.e. Pair M-T1(55) and M-T2(35) and Pair C-T1(55) and C-T2(35)) were chosen to investigate the degree of tone dispersion in these two languages. TAD assumes that the degree of tonal dispersion displayed by a language is correlated with both its tone-space size and the size of its tonal inventory, so it predicts that the degree of tone dispersion of mid-rising tone relative to high-level tone should be smaller in Cantonese (bigger inventory and smaller onset tone space) than in Mandarin (smaller inventory and bigger onset tone space). However, the results show that the F0 difference between high-level tone and mid-rising tone are not significantly different in Mandarin and Cantonese, indicating that the degrees of dispersion is similar in these two languages.

Following Alexander (2010), this study also regards onset F0 difference between tones as an indicator of degree of dispersion at tone onset, which also indicates the sufficiency of tone

contrasts at the initial position. Larger F0 difference indicates higher degree of tone dispersion and of tone contrast sufficiency. According to TAD, a sound inventory of sufficient contrasts may exhibit more variations. Assuming C-F0 tends to increase variations of tones, insufficiency in tonal contrasts, represented by small onset F0 difference between tones, may need to exert more inhibitory influence on C-F0 to constrain the variation. By contrast, sufficient tone contrasts in the tone system are likely to allow more variations of tones, and thus have less restriction on C-F0. In other words, whether tone contrasts have an influence on C-F0 can be examined by investigating the relation between F0 differences between tones (as an indicator of degree of tone dispersion and of tone contrast sufficiency) and F0 differences following different consonant types (as an indicator of magnitudes of C-F0).

If tones contrasts influence C-F0, the prediction is that larger F0 differences between tones (suggesting higher degree of tone contrast sufficiency) is related with larger F0 differences following different consonant types (suggesting magnitudes of C-F0), whereas smaller F0 differences between tones (suggesting lower degree of tone contrast sufficiency) is related with smaller F0 differences following different consonant types (suggesting smaller magnitudes of C-F0). In other words, positive correlation between F0 differences between tones and F0 differences following different consonant types is predicted.

### **1.3.3 Question 3: cue trading?**

The trading relation of cues claims that the reduction of the value of one phonetic cue for a phonological contrast can be offset by a change in the value of another cue (Kirby, 2010, 2013; Repp, 1982). A model of trading relations proposes that the precision of identifying a phonological contrast is based on the statistical distribution of acoustic-phonetic cues to the contrast, and thus the contrast does not directly get categorically enhanced by the values of cues,

but through the degree of precision<sup>6</sup> influenced by the number of cues competing over some acoustic-phonetic space.

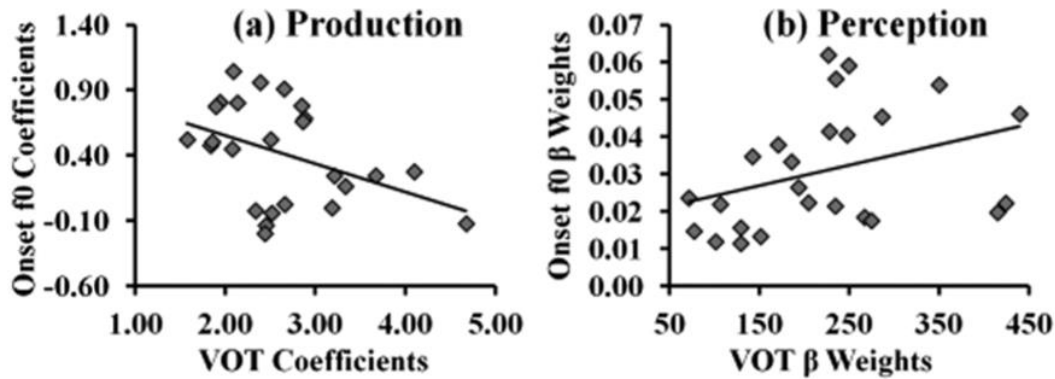


Figure 4. Relation between F0 and VOT coefficients in production (a) and perception (b) (from Shultz et al. (2012))

As discussed above, an example of such trade-off relations among cues comes from the study by Shultz et al. (2012) on cue weighting of consonant voicing in American English (Figure 4). Their production data shows that VOT coefficients<sup>7</sup> are inversely correlated with onset F0 coefficients, indicating that the speakers have a stronger weighting for VOT when the weighting of F0 is weak, and vice versa. However, the perception data from the same study do not show such a significant trade-off relation as in production – VOT coefficients are not significantly correlated with F0 coefficients in perception. However, one limitation of this study is that all data of voiced and voiceless were collapsed for analysis, so it is unclear whether the relations are the same in the two laryngeal categories respectively in perception and production.

Kirby & Ladd (2015) has also conducted a study on the relation between stop voicing and F0 perturbation in Italian and French. Similar to Shultz et al. (2012), an significant inverse

---

<sup>6</sup> According to Kirby (2010: 18), precision may be “reduced for a variety of reasons, including channel noise introduced by bias factors, or change in the system of contrast at the structural level, which may result in an increase or decrease in the number of categories competing over some acoustic-phonetic space”.

<sup>7</sup> A larger coefficient indicates a stronger weighting of a variable.

correlation between VOT and F0 has been found for Italian voiceless stops, but no significant correlation for French voiceless stops. For voiced stops in both languages, in contrast, post-release onset F0 was higher following longer voicing lead – a positive correlation was found. However, those results were not replicated in a later study Kirby & Ladd (2016) that had a better control of the intonation contexts: none of the covariates between VOT and F0 reached significance, and considerable individual differences in the direction of the correlation were observed. The results are compatible with the findings in Dmitrieva et al. (2015), where within-category VOT was found uncorrelated with onset F0 in Spanish.

Few studies have investigated the relation of VOT and C-F0 in Mandarin and Cantonese. VOT can act as a strong cue for aspiration in Cantonese (Khouw & Ciocca, 2007) and Mandarin (Chao, 1992). F0 is also reported to be a perceptual cue for aspiration (Korean: Kim et al. (2002); Cantonese: Francis et al. (2006)). If there is a trading relation between VOT and onset F0 for the aspiration contrast, there should be a negative correlation between VOT difference and F0 difference. The trading relation predicts that F0 values should be at an extreme level when VOT values are close to the ambiguous areas i.e. close to the VOT values that can induce most confusion with the other laryngeal category.

It is worth mentioning that there have been reports on the interaction of tones and VOT: high level M-T1(55) and high-falling M-T4(51) are reported to be associated with shorter VOT values than mid-rising M-T2(35) and falling-rising M-T3(214) by Chen & Ng (2005) and Liu et al. (2008).<sup>8</sup> The authors indicated that the longer VOT values in M-T2(35) and M-T3(214) may be related to the need to increase the tension in the larynx to prepare for a higher F0 in the later

---

<sup>8</sup> Liu et al. (2008) found that only M-T4(51) was associated with statistically significantly shorter VOT than M-T2(35) and M-T3(213) were. VOT before M-T1(55) was not crucially short.

rising direction during the production of the two tones. As for the cue trading issue, it is also interesting to investigate whether there is any interaction among tone, VOT and onset F0.

#### **1.4 An overview of the dissertation**

Following the introductory chapter, Chapter 2 provides a background review for the two target languages: Mandarin and Cantonese. Both the phonological representations and the phonetic properties of the consonants and the tones in these two languages will be introduced. Chapter 3 presents the methodology of the experiments. It presents the selection criteria of participants, the design of stimuli and the experiment procedure for the Mandarin group and the Cantonese group respectively, describes the measurement for duration values and F0 values, and discusses the statistical tests for analyzing the data.

Chapter 4 investigates Research Question 1. The results show that tones can influence C-F0 LENGTH (i.e. how far in the vowels C-F0 can extend) and C-F0 DIFF (i.e. how big the differences between F0 values after different types of consonants are) in both Mandarin and Cantonese. Tones can also condition C-F0 DIRECTION (i.e. whether F0 values are raised or lowered as consequences of C-F0) in Cantonese, but not in Mandarin. In general, C-F0 extended farther in tones associated with high pitch. There was a consistent direction of C-F0 across all Mandarin tones and C-T1(55) and C-T3(33): F0[aspirated]<sup>9</sup> was higher than F0[unaspirated] and F0[sonorant]. F0[unaspirated] started from the middle and converge with the lowest F0[sonorant]. The results indicate a robust aspiration raising and a weaker voiceless raising effect in both Mandarin and Cantonese.

Chapter 5 examines Research Question 2. Two hypotheses were suggested, which predicted that the pattern of C-F0 magnitudes may either follow the pattern of Smallest TONE

---

<sup>9</sup> F0[consonant type] represents F0 values after the consonant type in the square brackets.

DIFF (i.e. the F0 difference between the target tone and its closest tone in the tone inventory) or Overall TONE DIFF (i.e. the average value of a target tone's distances from all other tones in the tone inventory). Two indicators of C-F0 magnitudes were examined: C-F0 LENGTH and C-F0 DIFF within the first 10ms. Evidence was found to support the hypothesis that proposed that C-F0 LENGTH had the same pattern as that of Smallest TONE DIFF in different tone contexts. However, neither hypothesis made the right predictions on the C-F0 DIFF within the first 10ms. Finally, the cross-linguistic comparison provides support to the hypothesis that higher degree of tone competition may restrict C-F0.

Chapter 6 focuses on Research Question 3. The results showed that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue. However, the findings do not provide evidence for a cue trading relation between VOT and onset F0 in Mandarin or Cantonese.

Chapter 7 has the conclusion and a general discussion of the implications of the results.

## **2. Language background: Mandarin and Cantonese**

### **2.1 Introduction**

Standard Mandarin and Cantonese are selected as target languages to address the research questions related to C-F0 for the following reasons: Firstly, as will be introduced in the following sub-sections 2.2 – 2.3, these two languages are uncontroversially regarded as aspirating languages and thus providing a testing ground for the reasons for inconsistent reports of aspiration effects on vowel F0. Specifically, there are disagreements on the reports of C-F0 in Mandarin and Cantonese as stated in the previous chapter. For Mandarin, Chao (1992) found that pitch contours after aspirated consonants were generally higher than those after unaspirated ones, whereas Xu & Xu (2003) reported that the mean F0 values after aspirated consonants were lower than after unaspirated consonants in the mid-rising tone and in the falling-rising tone. For Cantonese, Zee (1980) found higher onset F0 values after aspirated /p<sup>h</sup>/ than after the unaspirated /p/, but Francis et al. (2006) observed that the onset F0 after aspirated consonants started lower than after unaspirated consonants. These inconsistent observations provide the motivation for the current study to conduct a more comprehensive probe into C-F0 in Mandarin and Cantonese. Furthermore, these two languages both have the same kind of consonants, i.e. the voiceless aspirated consonants /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/, the voiceless unaspirated consonants /p, t, k/ and the sonorant consonants /m, n, (ŋ), l/, which provide the basis for a cross-linguistic comparison of the C-F0 patterns considering the phonological features of the consonants are more controlled.

The second main reason is that Mandarin and Cantonese have different structures of tone inventories, which allows comparisons of how tones influence C-F0 within the language and across languages: Mandarin has four contrastive tones while Cantonese has six, which provides the testing ground for how different pitch levels (low vs. mid vs. high) and contour shapes (rising



vs. level vs. falling vs. falling-rising) may condition C-F0. Moreover, the different sizes of tone inventories and the likely different degrees of crowdedness (i.e. phonetic distance between tones) allow the investigation on how tone inventory may influence C-F0.

In this chapter, the consonant inventory and the tone system of Mandarin will be introduced in section 2.2 and those of Cantonese in section 2.3. Both the phonological representations and the phonetic properties of the consonants and the tones will be introduced.

## **2.2 Mandarin**

Standard Mandarin is the official language of China and is based on the Beijing dialect. It serves the purpose of intelligible communication among speakers of the several mutually unintelligible varieties of Chinese languages. It is widely spoken throughout China mainly due to the mandatory use in all levels of education and in media. There is often variation in the pronunciation and lexical items of Standard Mandarin correlated with different social variables, such as region, age, education and so on. The consonants of Standard Mandarin are introduced in the following sub-section 2.2.1 and the tones in 2.2.2.

### **2.2.1 Mandarin consonants**

Mandarin is an aspirating language where the laryngeal contrast for obstruents is typically the contrast between voiceless unaspirated and voiceless aspirated. Table 3 shows the consonant inventory of Mandarin. Mandarin has a pair of voiceless aspirated and voiceless unaspirated obstruents and a nasal consonant for each of the three places of articulation: the bilabial consonants /p<sup>h</sup>, p, m/, the alveolar consonants /t<sup>h</sup>, t, n/ and the velar consonants /k<sup>h</sup>, k, ŋ/. The lateral approximant /l/ is included in the stimuli, for Cantonese also has the same sonorant and thus a cross-linguistic comparison is allowed.

As for the relevant phonotactic rules, velars are not allowed to precede high front vowels, so /\*k<sup>h</sup>i, \*ki/ do not exist in Mandarin. Furthermore, the velar nasal /ŋ/ cannot appear in the onset consonant position in Mandarin, whereas Cantonese allows /ŋ/ to be an onset.

	Bilabial		Labiodental		Alveolar		Alveolo-palatal		Retroflex		Velar	
Stop	p <sup>h</sup>	p			t <sup>h</sup>	t					k <sup>h</sup>	k
Nasal		m				n						ŋ
Fricative				f		s		ɕ		ʂ		x
Affricate					ts <sup>h</sup>	ts	tɕ <sup>h</sup>	tɕ	tʂ <sup>h</sup>	tʂ		
Lateral Approximant						l						
Retroflex Approximant										ɻ		

Table 3. The Consonant Inventory of Mandarin. For each column, the segments on the left of the dash lines are voiceless aspirated, and the ones on the right are voiceless unaspirated or voiced.

There are mainly two kinds of proposals for the laryngeal contrasts of consonants in aspirating languages, namely the traditional view and the non-traditional view, or the laryngeal realism view, as in Beckman, Jessen, & Ringen's (2013) terms. The traditional view regards the contrast in both true voice and aspirating languages to be represented by the feature [voice] (Keating 1984; Kingston & Diehl 1994; Wetzels & Mascaró 2001), whereas the non-traditional view argues that in aspirating languages, the laryngeal feature of contrast for stops is [spread glottis], not [voice] (Beckman, Jessen, & Ringen, 2009; Beckman et al., 2013; Harris, 1994; Honeybone, 2005; Iverson & Salmons, 1995; Jessen & Ringen, 2002). However, most of the discussion has focused on Germanic languages, while Mandarin and Cantonese are much less controversially considered to have a clear aspiration contrast (Jessen, 2001). Moreover, empirical evidence from a recent study (Deterding & Nolan, 2007) shows that though there is little phonetic difference in the aspiration of stops in isolated contexts in English and Mandarin, passive voicing (i.e. the voicing during the closure due to coarticulatory pressure from sonorant

consonants or vowels) only occurs in English but not in Mandarin. The absence of passive voicing in Mandarin further supports that it is uncontroversially an aspirating language.

Acoustically, following the grouping by Lisker & Abramson (1964), Mandarin, Cantonese and the more controversial case of English fell into the same two-way contrast group of languages that have 0 to +25ms VOT continuum for voiceless unaspirated stops and +60 to +100 ms for voiceless aspirated stops (cf. (Cho & Ladefoged, 1999; Keating, 1984; Rochet & Fei, 1991)). On the other hand, Chao & Chen (2008) found Mandarin /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ had longer and wider range of VOT values than English /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ and concluded that Mandarin was categorized into the ‘highly aspirated’ group along the VOT continuum while English into the aspirated group, according to the categorization for voiceless aspirated stops by Cho & Ladefoged (1999).

### 2.2.2 Mandarin tones

Standard Mandarin has four lexical tones and a neutral tone (Table 4). The four lexical tones are the high-level M-T1(55), the mid-rising M-T2(35), the falling-rising M-T3(214) and the high-falling M-T4(51). The neutral tone M-T0 lacks a fixed pitch value or contour shape because its value and shape depend on the preceding tone. M-T0 is not considered in this study for it does not appear in the monosyllabic environment.

Tone number	M-T1	M-T2	M-T3	M-T4	M-T0
Pitch number	55	35	214	51	-
Description	high-level	mid-rising	low-falling-rising	high-falling	neutral
Example: transcription: character/ meaning	ma1(55): 妈 ‘mother’	ma2(35): 麻 ‘numb’	ma3(214): 马 ‘horse’	ma5(51): 骂 ‘scold’	-

Table 4. The Tone Inventory of Mandarin

Regarding the phonological representations of tones, there is a debate on whether tones have compositional sequences of multiples tone levels or unitary contours. Early theories considered

contour tones to be phonological units with sequences of level tones (Pike, 1948; Wang, 1967). Duanmu (1994) has also argued that in some Chinese dialects, tones are considered as sequences of tone features, based on the arguments that some tones act as a unit in initial association and can spread as a unit. On the other hand, Wan (2007) and Wan & Jaeger (1998) argued that Mandarin tones are unitary underlyingly based on their findings of similarity between tones and segments in speech errors – both tones and segments behave as single units in the majority of errors. It does not make a difference to this dissertation which representation model is adopted.

Acoustically, F0 height and F0 contour have been found to be the primary acoustic correlates to Mandarin tones (Liang & van Heuven 2004; Alexander 2010). Xu (1997) reported that M-T1(55) stayed as a high-level contour; M-T2(35) starts low and rises after about one-fifths of the vowel; M-T3(214) was found to start with onset pitch values lower than those of M-T2(35), fell until the middle of the vowel, and then rose sharply to the end of the syllable; M-T4(51) started high and then falls sharply at the fifth of the vowel. However, the production results of the current study show different Mandarin tone trajectories of M-T2(35) and M-T3(214) from what the previous literature has reported (Figure 5). M-T2(35) started at a low F0 onset and did not rise as high as M-T1(55) at the end of the syllable. M-T3(214) did not rise in the middle of the vowel.

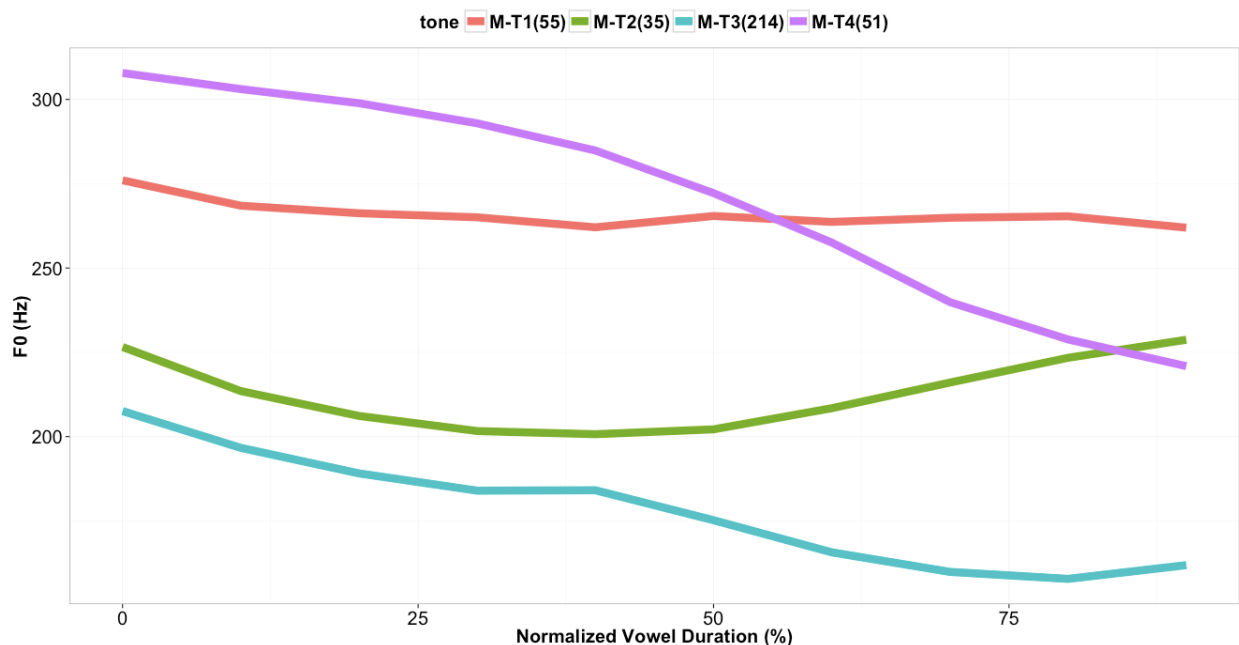


Figure 5. Average F0 values in Mandarin Tone contours within normalized vowels. The 11<sup>th</sup> end point is excluded due to the large amount of undefined measurement from Praat. The data were produced by 15 Mandarin female speakers.

Phonologically, M-T3(214) is considered as underlying low in some phonological analyses (Duanmu, 1999, 2007; Yip, 1980). Phonetically, the low pitch contour was described by Duanmu (1999) as either 211 or 11 in the Chao system (Chao, 1930, 1968) and was observed to be a low tone before another tone in production (Chao, 1968; Duanmu, 2007; Durvasula, Huang, Uehara, Luo, & Lin, to appear; Lin, 2007). Similar to the previous observation, the incomplete realization of M-T3(21) in our data is likely due to the fact that all target syllables are followed by a zi4(51) syllable (to form the meaning “\_\_\_\_\_ character”), thereby creating a disyllabic context that triggers the ‘half-third’ sandhi condition where the pitch 214 becomes 21 when followed by any Mandarin tone except M-T2(35) (Zhang & Lai, 2010).

Contextual factors may influence the phonetic realization of Mandarin tones. Xu (1997) observed that M-T2(35) and M-T3(214) tend to raise the maximum as well as the overall F0 of the preceding tone, except in cases where the preceding tone is M-T3(214). Zhang & Lai (2010)

reported that the disyllabic sequences of ‘M-T2(35) followed by M-T4(51)’ and ‘M-T4(51) followed by M-T4(51)’ had a narrower F0 range in the first syllable than other tone groups. Shih (1987), on the other hand, found M-T1(55), M-T2(35) and M-T4(51) were higher when followed by M-T3(214) than by other tones.

The primary perceptual cues for differentiating Mandarin tones include F0 height, overall F0 contours (Howie, 1976; Xu, 1997), amplitude contour (Whalen & Xu, 1992), voice quality (Garding et al., 1986), and duration (Blicher et al., 1990). The F0 contour has been reported to be the most important cue over other acoustic features for native speakers to judge Mandarin tones (Gandour & Harshman, 1978; Massaro et al., 1985). Listeners may also weigh the cues differently when perceiving different tones. For example, Fu & Zeng (2000) found that discrimination of M-T3(214) heavily relies on the duration cue, discrimination of M-T3(214) and M-T4(51) on the amplitude cue, and recognition of all tones on the pitch cue.

As mentioned earlier in the previous chapter, Mandarin tones are observed to have influence on VOT of the preceding consonants: some studies find that M-T1(55) and M-T4(51) are associated with shorter VOT values, while M-T2(35) and M-T3(214) with longer ones. The longer VOT values in the raising contour (i.e. M-T2(35) and M-T3(214)) may be related to the need to increase the tension in the vocal folds to prepare for a higher F0 during the production, when pitch level is found to increase at certain points and the anticipated increase of tension in the vocal folds and pitch may have delayed the onset of vibration (Wang, 2013). There is inconsistency in the report of the relation between Mandarin tones and VOT: Chao (1992), for example, has not found a major effect from Mandarin tones on VOT in either production or perception.

## 2.3 Cantonese

Cantonese is a Chinese language spoken mainly in Guangdong Province, along with some areas in Guangxi Province, Hong Kong and Macau. In Mainland China, Cantonese is the main lingua franca that native speakers of Cantonese use in daily communication. It is worth noting that most Cantonese speakers also speak Mandarin near-natively due to the mandatory use of Mandarin in schools and media.

This current study is interested in Cantonese that is spoken in Shenzhen, a major city of Guangdong province that is located immediately north of Hong Kong. Shenzhen was designated to be China's first Special Economic Zone in the 1980s. Based on the impressionistic intuition of the author, there is no noticeable difference between the Cantonese variety spoken in Shenzhen and that in Hong Kong in terms of pronunciation, which is probably due to the dominant influence of the Hong Kong culture since the 1980s<sup>10</sup>. The consonants and tones of Cantonese spoken in Shenzhen and Hong Kong are introduced in the following sub-sections.

### 2.3.1 Cantonese consonants

As with Mandarin, there is a two-way aspiration contrast for Cantonese obstruents (Hashimoto, 1972; Lisker & Abramson, 1964). Table 5 shows the consonant inventory of Cantonese. Cantonese also has the bilabial consonants /p<sup>h</sup>, p, m/, the alveolar consonants /t<sup>h</sup>, t, n, l/ and the velar consonants /k<sup>h</sup>, k, ŋ/. Unlike Mandarin, Cantonese allows the velar nasal /ŋ/ to be an onset consonant.

---

<sup>10</sup> The similarity of Shenzhen Cantonese to Hong Kong Cantonese is emphasized in this dissertation, since it provides us the rationale to directly apply the description of the sound system of Hong Kong Cantonese to Shenzhen Cantonese, while there is rarely any descriptive study on Shenzhen Cantonese. It is also worth mentioning that Shenzhen Cantonese has noticeable difference from Guangzhou Cantonese, for the Shenzhen variety does not have a High-Falling tone (53) as a allotone of the High-Level C-T1(55), while Guangzhou Cantonese has.

	Bilabial		Labiodental		Alveolar		Palatal		Velar		Glottal	
Stop	p <sup>h</sup>	p			t <sup>h</sup>	t			k <sup>h</sup>	k	ʔ	
Nasal		m				n				ŋ		
Fricative				f		s					h	
Affricate					ts <sup>h</sup>	ts						
Approximant						l		j				

Table 5. The Consonant Inventory of Cantonese. For each column, the segments on the left of the dash lines are voiceless aspirated, and the ones on the right are voiceless unaspirated or voiced.

Acoustically, similar to Mandarin, VOT is the primary acoustic correlate for specifying the aspiration feature in Cantonese (Cheung, 1986). The aspirated stops are associated with long VOT, whereas the unaspirated stops are associated only with short VOT (Cheung, 1986; Lisker & Abramson, 1964; Tsui & Valter, 2000). Khouw & Ciocca (2007) also suggested that a long VOT indicated the articulatory release and the onset of vocal fold vibration were separated by a longer interval than that found for the unaspirated stops, which allowed the aspirated stops to have breathy noise produced when air passed through the partially closed folds during the long interval.

Perceptually, Tsui & Valter (2000) showed that VOT was an important perceptual cue for the aspiration contrast of initial stops: when the VOT of the aspirated stops was reduced, a majority of listeners perceived the stops as unaspirated. On the other hand, it was also likely that both VOT and aspiration noise were important cues for the aspiration contrast, as they found no significant change in the perception of the unaspirated stops when the VOT interval was lengthened without the addition of aspiration noise.

### 2.3.2 Cantonese tones

Cantonese has six long lexical tones (Table 6). The six tones are the high-level C-T1(55), the high-rising C-T2(25), the mid-level C-T3(33), the low-falling C-T4(21), the mid-rising C-T5(23) and the low-level C-T6(22).



Tone number	C-T1	C-T2	C-T3	C-T4	C-T5	C-T6
Pitch number	55	25	33	21	23	22
Description	high-level	high-rising	mid-level	low-falling	mid-rising	low-level
Example: transcription: character/ meaning	ji1(55): 医 ‘cure’	ji2(25): 椅 ‘chair’	ji3(33): 意 ‘idea’	ji4(21): 儿 ‘child’	ji5(23): 耳 ‘ear’	ji6(22): 二 ‘two’

Table 6. The Tone Inventory of Cantonese

As with Mandarin, there is a debate on whether Cantonese contour tones can be represented as a unitary entity or as a sequence of tone features. Barrie (2007) proposed that a unitary tone target account with only the tonal onset (i.e. [ $\pm$ upper]) and the direction of tone trajectories (i.e. [ $\pm$ raise]) specified. The tonal offsets are featurally unspecified, but are implied by featural specification, e.g. if the tonal onset is [-upper] and the direction is [+raise], it implies that the tonal offset is [+upper]. Lee (2012), on the other hand, argued that Cantonese contour tones should be specified as a sequence of two separate features, based on evidence from a tonal morpho-phonological process that induced a derived rising tone, for which he argued that the formal analysis requires the tonal offsets of contour tones to be fully specified for delinking a tone feature.

Acoustically, previous studies have shown that F0 height and F0 contour are the primary correlates for the tonal distinction in Cantonese (Khouw & Ciocca, 2007; Vance, 1977). C-T4(21) also has creakiness as an optional supplementary feature (Yu & Lam, 2014). Perceptually, listeners also rely on pitch height, pitch direction, and magnitude of change for tonal distinctions in Cantonese (Gandour, 1981; Khouw & Ciocca, 2007). The primary perceptual cue for the three level tones (C-T1(55), C-T3(33), C-T6(22)) is pitch height. Mok et al. (2013) reported the production of C-T1(55) by a native speaker of Cantonese was well separated from all other tones by being at the top of the pitch range. They also reported a larger

difference in F0 between C-T1(55) and C-T3(33) than between C-T3(33) and C-T6(22) (also see data in Khouw and Ciocca, 2007). C-T2(25) and C-T5(23) differed in the magnitude of change: C-T2(25) rose to the highest pitch level, whereas C-T5(23) only had a minimal rise. C-T4(21) and C-T6(22) differed slightly in contour with C-T4(21) having a small fall throughout the syllable.

The production results of the current study show some slight difference from the description in the previous literature (Figure 6). The onset F0 difference (around 10 Hz) between C-T3(33) and C-T6(22) was much smaller than the reported onset F0 difference (around 50 Hz) by the normal speaker (non-merger) in Mok et al. (2013). The offset F0 values between C-T5(23) and C-T6(22) merged at the end, whereas C-T5(23) has been reported to rise to a higher pitch than C-T6(22) at the end.

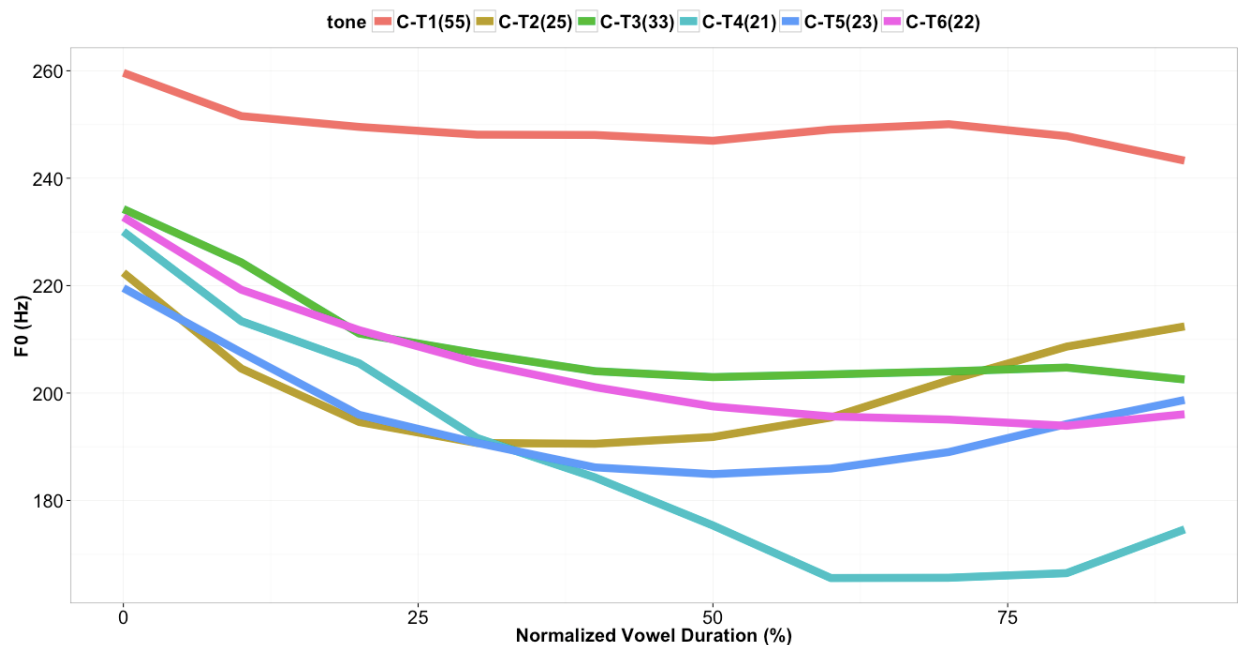


Figure 6. Average F0 values in Cantonese Tone contours within normalized vowels. The 11<sup>th</sup> end point is excluded due to the large amount of undefined measurement from Praat. The data were produced by 15 Cantonese female speakers.

Our general results may have captured the on-going tone mergers in Modern Cantonese reported in a number of recent studies (Fung & Wong, 2010; Mok, Zuo, & Wong, 2013; Mok & Wong, 2010). The reported tone merger in production and perception are C-T2(25) vs. C-T5(23), C-T3(33) vs. TC-6(22), C-T4(21) vs. C-T6(22). In our production data based on 15 female speakers, C-T3(33) is merged with TC-6(22), and no other cases of tone merger are found (see Chapter 4).

### **3. Methods**

#### **3.1 Introduction**

A series of production experiments were conducted to address the research questions repeated below:

Questions 1: Is C-F0 conditioned by lexical tones?

Questions 2: Is C-F0 conditioned by F0 distance between tones?

Questions 3: Is there a cue trading relation between F0 and VOT?

The following subsections present the methodology of the experiments. Sections 3.2 and 3.3 present the selection criteria of participants, the design of stimuli and the experimental procedure for the Mandarin group and the Cantonese group respectively. Section 3.4 describes the measurement for duration values and F0 values. Section 3.5 delineates the statistical tests for analyzing the data.

#### **3.2 Mandarin group**

##### **3.2.1 Participants of native Mandarin**

Fifteen female native speakers of Mandarin participated in the Mandarin production experiment. None of the participants reported hearing or speech disorders. To better control for phonetic and sociolinguistic variation related to gender, this study only reports the data from the female speakers. The age of the female Mandarin speakers ranged from 19 to 33. Two of them were graduate students from Michigan State University, and thirteen were recruited in Shenzhen city. They also varied in their degree of English proficiency. The participants originated from two regions of China: seven from Beijing and eight from Shenzhen. All participants were monolingual and only spoke standard Mandarin natively at home or work. They did not speak any other Chinese languages. Although these Mandarin speakers were from different regions, the

author, as a near-native speaker of Mandarin, has not observed any noticeable dialectal variations in their speech. The lack of regional differences could be for two reasons: Firstly, the participants had a relatively high education background – all were at or above university level. There may be a relation between the level of one’s education and how deaccented or standardized one’s Mandarin is. Secondly, the experiments were conducted in a formal context – reading sentences on the computer screen, so the participants were very conscious that they should speak standard Mandarin. Appendix A shows the demographic information of the participants.

### **3.2.2 Mandarin stimuli**

The selection of the Mandarin stimuli (n=79) was aimed at obtaining comprehensive data that adequately represent lexical tones, laryngeal types, place of articulation and vowel types (Appendix B). The speech materials covered four Mandarin lexical tones: M-T1(55), M-T2(35), M-T3(214) and M-T4(51). There were three groups of initial consonants that differed in three voicing types and three places of articulations: voiceless aspirated obstruents /t<sup>h</sup>, k<sup>h</sup>, p<sup>h</sup>/, voiceless unaspirated obstruents /t, k, p/ and sonorant consonants /l, n, m/.

The stimuli included four types of vowel types: /i, a, u, ə/. The wide range of vowel selection provides a basis for an investigation of the interaction of vowel types and C-F0, and enhances the possibilities to compare the same vowel(s) across different language groups. The three vowels /i, a, u/ followed bilabial and alveolar consonants, and the two vowels /u, ə/ followed velars. The syllables /k<sup>h</sup>a, ka/ were not selected, for there are many accidental gaps, i.e. /\*k<sup>h</sup>a2(35), \*k<sup>h</sup>a4(51), \*ka2(35), \*ka3(214)/, making it not possible to represent all four lexical tones. The syllables /k<sup>h</sup>i, ki/ were also excluded since they are disallowed by Mandarin phonotactics – velars cannot be followed by high front vowels.

All Mandarin stimuli were represented in Chinese characters and were CV syllables. Each stimulus was embedded in the carrier phrase “wɔ3(214) ʂwɔ1(55) \_\_\_\_\_ tsu4(51) san1(55) ts<sup>h</sup>i4(51)”, meaning ‘I say \_\_\_\_\_ character for three times’.

In total, 3555 tokens (79 words × 3 repetitions × 15 participants) were gathered from the native Mandarin speakers. However, 3375 tokens were measured, and 180 were excluded due to participants’ mispronunciation and poor recording quality identified by the author during the manual segmentation process, or elicited as undefined values by PRAAT measurement (Boersma & Weenink, 2015).

### **3.2.3 Procedure for the Mandarin experiment**

The experiments took place in two locations: in the Phonology-Phonetics Experimental lab at the linguistics department at Michigan State University, and in a soundproof studio in Shenzhen city. Recordings were collected using a Logitech Desktop microphone (frequency response – 100Hz-16KHz) and a MacBook laptop. There were instructions and a practice experiment before the real experiment. Stimuli were randomly presented through PsychoPy (Peirce, 2007), which recorded participants’ response at a 48,000 Hz sampling rate, with a 24-bit resolution in 1-channel. Participants were asked to produce three repetitions of each sentence presented on the screen at a normal pace. Each sentence was presented for 8 seconds and participants had to finish the three repetitions within that amount of time. Thus, the speech rate was relatively controlled.

## **3.3 Cantonese group**

### **3.3.1 Participants of Cantonese**

Fifteen female native speakers of Cantonese participated in the Cantonese experiment (Appendix A). None of them reported hearing or speech disorders. As with the Mandarin data, this study only reports the production data from the female Cantonese speakers, in order to control for the

phonetics and sociolinguistic variation related to gender. The age of the Cantonese female speakers ranged from 20 to 35. Three of them were graduate students from Michigan State University, and the other twelve were recruited in Shenzhen city. All speakers originated from Shenzhen city. They all spoke Cantonese natively and Mandarin near-natively. None of the participants spoke any other Chinese dialects, and they varied in their proficiency of English. The author, as a native speaker of Cantonese, did not notice any noticeable dialectal differences of the speakers' Cantonese pronunciation. It is also worth noting that the Cantonese participants were not recruited from Guangzhou, where Guangzhou Cantonese is spoken. This is so that there is a better control of the high-level tone contexts between Mandarin and Cantonese: Guangzhou Cantonese speakers have two allotones for C-T1(55) – a high-falling tone (53)<sup>11</sup> and a high-level tone (55) (Bauer & Benedict, 1997; Hashimoto, 1972; Yip, 2002), but this study is interested in comparing the high-level contexts between Mandarin and Cantonese, so the allotonic variants in Guangzhou Cantonese could have been a confounding factor. Cantonese speakers in Shenzhen do not have the high-falling allotone (53), which may have been influenced by Hong Kong Cantonese that has lost the high-level and high-falling distinction.

### **3.3.2 Cantonese stimuli**

As with the selection of the Mandarin experimental stimuli, the Cantonese stimuli were selected to be adequately comparable to represent Cantonese lexical tones, voicing types, place of articulation and vowel types (Appendix C). The speech materials covered six Cantonese lexical tones: C-T1(55), C-T2(35), C-T3(33), C-T4(21), C-T5(23) and C-T6(22). There are three groups

---

<sup>11</sup> The high falling tone is also documented in a tone sandhi environment (Bauer & Benedict, 1997; Hashimoto, 1972). A high falling (53) becomes a high level (55) before a high falling or a high level.

of initial consonants that differ in three voicing types and three places of articulations: voiceless aspirated obstruents /t<sup>h</sup>, k<sup>h</sup>, p<sup>h</sup>/, voiceless unaspirated obstruents /t, k, p/ and sonorants /l, n, m, ŋ/.

A list of 77 words was recorded, including four types of vowels: /a, ɔ, e, o/. Cantonese phonotactics prevent unaspirated onset consonants /t, k, p/ from occurring in syllables with tones C-T5(23) or C-T4(21). Also, aspirated onsets /t<sup>h</sup>, k<sup>h</sup>, p<sup>h</sup>/ do not appear with C-T6(22). Besides these systematic gaps due to phonotactics, there can also be accidental gaps that are possible words, but are not present in the Cantonese lexicon (Kirby & Yu, 2007). Due to a large number of lexical gaps for the co-occurrence of different lexical tones and segments, the selection of stimuli was designed in such a way that voicing type, place of articulation (including alveolar, velar and bilabial) and vowel type (including low vs. mid and round vs. unrounded categories) had comparable and adequate representations in the data to study C-F0. All Cantonese stimuli were represented in Chinese characters and carried CV syllables. Each stimulus was in embedded in the carrier phrase “ŋɔ5(23) kəŋ2(35) \_\_\_\_\_ tsi6(22) sam1(55) ts<sup>h</sup>i3(33)”, which means ‘I say \_\_\_\_ character for three times’.

In total, 3465 tokens (77 words × 3 repetitions × 15 participants) were collected from the Cantonese speakers. However, 3322 tokens were measured and 143 were not included during the segmentation process due to mispronunciation and poor recording quality, or undefined values in PRAAT measurement.

### **3.3.3 Procedure for the Cantonese experiment**

The locations and equipment for the Cantonese experiment were the same as the Mandarin ones. Instructions were given in Cantonese by the author, who is a native speaker of Cantonese. Participants were also asked to produce three repetitions of each sentence presented on the screen



at a normal pace. They had to finish the three repetitions within 8 seconds. There was a practice session before the actual experiment.

### **3.4 Measurements**

#### **3.4.1 Segmentation labeling and measurements of durations**

All segmentation was done by the author manually in PRAAT. The boundaries of parts of vowels and consonants were primarily determined by visual inspection of the spectrogram in PRAAT. A secondary criterion for the segmentation, especially for the boundaries between consonant closures and releases, was to look at waveforms, which presents information of amplitudes of the segments, improving the accuracy of durational measurement (Turk et al., 2006).

Figures 7 and 8 respectively show the segmentation of the Mandarin word p<sup>h</sup>a1(55) with a voiceless aspirated onset, and the word pa1(55) with a voiceless unaspirated onset. The consonant closure duration, labeled as “c” in the second tier, was marked from the end of the preceding segment in the carrier phrase until the beginning of the release, i.e. where the burst began in the spectrogram and the waveform. The VOT was labeled as “r” in the second tier, measured from the point right after the oral closure until the onset of voicing of the following vowel, which is labeled as the vowel itself (e.g. “a” in Figures 7 and 8) in the first tier. The onsets of the first and the second formants of the vowels in the spectrogram were used to determine the beginning points of vowels, and the offsets of the first and the second formants define the ending points of vowels.

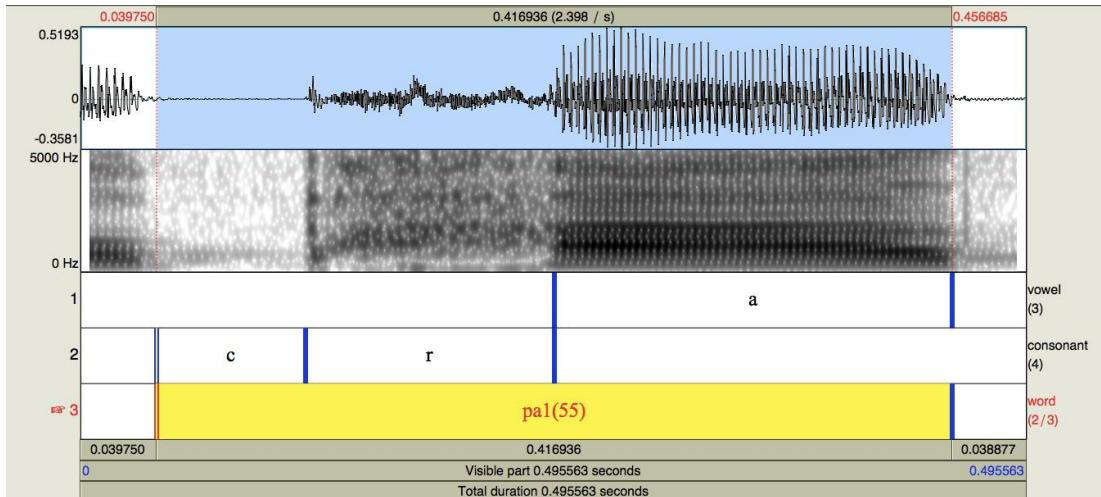


Figure 7. segmentation for  $p^h1(55)$  in Mandarin. ‘c’ = closure; ‘r’ = release; ‘a’=the vowel /a/

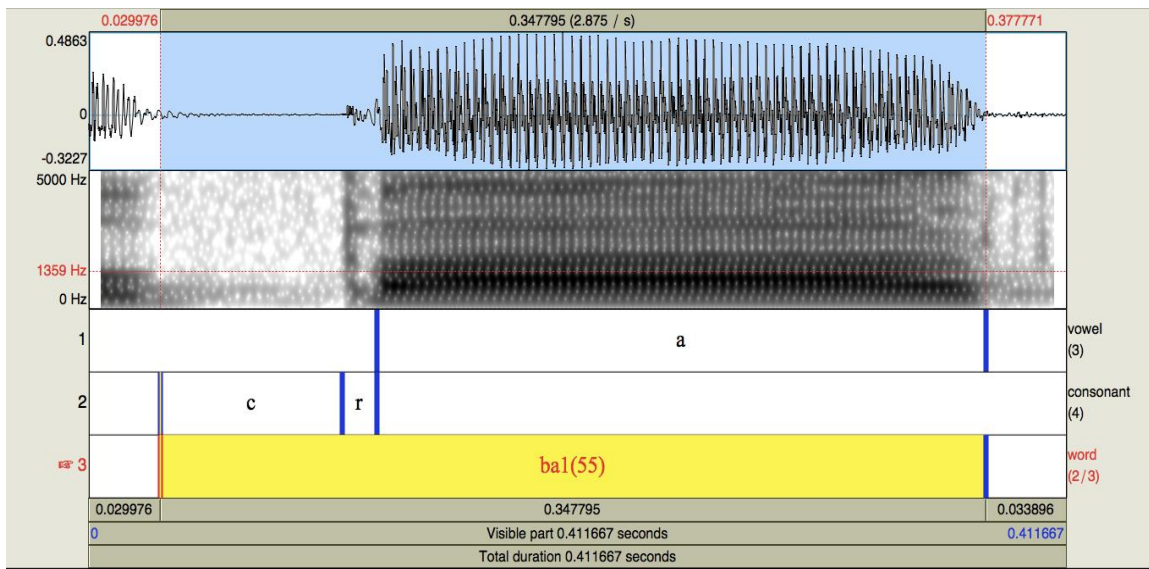


Figure 8. segmentation for  $pal(55)$  in Mandarin. ‘c’ = closure; ‘r’ = release; ‘a’=the vowel /a/

Figure 9 shows how the word  $mal(55)$  with a sonorant onset was segmented. The sonorant consonant was marked from the end of the preceding segment in the carrier until the voicing onset of the vowel.

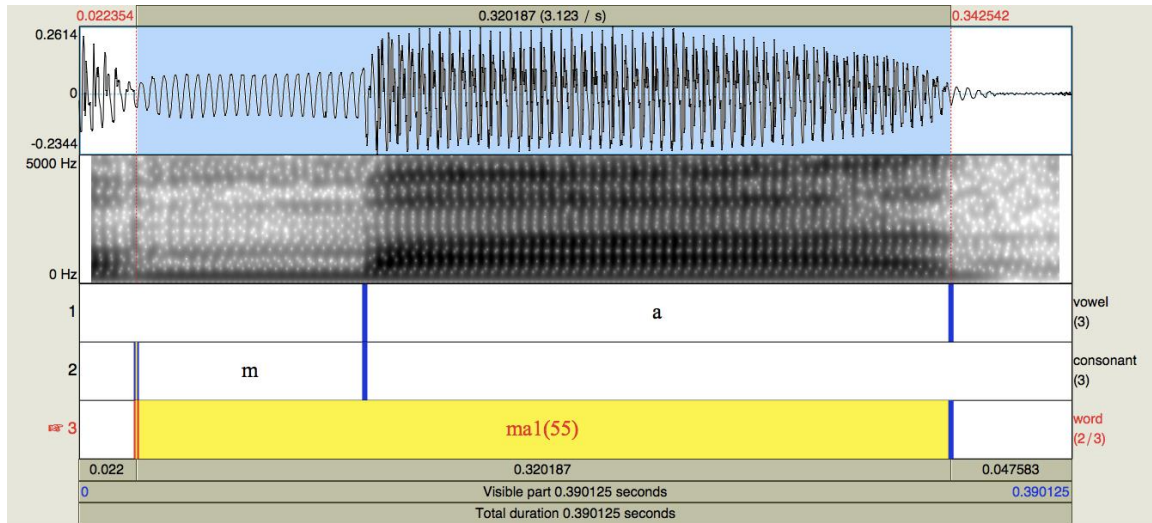


Figure 9. segmentation for mal(55) in Mandarin. ‘m’ = nasal /m/; ‘a’=the vowel /a/

A PRAAT script created by Karthik Durvasula and myself was used to track the labeled target and extract values of closure durations, VOT and vowel durations. The same script was used to yield values of F0 values, which will be elaborated in the following sub-sections.

### 3.4.2 Measurements of F0

F0 values were exacted with a 500 Hz pitch ceiling and at every 5ms within 100 ms along the F0 trajectory of the vowels. All F0 values were then normalized into z-scores (further discussed below) to reduce between-speaker and between-token variance (Rose, 1987). Two kinds of F0 values will be analyzed: (1) within absolute duration: studies on pitch perturbation have found that the most prominent perturbation effects are expected within the first 100ms in non-tonal languages, and first the first 30-50 ms in tonal languages (Francis et al., 2006). Based on these previous findings, the mean normalized F0 values following the voicing onset were analyzed within the first 30 ms of the vowels; (2) within normalized duration: it is possible that each speaker has different vowel lengths and speech rates, which give rise to different tonal alignments (Xu 1998; 2001). If the C-F0 is influenced by articulatory factors such as timings of gestures, it is important to also investigate the time-normalized F0 values. Time-normalized F0

values were measured at 11 equal-distant points of the vowel, between 0%-100%, in steps of 10%.

### **3.5 Statistic analysis**

Prior to statistical analysis, by-speaker raw F0 and duration values were transformed into z-scores using the `scale()` function in R (R Core Team, 2015), to facilitate comparison of total degree of pitch change across subjects and tokens. The above normalization was by subject.

For Question 1, on whether the F0 difference and direction C-F0 is conditioned by F0 lexical tones, the analysis was conducted on F0 values after three CONSONANT types (i.e. aspirated vs. unaspirated vs. sonorant) at each TIMEPOINT (i.e. 0, 5, 10, 15, 20, 25 and 30 ms) of the vowel in different TONE contexts (i.e. six Cantonese tones and four Mandarin tones) in each LANGUAGE (i.e. Mandarin vs. Cantonese). Within-subject Repeated Measures Analyses were conducted with CONSONANT, TIMEPOINT, TONE and LANGUAGE as the independent fixed-effect variables. The effects were checked for significance at the thresholds  $\alpha=0.05$ , 0.01 and 0.001.

For Question 2, on whether C-F0 is influenced by distance between tones, the analysis was run on TONE DIFF (i.e. F0 difference between tones) and C-F0 DIFF (i.e. F0 difference between consonant types) in each LANGUAGE (i.e. Mandarin vs. Cantonese). Within-subject Repeated Measures Analyses were conducted with TONE DIFF, C-F0 DIFF, TIMEPOINT and LANGUAGE as the independent fixed-effect variables. The effects were checked for significance at the thresholds  $\alpha=0.05$ , 0.01 and 0.001.

For Question 3, ANOVA and follow-up t-tests were run to investigate whether VOT and onset F0 within the first 10ms of the vowel are cues to the voicing contrast. Later, linear regression was run on normalized VOT values and normalized onset F0 values following different consonant categories (i.e. aspirated and unaspirated) in Mandarin and Cantonese.

## **4. Results for Question 1: C-F0 conditioned by tones?**

### **4.1 Introduction**

This chapter addresses Research Question 1 stated in Chapter 1: is C-F0 conditioned by tones? More specifically, are the magnitudes and the directions of C-F0 influenced by different tones in Mandarin and Cantonese? Do tone properties (i.e. pitch height and contour shape) have consistent effects on the patterns of C-F0?

Two dimensions of C-F0 are evaluated. (1) The first one is its magnitude, which has two indicators: C-F0 LENGTH (i.e. how far in the vowels C-F0 can extend) and C-F0 DIFF (i.e. how big the differences between F0 values after different types of consonants are). (2) The second dimension is C-F0 DIRECTION: that is, to compare the values of F0[aspirated] or F0[unaspirated] and the values of F0[sonorant] baseline. C-F0 is in a lowering direction if F0 values after obstruents are lower than the baseline. On the contrary, it is in a raising direction when F0 values after obstruents are higher than the baseline. To preview the results in this chapter, this chapter will show that tones can influence C-F0 LENGTH and C-F0 DIFF in both Mandarin and Cantonese. Moreover, there is also evidence showing that tones may condition C-F0 DIRECTION, but only in Cantonese, and not in Mandarin.

The results of Mandarin are shown in Section 4.2: Section 4.2.1 confirms that four distinctive Mandarin tones were produced. Section 4.2.2 presents the findings of C-F0 in each Mandarin tone, suggesting that the magnitudes of C-F0 are conditioned by Mandarin tones: In terms of C-F0 LENGTH, the effect is longer in M-T1(55) and in M-T4(51) than in M-T3(214). However, the max C-F0 DIFF is slightly bigger for M-T3(214) than for M-T1(55) and M-T4(51). No statistically significant C-F0 is found in M-T2(35). Nevertheless, the general C-F0 DIRECTION does not vary in different Mandarin tone contexts. A general aspiration raising

effect is found across all Mandarin tones. Section 4.2.3 provides a summary for the Mandarin data.

The results of Cantonese are reported in Section 4.3: Section 4.3 finds that five distinctive Cantonese tones were produced, with the merger of C-T3(33) and C-T6(22). Section 4.2.2 presents the findings of C-F0 in each Cantonese tone, suggesting that both the magnitudes and direction of C-F0 are conditioned by Cantonese tones: In terms of C-F0 LENGTH, the effect is longer in C-T1(55) and C-T2(25) than in C-T3(33) and C-T5(23). No statistically significant C-F0 is found in C-T4(21). As for C-F0 DIFF, the maximum F0 difference is slightly smaller for C-T2(25) and C-T5(23) than for C-T1(55) and C-T3(33). The general C-F0 DIRECTION also varies in different Cantonese tone contexts. F0[unaspirated] is in the middle in C-T1(55) and C-T3(33), but the lowest in C-T2(25). Section 4.3.3 provides a summary for the Cantonese data.

A summary of the results in both languages is provided in Section 4.4.

## **4.2 C-F0 conditioned by tones in Mandarin?**

Tone trajectories in Mandarin are examined in Section 4.2.1 to confirm that participants produced four distinctive Mandarin tones. Section 4.2.2 examines whether C-F0 is conditioned by tones in Mandarin. The results show that the magnitudes of C-F0, i.e. C-F0 LENGTH and C-F0 DIFF, are conditioned by Mandarin tones, but C-F0 DIRECTION is not. Section 4.2.3 provides a section summary.

### **4.2.1 Overall Mandarin tone trajectories**

The F0 values are extracted at the beginning of the vowel and every subsequent 10% of the vowel. The F0 values at the endpoints of the vowels are excluded due to a large amount of undefined measurements. The raw and normalized F0 values were averaged for each of the

fifteen Mandarin participants at each of the time points for the four tones. The normalized F0 values is the z-scores by subject using the following function:

$$z = (x - \mu) / sd$$

where  $z$  is a z-score,  $x$  is the raw F0 values,  $\mu$  is the mean raw F0 values by subject, and  $sd$  is the standard deviation by subject.

In the following reports of the results, the F0 of the first 90% will be shown in raw values so it can provide a reference compared to Mandarin tone trajectories in previous studies that were mostly described in raw values. The reports and analysis of C-F0 will be conducted with normalized z-scored F0, which makes cross-linguistic comparisons more accessible.

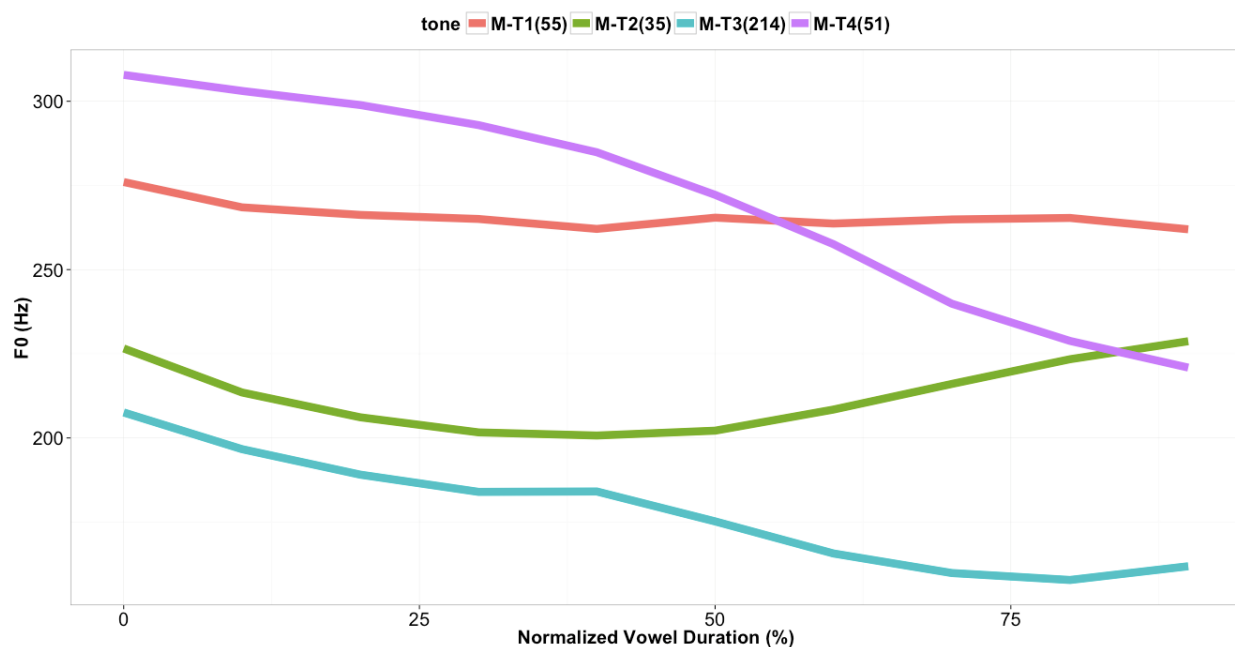


Figure 10. Four Mandarin tone contours within vowels in normalized duration

Figure 10 shows the raw F0 values (in Hz) for the four Mandarin tones within the first 90% of the vowels. The F0 values of M-T1(55) maintain a high level. M-T4(51) starts higher than the high-level M-T1(55) and ends at a similar F0 height as M-T2(35). The first half of M-T2(35) and M-T3(214) are very close and they start diverging after about 40% of the duration.

As mentioned earlier, the incomplete M-T3(214) can be due to the fact that all target syllables are followed by a zi4(51) syllable (to form the meaning “\_\_\_\_\_ character”), creating a disyllabic context that triggers the ‘half-third’ sandhi condition where the pitch 214 becomes 21 when followed by any Mandarin tone except M-T2(35) (Zhang & Lai, 2010).

Tone	Time Value	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
M-T1(55)	Mean	275.0	268.2	265.5	264.2	261.6	264.4	263.0	264.2	265.0	263.3
	SD	35.9	34.1	34.4	34.5	37.3	35.7	34.6	36.9	36.1	34.5
M-T2(35)	Mean	225.2	214.4	206.1	201.7	200.7	202.1	208.4	216.0	223.4	229.1
	SD	27.2	27.5	28.3	28.7	28.3	28.1	24.5	23.5	24.1	23.3
M-T3(214)	Mean	204.0	198.6	195.6	189.1	181.4	173.0	168.4	161.7	160.5	157.8
	SD	10.9	10.3	18.0	21.0	27.0	27.1	29.3	23.6	25.5	23.0
M-T4(51)	Mean	308.5	303.1	298.9	292.9	284.8	272.2	257.6	239.4	227.2	219.7
	SD	36.4	33.2	30.3	26.1	23.3	20.1	19.5	26.1	31.8	34.4
ANOVA	F(3, 42)	30.02	62.73	90.32	92.92	74.61	66.02	47.43	43.94	36.11	29.45
	P	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		***	***	***	***	***	***	***	***	***	***
	ges	0.74	0.79	0.71	0.73	0.70	0.70	0.67	0.66	0.63	0.60

Table 7. Mean F0 values(Hz) and the standard deviation at every 10% of the vowel. The lower portion of the table includes results of ANOVAs at each time point where the dependent variable was the mean F0, and the within-subjects independent variable was TONE (4 tones).

Table 7 shows the mean F0 values and the standard deviation at every 10% of the total vowel duration. ANOVA tests show that at each 10% time point where the dependent variable was the mean F0, and the within-subjects independent variable was TONE, the mean F0 values were significantly different among the four tones at each 10% time point.



Tone Pair	Time Value	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
T1-T4	t (14)	-6.82	-7.23	-6.42	-5.35	-3.73	-1.64	0.32	2.06	3.12	3.12
	p	<0.001***	<0.001***	<0.001***	<0.001***	0.006**	0.142	0.771	0.065	0.016*	0.016*
T2-T3	t(14)	-0.45	-0.66	0.33	1.42	1.62	2.03	3.34	7.02	7.71	8.14
	p	0.736	0.592	0.792	0.194	0.129	0.068	0.014*	<0.001***	<0.001***	<0.001***

Table 8. T-tests for Tone Pair M-T1(55) vs. M-T4(51) and M-T2(35) vs. M-T3(214) at every 10% of the vowel.

Table 8 presents the results of the following t-tests for two tone pairs that are most likely to merge: M-T1(55) vs. M-T4(51) and M-T2(35) vs. M-T3(214). The two pairs are selected among all possible comparisons between two tones because native and non-native speakers have high confusion rates for these two tone pairs, as observed in previous literature (Huang, 2001; Wang et al., 2006; Hao, 2012; Li, 2016). The results show that M-T1(55) and M-T4(51) were significantly different except between 50% - 70% of the vowel. M-T2(35) and M-T3(214) were significantly different starting from 60% till the end. The statistical tests indicated that the participants produced four distinctive Mandarin tones.

## 4.2.2 C-F0 in Mandarin tones

### 4.2.2.1 General results

Figure 11 shows the overall normalized F0 values within the first 50ms after different types of consonants in each Mandarin tone. The F0 difference following aspirated and sonorant in M-T2(35) is smaller than the other three tones. The immediate following four subsections 4.2.2.2 – 4.2.2.5 will present the results in each tone in more details.

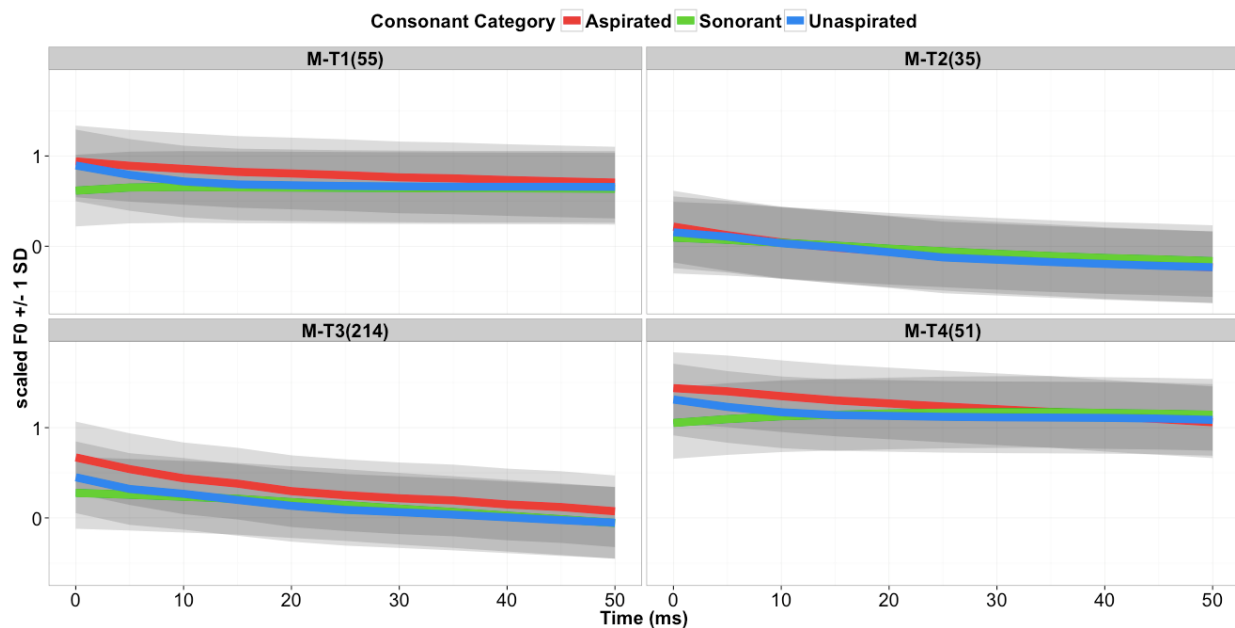


Figure 11. Normalized F0 trajectories within the first 50ms of the vowels in four Mandarin tones. Shading shows 1 SD around the mean

Table 9 presents the results of ANOVA tests examining whether the within-subjects independent variables CONSONANT, TONE and VOWEL influenced the dependent variable, normalized F0 values by subjects. The results show that CONSONANT had a significant effect on F0 from the beginning until the first 20ms. By contrast, VOWEL significantly influenced F0 from 15ms until the end of the target range. Not surprisingly, TONE had a crucial impact on F0 for the entire 35ms. None of the interactions were significant.

The results have captured the general observation that C-F0 usually does not extend far in tonal languages. The later appearance of vowel effects may indicate that Mandarin speakers have a later tongue body movement, which can pull mechanically on the vocal folds via the hyoid bone and change the vocal fold tension and rate of vibration (Kingston, 2007). Further study can be done to see whether this is a language-specific observation about Mandarin without any direct articulatory explanation.

Duration	0ms		5ms		10ms		15ms	
	F	P	F	P	F	P	F	P
C	13.30	<0.001** *	9.74	<0.001** *	6.29	<0.001** *	4.66	0.009 *
T	75.68	<0.001** *	108.60	<0.001** *	135.55	<0.001** *	150.51	<0.001** *
V	1.83	0.141	2.06	0.104	2.21	0.085	2.80	0.039 *
C*T	0.47	0.830	0.78	0.583	0.57	0.748	0.54	0.776
C*V	1.62	0.152	1.62	0.152	1.51	0.182	1.91	0.089
T*V	1.86	0.056	0.93	0.495	0.88	0.538	1.02	0.418
C*T*V	0.57	0.895	0.34	0.990	0.30	0.994	0.36	0.988

Duration	20ms		25ms		30ms		35ms	
	F	P	F	P	F	P	F	P
C	3.19	0.041 *	2.12	0.119	1.45	0.234	1.11	0.327
T	166.08	<0.001** *	170.26	<0.001** *	172.76	<0.001** *	176.65	<0.001** *
V	2.70	0.044 *	3.22	0.022 *	4.11	0.007 **	4.32	0.004 **
C*T	0.39	0.885	0.33	0.936	0.34	0.911	0.44	0.847
C*V	1.60	0.157	1.61	0.155	1.10	0.356	0.84	0.520
T*V	1.01	0.430	1.02	0.414	1.35	0.204	1.46	0.156
C*T*V	0.33	0.992	0.48	0.947	0.43	0.970	0.40	0.977

Table 9. ANOVA where the independent variables are CONSONANT, TONE and VOWEL, and the dependent variable is normalized F0 values in Mandarin.

T = Tone; C= Consonant; V = Vowel

#### 4.2.2.2 C-F0 in M-T1(55)

Figure 12 shows the normalized F0 values at every 5ms within the first 35ms of vowels that carry the high-level M-T1(55). F0[aspirated] is higher than F0[unaspirated] and F0[sonorant].

F0[unaspirated] falls from high pitch and converge with F0[sonorant] after 10ms.

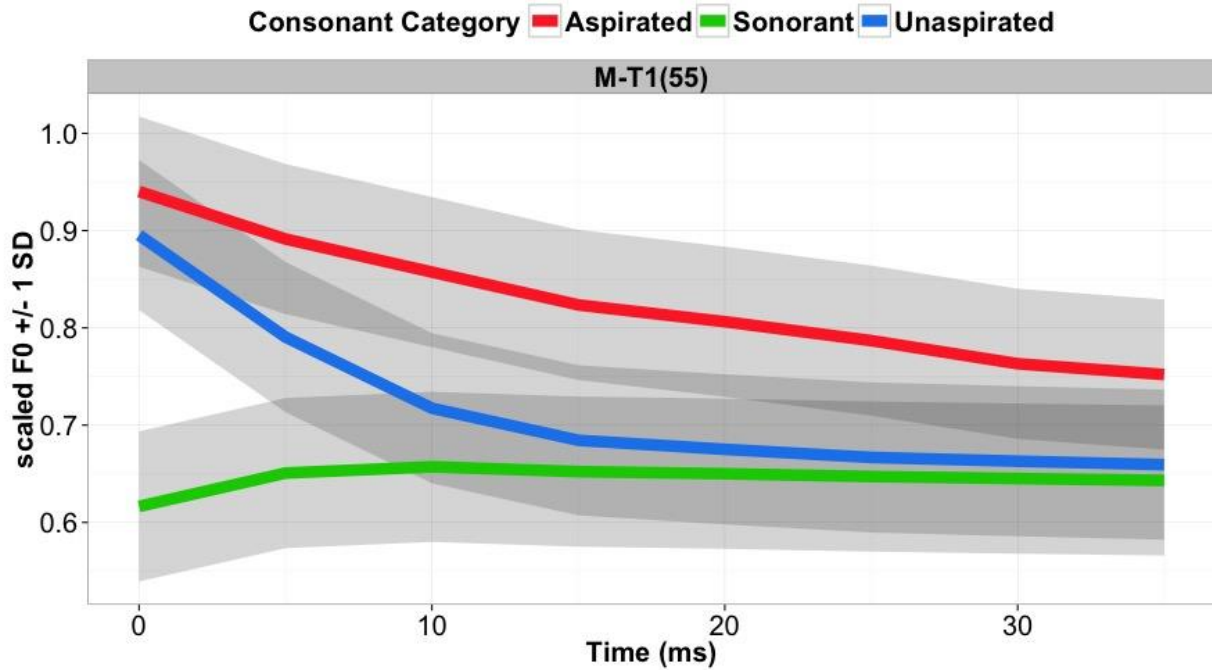


Figure 12. Normalized F0 within the first 35ms of the vowel in M-T1(55). Shading shows 1 SD around the mean value

Table 10 presents ANOVA and t-tests for the F0 difference in M-T1(55). ANOVA was conducted at every 5ms where the dependent variable was the mean normalized F0, and the within-subjects independent variable was CONSONANT. Significant F0 difference was found after the three types of initial consonants at every time point of the 35ms. T-tests showed that the difference between F0[unaspirated] and F0[sonorant] was only crucial within the first 5ms. F0[aspirated] was significantly higher than F0[sonorant] and F0[unaspirated] during the entire 35ms.

The largest mean C-F0 DIFF(T-N) (i.e. difference between F0[unaspirated] and F0[sonorant]) was at 0ms (MeanDiff<sub>zscore</sub>=0.29, MeanDiff<sub>raw</sub>=18.08Hz). The largest mean C-F0 DIFF(TH-N) (i.e. difference between F0[aspirated] and F0[sonorant]) was at 0ms (MeanDiff<sub>zscore</sub>=0.34, MeanDiff<sub>raw</sub>=25.14Hz). The max C-F0 DIFF(T-TH) (i.e. difference

between F0[unaspirated] and F0[aspirated]) was at 10ms (MeanDiff<sub>zscore</sub>=0.19, MeanDiff<sub>raw</sub>=11.67Hz).

Duration	0ms			5ms			10ms		
ANOVA	F	P		F	P		F	P	
	(2, 28) =17.41	<0.001***		(2, 28) =13.45	<0.001***		(2, 28) =10.57	<0.001***	
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	3.61	0.002**	0.29	2.06	0.058*	0.27	0.89	0.386	0.03
TH – N	5.86	<0.001***	0.34	4.78	<0.001***	0.27	4.11	<0.001***	0.21
T – TH	5.85	<0.001***	-0.16	3.09	0.007**	-0.18	3.39	0.004* *	-0.19
Duration	15ms			20ms			25ms		
ANOVA	F	P		F	P		F	P	
	(2, 28) =8.15	0.002**		(2, 28) =6.99	0.003* *		(2, 28) =5.96	0.007* *	
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	0.48	0.638	0.02	0.31	0.759	0.01	0.22	0.821	0.01
TH – N	3.36	0.004**	0.18	2.96	0.010* *	0.16	2.71	0.010* *	0.14
T – TH	3.25	0.005**	-0.16	3.24	0.005* *	-0.15	3.18	0.007* *	0.14
Duration	30ms			35ms					
ANOVA	F	P		F	P				
	(2, 28) =4.59	0.019* *		(2, 28) =3.93	0.031* *				
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.			
T – N	0.18	0.856	0.01	0.147	0.884	0.00			
TH – N	2.41	0.030* *	0.12	2.21	0.044* *	0.11			
T – TH	2.86	0.012* *	0.11	2.70	0.017* *	0.11			

Table 10. In the context of M-T1(55), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded.

T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant

The results indicate a raising effect from aspiration and voiceless stops in general in the high-level M-T1(55). This is inconsistent with the findings reported in Xu & Xu (2003), where a significant aspiration raising effect is not found. Instead, they observe that F0 values at the voice onset are higher for unaspirated than for aspirated across the four tone conditions, though the difference in M-T1(55) was not observable. However, their results are based on the average F0 within 105ms from the syllable onsets, so the effect magnitudes may have been washed away due to the long duration.

#### 4.2.2.3 C-F0 in M-T2(35)

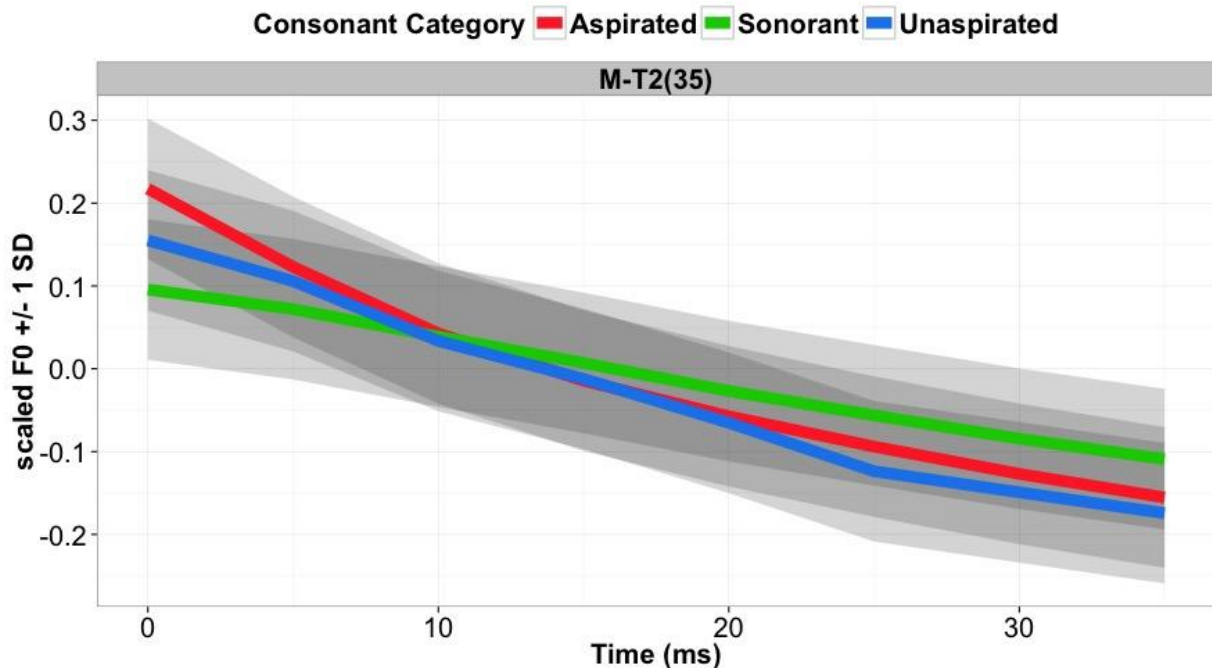


Figure 13. Normalized F0 within the first 35ms of the vowel in M-T2(35). Shading shows 1 SD around the mean.

Figure 13 shows the normalized F0 within the first 35ms of the vowels that carry M-T2(35). The F0 difference among the three consonant types is smaller than that shown in M-T1(55). Within the first 10ms, F0[aspirated] is slightly higher than F0[unaspirated], which is in turn slightly higher than F0[sonorant].

Duration	0ms		5ms		10ms		15ms	
ANOVA	F	P	F	P	F	P	F	P
	(2, 28) =0.41	0.668	(2, 28) =0.65	0.526	(2, 28) =0.41	0.667	(2, 28) =0.77	0.472
Duration	20ms		25ms		30ms		35ms	
ANOVA	F	P	F	P	F	P	F	P
	(2, 28) =1.11	0.343	(2, 28) =2.11	0.140	(2, 28) =2.59	0.092	(2, 28) =2.86	0.07

Table 11. In the context of M-T2(35), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values.

Though the C-F0 DIRECTION in the context of M-T2(35) is the same as in M-T1(55), ANOVA tests in Table 11 did not reveal any statistically significant F0 difference after the three consonants.

The results indicate that C-F0 is not robust in M-T2(35), which is also contradictory to the findings in Xu & Xu (2003). They found F0[aspirated] was significantly lower than F0[unaspirated] in the Mandarin rising tone context, which they explained was due to the constant subglottal pressure decreases markedly during the aspiration period. The inconsistency between our results and Xu & Xu's may be due to that different durations were studied: they were studying the average F0 within 105ms from the syllable onsets, so the relatively stronger C-F0 in M-T2(35) may be due to F0 difference accumulates as longer duration is involved. However, there was not an obvious accumulative F0 difference in M-T2(51) in our data – the F0 difference did not seem to get obviously bigger later in the vowels.

#### 4.2.2.4 C-F0 in M-T3(214)

Figure 14 shows the normalized F0 values in M-T3(214). F0[aspirated] is higher than F0[unaspirated] and F0[sonorant]. F0[unaspirated] starts from the middle and rapidly converge with F0[sonorant].

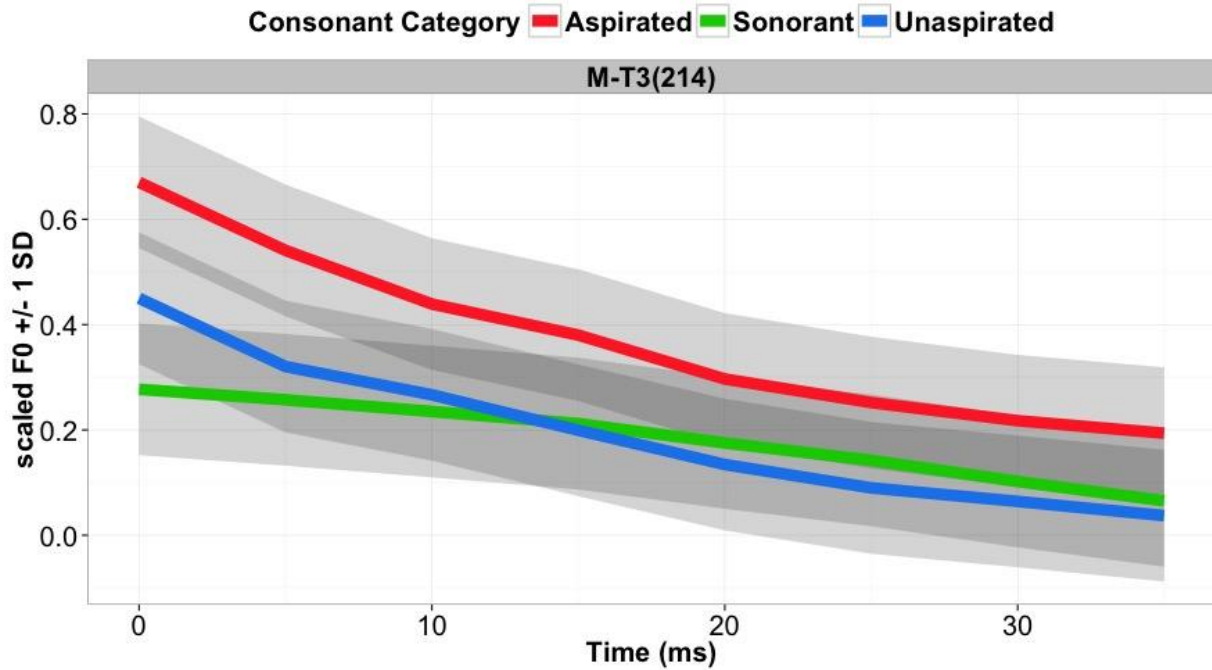


Figure 14. Normalized F0 trajectories within the first 50ms of the vowel in M-T3(214). Shading shows 1 SD around the mean.

Table 12 presents ANOVA and t-tests for the F0 difference in M-T3(214). ANOVA found that overall C-F0 LENGTH for M-T3(214) is 15ms. That is, a significant F0 difference after the three types of initial consonants was found within the first 15ms. T-tests found no crucial difference between F0[unaspirated] and F0[sonorant]. F0[aspirated] was significantly higher than F0[sonorant] within 5ms. F0[aspirated] was also significantly higher than F0[unaspirated] between 10-15ms. The C-F0 DIRECTION of M-T3(214) in an early period was the same as in M-T1(55) and M-T2(35).

The max mean C-F0 DIFF(TH–N) (i.e. difference between F0[aspirated] and F0[sonorant]) was at 0ms (MeanDiff<sub>zscore</sub>=0.55, MeanDiff<sub>raw</sub>=41.21Hz). The max C-F0 DIFF(T–TH) (i.e. difference between F0[unaspirated] and F0[aspirated]) was at 10ms (MeanDiff<sub>zscore</sub>=0.28, MeanDiff<sub>raw</sub>=18.35Hz).



Duration	0ms			5ms			10ms		
ANOVA	F	P		F	P		F	P	
	(2, 26) =3.56	0.043*		(2, 28) =3.87	0.033*		(2, 28) =3.96	0.030*	
T-tests Pairs	t (13)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	0.69	0.502	0.08	0.52	0.611	0.03	0.13	0.897	0.00
TH – N	2.20	0.046*	0.55	2.09	0.054*	0.38	1.96	0.070	0.27
T – TH	1.76	0.100	–0.25	1.94	0.071	–0.25	2.10	0.054*	–0.28
Duration	15ms			20ms			25ms		
ANOVA	F	P		F	P		F	P	
	(2, 28) =3.26	0.053*		(2, 28) =2.71	0.083		(2, 28) =2.03	0.149	
T-tests Pairs	t (14)	P	Mean Diff.						
T – N	1.27	0.225	0.04						
TH – N	1.55	0.142	0.18						
T – TH	2.10	0.053*	–0.22						
Duration	30ms			35ms					
ANOVA	F	P		F	P				
	(2, 28) =1.77	0.188		(2, 28) =1.50	0.241				

Table 12. In the context of M-T3(214), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded.

T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant

The pattern of C-F0 in M-T3(214) are very similar to that in M-T1(55), though C-F0 LENGTH is shorter in the former than the latter. However, the max C-F0 DIFF(TH–N) and the max C-F0 DIFF(T–TH) are higher in M-T3(214) than in M-T1(55). The result is still inconsistent to the findings in Xu & Xu (2003), where a significant aspiration lowering effect is not found in M-T3(214).

#### 4.2.2.5 C-F0 in M-T4(51)

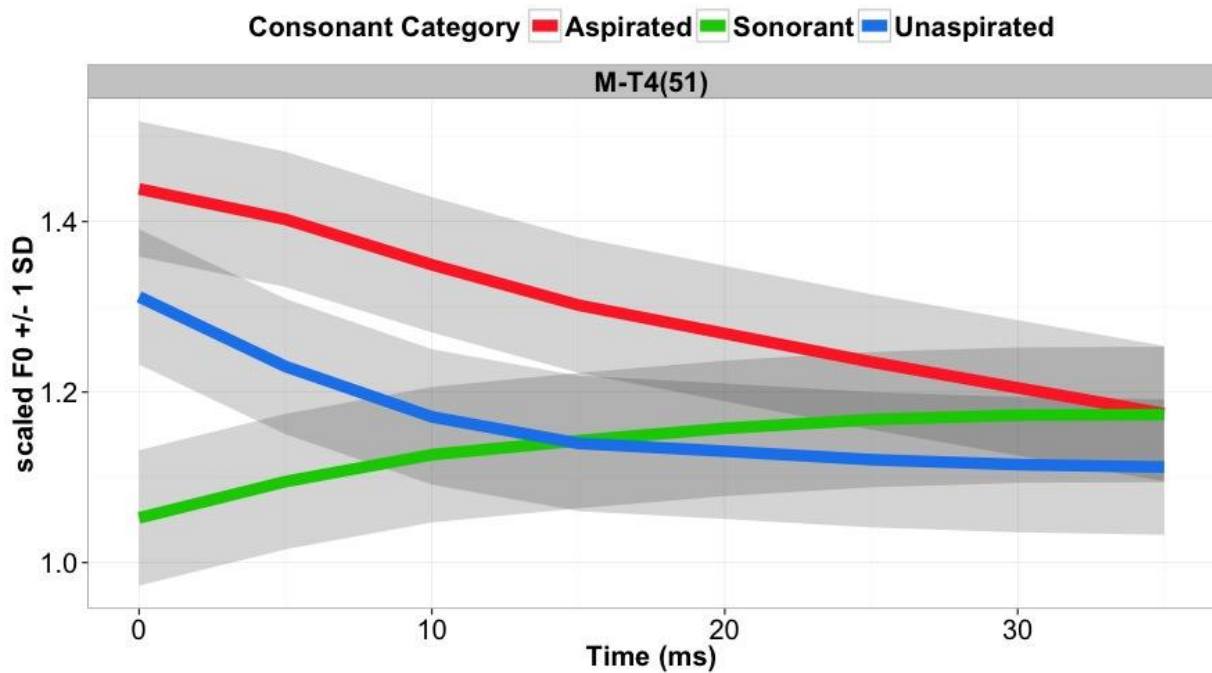


Figure 15. Normalized F0 trajectories within the first 50ms of the vowels in M-T4(51). Shading shows 1 SD around the mean

Figure 15 shows the normalized F0 values within the first 35ms of the high falling M-T4(51). The C-F0 DIRECTION is very similar to that in the other three tones. That is, F0[aspirated] is higher than F0[unaspirated] and F0[sonorant]. F0[unaspirated] starts from the middle and converge with F0[sonorant] after 10ms.

In Table 13, the ANOVA found that overall C-F0 LENGTH for M-T4(51) was 30ms. T-tests showed that the difference between F0[unaspirated] and F0[sonorant] was only crucial within the first 5ms. F0[aspirated] was significantly higher than F0[sonorant] within 20ms and than F0[unaspirated] within 30ms.

Duration	0ms			5ms			10ms		
ANOVA	F	P		F	P		F	P	
	(2, 28) =24.97	<0.001 ***		(2, 28) =21.93	<0.001 ***		(2, 28) =15.72	<0.001 ***	
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	4.16	<0.001 ***	0.20	2.94	0.011*	0.11	1.00	0.33	0.03
TH – N	6.99	<0.001 ***	0.41	6.01	<0.001 ***	0.31	4.86	<0.001 ***	0.21
T – TH	3.19	0.006* *	0.18	3.84	0.002* *	0.20	4.05	<0.001 ***	0.21
Duration	15ms			20ms			25ms		
ANOVA	F	P		F	P		F	P	
	(2, 28) =12.87	<0.001* **		(2, 28) =8.59	<0.001 ***		(2, 28) =5.20	0.012*	
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	0.17	0.867	0.01	0.91	0.375	0.03	1.50	0.156	0.05
TH – N	4.27	<0.001* **	0.15	3.05	0.008*	0.10	1.75	0.101	0.06
T – TH	4.24	<0.001* **	0.18	3.83	0.002* *	0.15	3.15	0.007* *	0.11
Duration	30ms			35ms					
ANOVA	F	P		F	P				
	(2, 28) =3.19	0.056*		(2, 28) =2.40	0.109				
T-tests Pairs	t (14)	P	Mean Diff.						
T – N	1.80	0.092	0.06						
TH – N	0.62	0.542	0.02						
T – TH	2.42	0.029* *	0.08						

Table 13. In the context of M-T4(51), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded.

T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant

The largest mean C-F0 DIFF(T-N) (i.e. difference between F0[unaspirated] and F0[sonorant]) was at 0ms (MeanDiff<sub>zscore</sub>=0.20, MeanDiff<sub>raw</sub>=14.20Hz). The largest mean C-F0 DIFF(TH-N) (i.e. difference between F0[aspirated] and F0[sonorant]) was at 0ms (MeanDiff<sub>zscore</sub>=0.41, MeanDiff<sub>raw</sub>=30.25Hz). The max C-F0 DIFF(T-TH) (i.e. difference between F0[unaspirated] and F0[aspirated]) was at 10ms (MeanDiff<sub>zscore</sub>=0.21, MeanDiff<sub>raw</sub>=15.20Hz).

#### 4.2.3 Summary of results-Q1 for Mandarin

C-F0 \ Tone	M-T1(55)	M-T2(35)	M-T3(214)	M-T4(51)
LENGTH (T- N)	5ms	Non-significant throughout	Non-significant throughout	5ms
Max DIFF <sub>z-score</sub> (T- N)	0.29	Non-significant throughout	Non-significant throughout	0.20
LENGTH (TH- N)	35ms	Non-significant throughout	5ms	20ms
Max DIFF <sub>z-score</sub> (TH- N)	0.34	Non-significant throughout	0.55	0.41
LENGTH (T - TH)	35ms	Non-significant throughout	5ms	30ms
Max DIFF <sub>z-score</sub> (T - TH)	0.19	Non-significant throughout	0.28	0.21
DIRECTION	F0[aspirated] > F0[unaspirated] > F0[sonorant]			

Table 14. Summary of Results-Q1 for Mandarin

Section 4.2.1 has shown that the Mandarin participants produced four distinctive Mandarin tones. Table 14 summarizes the results for Question 1 for the Mandarin part (Section 4.2.): the magnitudes of C-F0 did differ in different Mandarin tonal contexts. In terms of C-F0 LENGTH, the effect was much longer in tones starting with a high pitch than the ones with low pitch. To be more specific, C-F0 maximally extended to 35ms in M-T1(55) and 30ms in M-T4(51), but only 5ms in M-T3(214). Furthermore, no clear C-F0 was found in M-T2(35). In terms of C-F0 DIFF, the maximum F0 difference was slightly bigger for M-T3(214) than for M-T1(55) and M-T4(51). Nevertheless, the general C-F0 DIRECTION did not vary in different Mandarin tone contexts.

F0[aspirated] was generally higher than F0[unaspirated] and F0[sonorant]. F0[unaspirated] started from the middle and converged with F0[sonorant].

### **4.3 C-F0 conditioned by tones in Cantonese?**

Tone trajectories in Cantonese are presented in Section 4.3.1. A series of statistical tests will show that the mid-level C-T3(33) is almost completely merged with the low-level C-T6(22), while other tone pairs are either not merged or just partially merged. Section 4.3.2 presents the results of C-F0 in each Cantonese tone. The findings show that both the magnitudes and direction of C-F0 are conditioned by Cantonese tones. Section 4.3.3 provides a section summary.

#### **4.3.1 Overall Cantonese tone trajectories**

Like Mandarin, in the following reports of the results for Cantonese tones, the raw F0 values of the first 90% will be provided so it is more attainable to make comparison to Cantonese tones measured in raw values in previous studies. The reports and analysis of C-F0 will be conducted with normalized z-scored F0, which makes cross-linguistic comparisons more feasible.

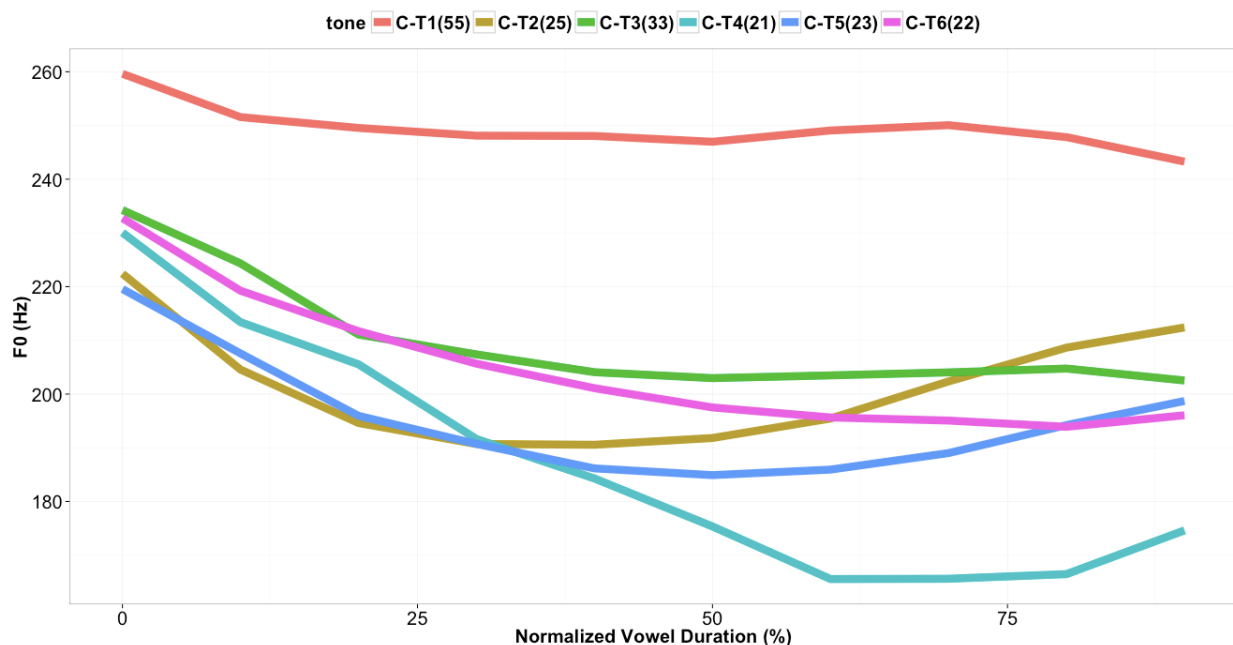


Figure 16. Six Cantonese tone contours within normalized vowels

The F0 values were averaged at every 10% of the vowels of the Cantonese stimuli produced by the fifteen Cantonese participants (Figure 16). Similar to M-T1(55), the F0 values of C-T1(55) were at a high level. The high-rising C-T2(25) did not rise as high as the end of C-T1(55). C-T5(23) had a slightly lower rising contour comparing to C-T2(25). C-T3(33) and C-T6(22) had very similar slightly falling contours. C-T4(21) fell to the lowest pitch among all other tones.

Similar to the method used for the Mandarin stimuli, Table 15 shows the mean F0 values and the standard deviation at every 10% of the total vowel duration of the Cantonese stimuli. ANOVA tests showed that at each 10% time point where the dependent variable was the mean F0, and the within-subjects independent variable was TONE, the mean F0 values were significantly different among the six tones at each 10% time point

Tone	Time Value	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
C-T1(55)	Mean	259.2	250.5	248.5	247.0	246.9	245.8	246.8	249.4	247.1	242.6
	SD	27.4	28.3	27.3	28.3	28.2	27.6	28.1	27.3	26.9	25.5
C-T2(25)	Mean	218.0	204.7	194.7	190.8	190.6	191.8	195.4	202.4	208.6	212.4
	SD	25.0	15.7	16.0	19.0	19.8	20.3	21.0	20.9	18.1	15.6
C-T3(33)	Mean	233.6	223.6	211.1	207.4	204.1	203.0	203.5	204.0	204.7	202.5
	SD	26.2	21.3	27.8	23.7	24.5	24.9	23.7	21.6	18.6	17.8
C-T4(21)	Mean	228.9	212.3	204.9	191.3	184.0	175.1	165.1	164.1	164.2	170.9
	SD	23.5	20.7	20.6	22.0	23.1	24.9	30.0	31.4	29.3	25.6
C-T5(23)	Mean	220.2	207.8	196.4	191.3	186.8	185.6	186.6	189.9	194.9	199.8
	SD	23.9	20.2	17.2	18.5	17.0	16.9	15.2	14.7	12.5	12.6
C-T6(22)	Mean	230.3	219.1	211.7	205.6	201.1	197.5	195.6	195.1	193.9	196.0
	SD	27.5	22.2	20.8	20.3	21.1	21.3	21.6	21.4	21.5	19.4
ANOV A	F(5,70)	26.82	26.78	33.47	34.33	37.31	37.53	36.98	33.29	16.20	33.88
	P	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***
	ges	0.19	0.32	0.39	0.44	0.48	0.50	0.53	0.54	0.45	0.52

Table 15. Mean F0 values(Hz) and the standard deviation at every 10% of the vowel. The lower portion of the table includes results of ANOVAs at each time point where the dependent variable was the mean F0, and the within-subjects independent variable was TONE (6 tones).

Based on the reports of on-going tone mergers in previous studies, four tone pairs were selected among all other possible combinations as the most likely tone merger pairs: C-T2(25) vs. C-T5(23), C-T3(33) vs. C-T5(23), C-T3(33) vs. C-T6(22) and C-T5(23) vs. C-T6(22). Table 16 presents the results of the t-tests for the four tone pairs: significant difference of F0 values were found between 50%-90% of the vowels for C-T2(25) and C-T5(23), between the beginning and 70% for C-T3(33) and C-T5(23), and between the beginning and 50% of C-T5(23) and C-T6(22). Therefore, these three tone pairs were unlikely to undergo any advanced tone merger. However, the F0 values were non-significantly different between C-T3(33) and C-T6(22) except at the time point of 80%, which indicates these two tones have more-or-less merged for our participants.

Tone Pair	Time Value	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
C-T2 vs. C-T5	t(14)	-0.93	-0.62	-0.25	0.89	1.82	2.44	2.57	3.55	3.91	4.33
	p	0.370	0.546	0.803	0.930	0.091	0.029 *	0.052 *	0.012 *	0.004 **	0.003 **
C-T3 vs. C-T5	t(14)	5.38	4.04	3.58	4.10	4.29	4.03	3.58	3.22	2.54	0.85
	p	<0.001 ***	0.004 **	0.012 *	0.004 **	0.004 **	0.005 **	0.012 *	0.024 *	0.024	0.410
C-T3 vs. C-T6	t(14)	1.82	1.16	0.13	0.647	0.876	1.36	1.91	2.34	2.73	1.58
	p	0.946	0.267	0.896	0.530	0.398	0.198	0.080	0.128	0.04 *	0.140
C-T5 vs. C-T6	t(14)	-5.31	-7.67	-6.16	-4.42	-4.26	-2.80	-1.29	-0.41	-0.77	1.40
	p	<0.001 ***	<0.001 ***	<0.001 ***	0.004 **	0.004 **	0.012 *	0.224	0.688	0.456	0.190

Table 16. T-tests for Tone Pair C-T2(25) vs. C-T5(23), C-T3(33) vs. C-T5(23), C-T3(33) vs. C-T6(22) and C-T5(23) vs. C-T6(22) at every 10% of the vowel

## 4.3.2 C-F0 in Cantonese tones

### 4.3.2.1 General results

Figure 17 shows the overall F0 values within the first 50ms after different consonant types in each Cantonese tone. The patterns of C-F0 of C-T1(55) and C-T3(33) were very similar to the general pattern found in Mandarin: F0[aspirated] was higher than F0[unaspirated] and F0[sonorant]. F0[unaspirated] fell from the middle to converge with F0[sonorant]. C-T6(33), which has merged with C-T3(33), shows the same C-F0 patterned with C-T3(33) for F0[unaspirated] and F0[sonorant]. Similarly, F0[sonorant] was close to F0[aspirated] in C-T2(25) and C-T5(23), and even higher than F0[unaspirated] in C-T2(25). F0[sonorant] was higher than F0[aspirated] in C-T4(21). The immediate following sections 4.2.2.2 – 4.2.2.4 will present the results in each tone in more details.



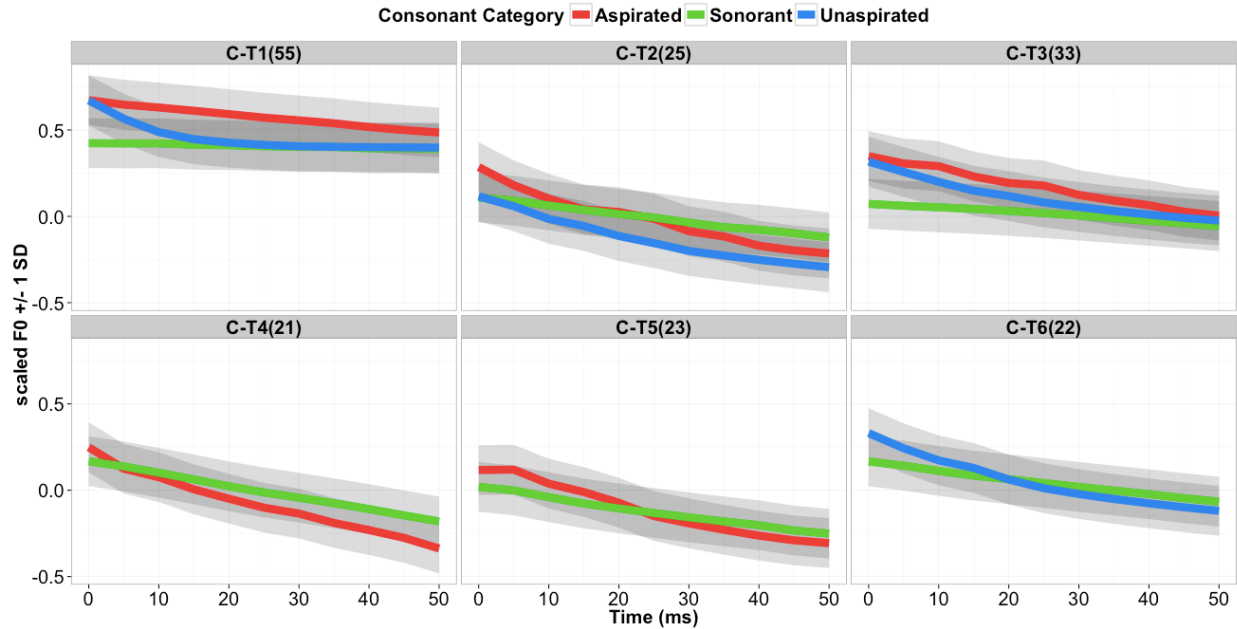


Figure 17. Normalized F0 trajectories within the first 50ms of the vowels in six Cantonese tones. Shading shows 1 SD around the mean

Duration	0ms		5ms		10ms		15ms	
	F	P	F	P	F	P	F	P
C	17.05	<0.001***	12.43	<0.001***	9.42	<0.001***	7.27	<0.001***
T	21.44	<0.001***	32.21	<0.001***	39.98	<0.001***	47.42	<0.001***
V	1.00	0.416	0.92	0.466	0.73	0.601	0.73	0.602
C*T	0.70	0.669	0.91	0.501	1.04	0.403	1.03	0.405
C*V	0.15	0.996	0.21	0.989	0.37	0.937	0.47	0.881
T*V	0.61	0.871	0.55	0.913	0.37	0.986	0.31	0.994
C*T*V	0.47	0.954	0.73	0.760	0.60	0.870	0.60	0.879
Duration	20ms		25ms		30ms		35ms	
	F	P	F	P	F	P	F	P
C	7.35	<0.001***	6.16	0.002**	4.66	0.009**	3.41	0.033*
T	61.02	<0.001***	66.43	<0.001***	62.51	<0.001***	67.06	<0.001***
V	1.01	0.411	1.15	0.329	1.19	0.311	1.28	0.270
C*T	1.13	0.340	1.13	0.336	0.62	0.743	0.52	0.816
C*V	0.48	0.869	0.55	0.815	0.21	0.989	0.23	0.984
T*V	0.31	0.994	0.32	0.992	0.66	0.823	0.74	0.749
C*T*V	0.08	0.867	0.80	0.869	0.752	0.731	0.77	0.713

Table 17. ANOVA tests for effects from the independent variables CONSONANT, TONE and VOWEL on the dependent variable normalized F0 values in Cantonese.

T = Tone; C= Consonant; V = Vowel

Table 17 presents the results of ANOVA tests investigating whether CONSONANT, TONE and VOWEL influenced F0. The results show that CONSONANT had a significant effect on F0 in the

entire 35ms (i.e., the entire statistically analyzed time window), which was longer than that in Mandarin (=20ms, in Table 15). VOWEL did not have any significant influence on F0, which was different from the Mandarin results, where there was a VOWEL effect starting from 15ms. TONE had a clear impact on F0 for the entire 35ms. None of the interactions were significant.

#### 4.3.2.2 C-F0 in C-T1(55)

Figure 18 shows the F0 values within the first 35ms of vowels that carry the high-level C-T1(55). Similar to M-T1(55), C-T1(55) has a higher F0[aspirated] than F0[unaspirated] and F0[sonorant]. F0[unaspirated] also falls from a high pitch and converges with F0[sonorant].

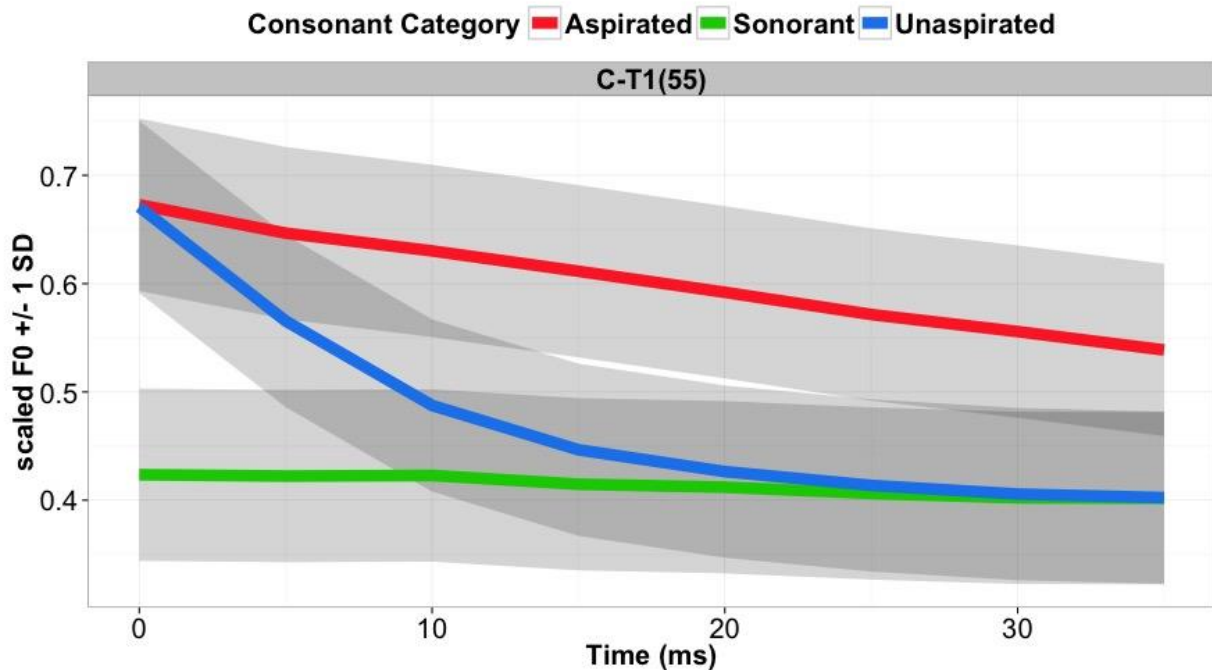


Figure 18. Normalized F0 trajectories within the first 35ms of the vowels in C-T1(55). Shading shows 1 SD around the mean.

Duration	0 ms			5 ms			10 ms		
ANOVA	F	P		F	P		F	P	
	(2, 26) =32.9 5	<0.001**		(2, 26) =29.4 3	<0.001* *		(2, 28) =24.3 9	<0.001* *	
T-tests Pairs	t (13)	P	Mean Diff.	t (13)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	6.75	<0.001**	0.22	4.41	<0.001* *	0.13	2.15	0.05*	0.07
TH – N	7.96	<0.001**	0.27	8.97	<0.001* *	0.24	8.00	<0.001* *	0.22
T – TH	1.17	0.262	-0.0 5	2.85	0.013*	-0.1 0	3.73	0.002 **	-0.1 5
Duration	15ms			20ms			25ms		
ANOVA	F	P		F	P		F	P	
	(2, 28) =28.3 0	<0.001**		(2, 28) =27.9 2	<0.001* *		(2, 28) =24.2 0	<0.001* *	
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	1.52	0.15	0.05	0.94	0.359	0.03	0.71	0.48	0.02
TH – N	7.40	<0.001**	0.21	6.62	<0.001* *	0.19	6.05	<0.001* *	0.18
T – TH	4.80	<0.001**	-0.1 6	5.43	<0.001* *	-0.1 6	5.72	<0.001* *	-0.1 6
Duration	30ms			35ms					
ANOVA	F	P		F	P				
	(2,28) =22.4 2	<0.001** *		(2,28) =19.3 9	<0.001* *				
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.			
T – N	0.55	0.59	0.02	0.41	0.69	0.01			
TH – N	5.58	<0.001**	0.17	5.22	<0.001* *	0.15			
T – TH	5.82	<0.001**	-0.1 5	5.44	<0.001* *	-0.1 3			

Table 18. In the context of C-T1(55), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded.

T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant

Table 18 presents ANOVA and t-tests for the C-F0 magnitudes in C-T1(55). The ANOVA found that overall C-F0 LENGTH for C-T1(55) was the entire 35ms. T-tests showed that the difference between F0[unaspirated] and F0[sonorant] was significant within the first 10ms. F0[aspirated] was significantly higher than F0[sonorant] during the entire 35ms, and than F0[unaspirated] starting from 5ms.

The max mean C-F0 DIFF(T–N) was at 0ms ( $\text{MeanDiff}_{\text{zscore}}=0.22$ ,  $\text{MeanDiff}_{\text{raw}}=10.27\text{Hz}$ ). The max mean C-F0 DIFF(TH–N) was at 0ms ( $\text{MeanDiff}_{\text{zscore}}=0.27$ ,  $\text{MeanDiff}_{\text{raw}}=17.75\text{Hz}$ ). The max C-F0 DIFF(T–TH) was at 15, 20 and 25ms ( $\text{MeanDiff}_{\text{zscore}}=0.16$ ,  $\text{MeanDiff}_{\text{raw}}=7.47\text{Hz}$ ).

#### 4.3.2.3 C-F0 in C-T2(25) and C-T5(23)

The patterns of C-F0 in C-T2(25) and C-T5(23) are presented in this sub-section together because it is interesting to compare C-F0 in two tones that share very similar contour shapes and pitch ranges.

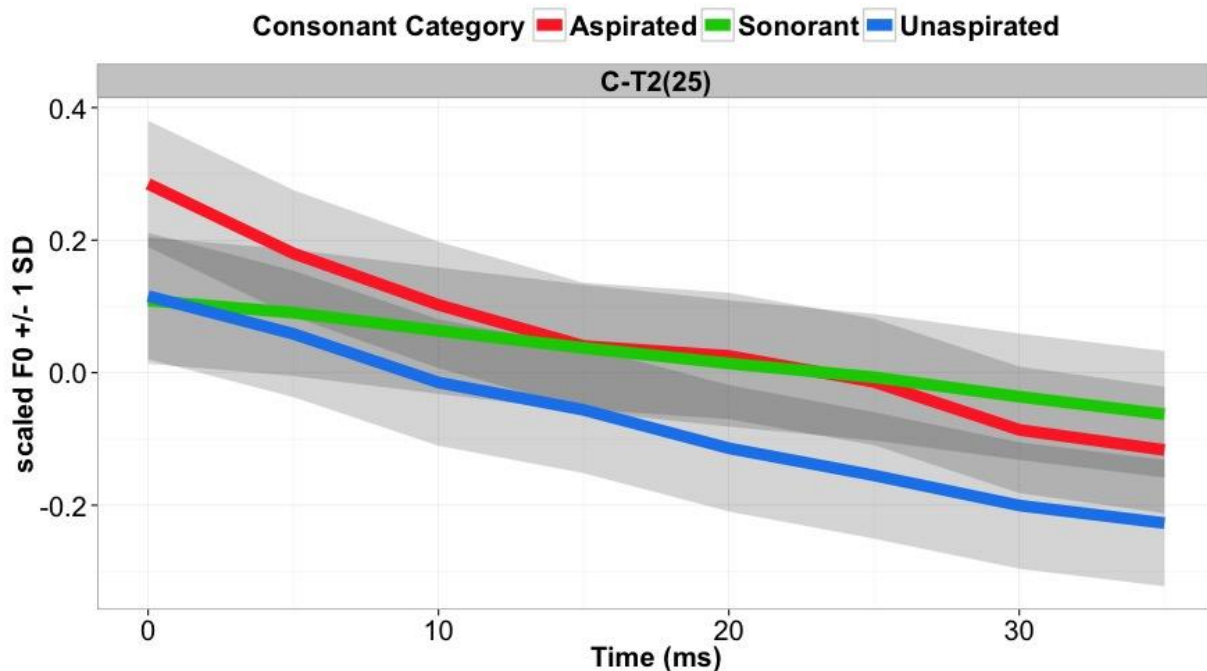


Figure 19. Normalized F0 trajectories within the first 35ms of the vowel in C-T2(25). Shading shows 1 SD around the mean

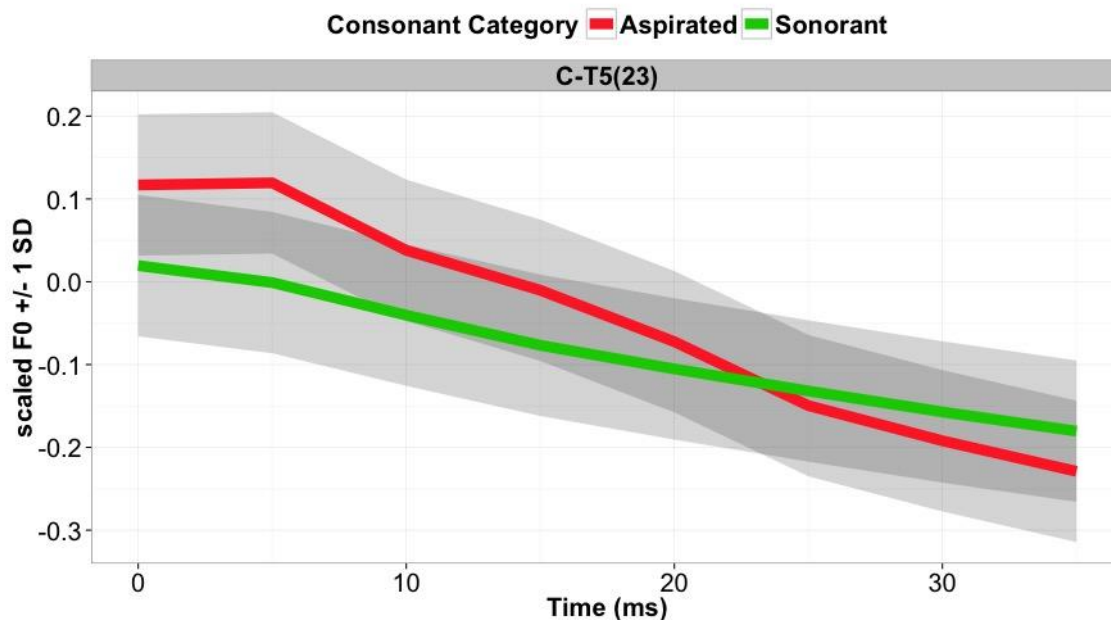


Figure 20. Normalized F0 trajectories within the first 35ms of the vowel in C-T5(23). Shading shows 1 SD around the mean

Figure 19 shows the F0 values within the first 35ms of the vowels that carry C-T2(25). Different from the general direction in Mandarin, though F0[aspirated] is still higher than both F0[sonorant] and F0[unaspirated], F0[unaspirated] is the lowest among the three trajectories in this tone context. Figure 20 shows the F0 trajectories in C-T5(23), with the absence of F0[unaspirated] due to lexical gaps in Cantonese. F0[aspirated] and F0[sonorant] converge later in C-T5(23) than in C-T2(25).

Tables 19 and 20 present ANOVA and t-tests for the C-F0 magnitudes in C-T2(25) and T5(23) respectively. ANOVA found that overall C-F0 LENGTH was between first 15~35ms for C-T2(25), and the first 5ms for C-T5(23). T-tests showed that F0[sonorant] and F0[aspirated] was higher than F0[unaspirated] between 15~35ms for C-T2(25). The difference between F0[aspirated] and F0[sonorant] was not significant for C-T2(25), but significant for the first 5ms of C-T5(23).

The max mean C-F0 DIFF(T-N) was at 30ms and 35ms (MeanDiff<sub>zscore</sub>=0.17, MeanDiff<sub>raw</sub>=13.47Hz) for C-T2(25). The max mean C-F0 DIFF(TH-N) was at 0ms (MeanDiff<sub>zscore</sub>=0.16, MeanDiff<sub>raw</sub>=9.82.Hz) for C-T5(23). The max C-F0 DIFF(T-TH) was at 15, 20 and 25ms (MeanDiff<sub>zscore</sub>=0.11, MeanDiff<sub>raw</sub>=8.78Hz) for C-T2(25).

Duration	0 ms			5 ms			10 ms		
ANOVA	F	P		F	P		F	P	
	(2, 22) =2.28	0.126		(2, 26) =1.44	0.254		(2, 26) =3.01	0.066	
Duration	15ms			20ms			25ms		
ANOVA	F	P		F	P		F	P	
	(2, 26) =3.27	0.054 *		(2, 28) =4.51	0.021 *		(2, 28) =4.74	0.018 *	
T-tests Pairs	t (13)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
T – N	2.83	0.014 *	-0.11	3.23	0.006 *	-0.13	3.52	0.004 *	-0.15
TH – N	0.04	0.971	0.00	0.36	0.722	0.02	0.67	0.516	-0.04
T – TH	2.19	0.047 *	-0.11	2.51	0.026 *	-0.11	2.33	0.036 *	-0.11
Duration	30ms			35ms					
ANOVA	F	P		F	P				
	(2, 28) =8.32	0.001 **		(2, 28) =9.08	0.001 **				
T-tests Pairs	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.			
T – N	3.94	0.001 *	-0.17	4.11	0.001 *	-0.17			
TH – N	1.99	0.068	-0.10	2.18	0.062	-0.10			
T – TH	2.37	0.033 *	-0.07	2.31	0.037 *	-0.06			

Table 19. In the context of C-T2(25), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded.

T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant

Duration	0 ms			5 ms			10 ms		
T-tests Pairs	t (11)	P	Mean Diff.	t (14)	P	Mean Diff.	t (14)	P	Mean Diff.
TH – N	2.56	0.026 *	0.16	2.6	0.021 *	0.13	1.14	0.273	
Duration	15ms			20ms			25ms		
T-tests Pairs	t (14)	P		t (14)	P		t (14)	P	
TH – N	0.65	0.525		0.18	0.856		0.49	0.629	
Duration	30ms			35ms					
T-tests Pairs	t (14)	P		t (14)	P				
TH – N	0.79	0.440		1.07	0.303				

Table 20. In the context of C-T5(23), t-tests at every 5ms time point for normalized F0 after aspirated obstruents and sonorants. Significant results in t-tests are shaded.

TH = Voiceless Aspirated Stop; N = Sonorant

#### 4.3.2.4 C-F0 in C-T3(33), C-T6(22) and C-T4(21)

The patterns of C-F0 in C-T3(33) and C-T6(22) are presented because they are merged tone categories. C-F0 in C-T4(21) is also shown in this sub-section because it has a similar contour shape and pitch height as the other two tones, which provides similar phonetic basis to investigate C-F0.

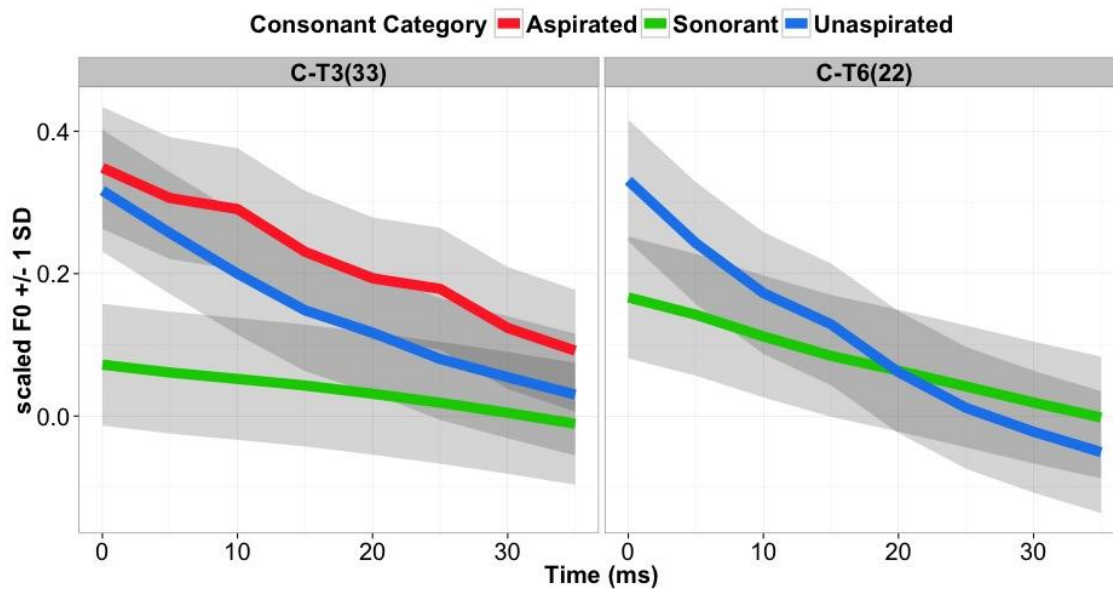


Figure 21. Normalized F0 trajectories within the first 35ms of the vowels in C-T3(33) and C-T6(22). Shading shows 1 SD around the mean.

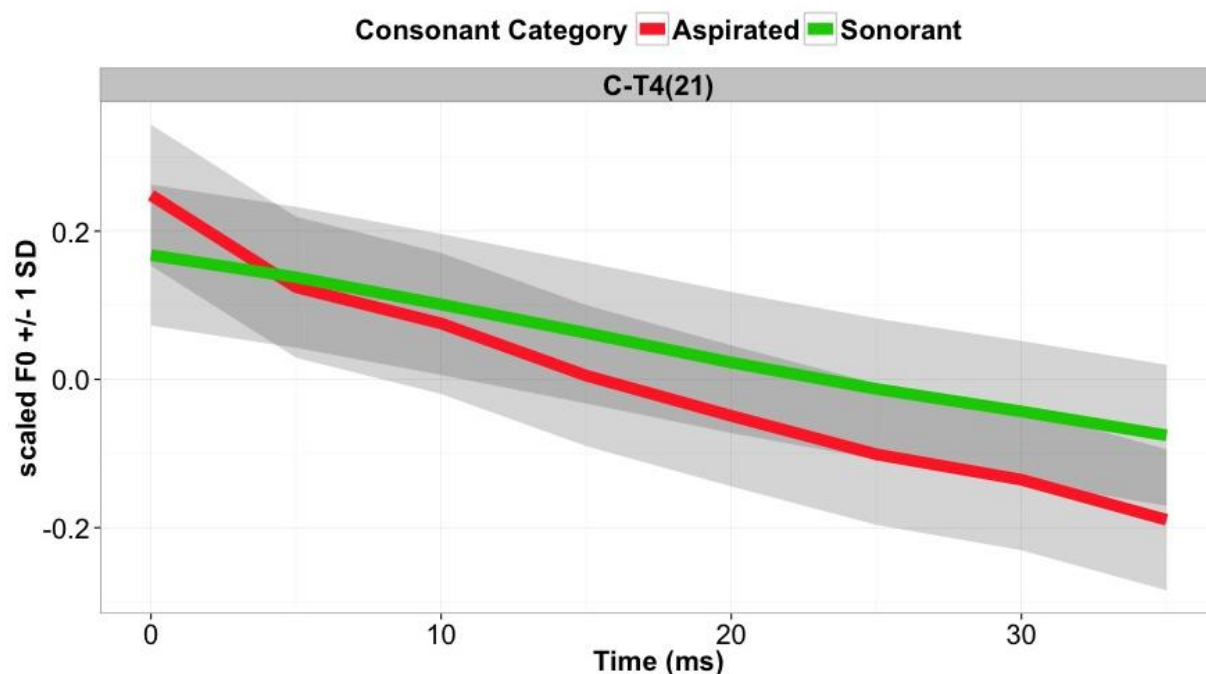


Figure 22. Normalized F0 trajectories within the first 35ms of the vowels in C-T4(21). Shading shows 1 SD around the mean.

Figures 21 and 22 respectively show the F0 values in the three Cantonese tones. Similar to the general direction in Mandarin, F0[aspirated] is higher than both F0[unaspirated], which is in turn higher than F0[sonorant] in C-T3(33). C-T6(22) has the same pattern for F0[unaspirated] and F0[sonorant] before 20ms. Figure 22, on the other hand, shows that F0[sonorant] is higher than F0[aspirated] in C-T4(21).

Table 21-23 respectively report ANOVA and t-tests for the C-F0 magnitudes in the three tones. None of the differences at any timepoint was significant for C-T6(22) or C-T4(21). For C-T3(33), ANOVA found that overall C-F0 LENGTH is the first 5ms. T-tests showed that F0[aspirated] was significantly higher than F0[unaspirated] at 0ms (MeanDiff<sub>zscore</sub>=0.26, MeanDiff<sub>raw</sub>=17.25Hz). No other significant difference was found for C-T3(33).



Duration	0 ms			5 ms			10 ms	
ANOVA	F	P		F	P		F	P
	(2, 22) =5.55	0.017 *		(2, 22) =3.95	0.044 *		(2, 24) =3.04	0.080
T-tests Pairs	t (11)	P	Mean Diff.	t (11)	P	Mean Diff.		
T – N	2.10	0.072	0.13	2.01	0.084	0.14		
TH – N	2.74	0.029 *	0.26	2.15	0.068	0.21		
T – TH	1.76	0.120	-0.13	1.33	0.22	0.8		
Duration	15ms			20ms		25ms		
ANOVA	F	P		F	P		F	P
	(2, 22) =1.48	0.260		(2, 22) =1.15	0.343		(2, 22) =0.896	0.430
Duration	30ms			35ms				
ANOVA	F	P		F	P			
	(2, 22) =0.608	0.558		(2, 22) =0.441	0.652			

Table 21. In the context of C-T3(33), ANOVA at every 5ms time point where the independent variable is CONSONANT and the dependent variable is normalized F0 values. Follow-up t-tests for normalized F0 after different consonants when the ANOVA result is significant. Significant results in t-tests are shaded.

T = Voiceless Unaspirated Stop; TH = Voiceless Aspirated Stop; N = Sonorant

Duration	0 ms		5 ms		10 ms		15 ms	
T-tests Pair	t (12)	P	t (14)	P	t (14)	P	t (14)	P
T – N	1.33	0.208	1.28	0.222	0.31	0.765	0.11	0.911
Duration	20 ms		25 ms		30ms		35ms	
T-tests Pair	t (14)	P	t (14)	P	t (14)	P	t (14)	P
T – N	1.11	0.287	1.42	0.178	1.68	0.116	1.95	0.072

Table 22. In the context of C-T6(22), t-tests at every 5ms time point for normalized F0 after unaspirated obstruents and sonorants. Significant results in t-tests are shaded.

T = Voiceless Unaspirated Stop; N = Sonorant

Duration	0 ms		5 ms		10 ms		15 ms	
T-tests Pair	t (14)	P	t (14)	P	t (14)	P	t (14)	P
TH – N	1.32	0.208	0.03	0.971	0.35	0.730	0.771	0.454
Duration	20 ms		25 ms		30ms		35ms	
T-tests Pair	t (14)	P	t (14)	P	t (14)	P	t (14)	P
TH – N	0.911	0.379	1.18	0.258	1.53	0.151	1.83	0.090

Table 23. In the context of C-T4(21), t-tests at every 5ms time point for normalized F0 after aspirated obstruents and sonorants. Significant results in t-tests are shaded.

TH = Voiceless Aspirated Stop; N = Sonorant

#### 4.3.3 Summary of results-Q1 for Cantonese

Tone C-F0	C-T1(55)	C-T3(33)	C-T2(25)	C-T4(21)	C-T5(23)
LENGTH (T- N)	10ms	Non-significant throughout	from 15 to 35 ms	N/A	N/A
Max DIFF <sub>z-score</sub> (T- N)	0.22	Non-significant throughout	0.11	N/A	N/A
LENGTH (TH- N)	35ms	<5ms	Non-significant throughout	Non-significant throughout	5ms
Max DIFF <sub>z-score</sub> (TH- N)	0.27	0.26	Non-significant throughout	Non-significant throughout	0.16
LENGTH (T - TH)	30ms	Non-significant throughout	from 15 to 35 ms	N/A	N/A
Max DIFF <sub>z-score</sub> (T - TH)	0.16	Non-significant throughout	0.17	N/A	N/A
DIRECTION	F0[aspirated] > F0[unaspirated] > F0[sonorant]		F0[aspirated] = F0[sonorant] > F0[unaspirated]	N/A	F0[aspirated] > F0[sonorant]

Table 24. Summary of Results-Q1 for Cantonese

Section 4.3.1 has shown that Cantonese participants produced five distinctive Cantonese tones, with C-T3(33) and C-T6(22) merged. Table 24 summarizes the results for Question 1 for the Cantonese part (Section 4.3.2): the magnitudes of C-F0 differ in different Cantonese tones. In terms of C-F0 LENGTH, as with the Mandarin results, the effect was still longer in tones associated with high pitch than the ones with low pitch. More specifically, C-F0 maximally

extended to 35ms in C-T1(55) and 35ms in C-T2(25), but only 5ms in C-T3(33) and C-T5(23). As for C-F0 DIFF, the maximum F0 difference was slightly smaller for C-T2(25) and C-T5(23) than for C-T1(55) and C-T3(33). The general C-F0 DIRECTION also varied in different Cantonese tone contexts. F0[aspirated] was generally higher than F0[sonorant], but F0[unaspirated] was in the middle in C-T1(55) and C-T3(33), but the lowest in C-T2(25).

#### **4.4 Overall summary for results-Q1**

Question 1 asks whether tones can condition C-F0. The results in this chapter show that tones can influence C-F0 LENGTH and C-F0 DIFF in both Mandarin and Cantonese. Tones can also condition C-F0 DIRECTION in Cantonese, but not in Mandarin.

As for confirming distinctive tones are produced: the Mandarin data show four contrastive tones, while Cantonese has five contrastive tones, with two tones, i.e. C-T3(33) and C-T6(22), that have merged.

The magnitudes of C-F0 are conditioned by tones in both Mandarin and Cantonese. As for C-F0 LENGTH in Mandarin, the effect was longer in M-T1(55) and M-T4(51) than in M-T3(214). In Cantonese, the effect was longer in C-T1(55) and in C-T2(25) than in C-T3(33) and C-T5(23). In general, C-F0 extended farther in tones associated with high pitch. As for the max C-F0 DIFF in Mandarin, it was slightly bigger for M-T3(214) than for M-T1(55) and M-T4(51). In Cantonese, the maximum difference was slightly bigger for C-T1(55) and C-T3(33) than for C-T2(25) and C-T5(23).

There was a consistent C-F0 DIRECTION across all Mandarin tones and C-T1(55) and C-T3(33): F0[aspirated] was higher than F0[unaspirated] and F0[sonorant]. F0[unaspirated] started from the middle and converged with the lowest F0[sonorant]. The results indicate a robust aspiration raising and a weaker voiceless raising effect in both Mandarin and Cantonese.

For C-T2(25), however, F0[unaspirated] was lower than F0[aspirated] and F0[sonorant]. It is unclear why the unaspirated stops have a lowering effect on F0 in the C-T2(25) context but not in other tone contexts in our data. One possible reason can be that C-T2(25) initiated with unaspirated stops have undergone more advanced tone merger with a lower contour as in C-T5(23), so F0[unaspirated] in C-T2(25) became lower than the other two consonant types. This speculation needs further research to confirm.

## **5. Results for Question 2: C-F0 conditioned by tone distance?**

### **5.1 Introduction**

This chapter addresses Research Question 2 stated in Chapter 1: is C-F0 conditioned by tone distance in a tone inventory? To be more specific, the previous chapter finds that the magnitudes of C-F0 differ in different tonal contexts, this chapter further asks whether the acoustic distance between tones is responsible for the different magnitudes of C-F0.

Three aspects are of interest in this chapter: (1) The first aspect is tone distance, represented by TONE DIFF and defined as the F0 distance between two tones in a certain period of time (Alexander 2010: 68). One type of tone distance studied here is the shortest tone distance of a target tone, i.e. the F0 difference between the target tone and its closest tone in the tone inventory. I call this type of tone distance Smallest TONE DIFF. The other type of tone distance calculates the average value of a target tone's distances from all other tones in the tone inventory. I call this Overall TONE DIFF.

(2) The second aspect is the magnitude of C-F0, which has two major indicators: C-F0 LENGTH (i.e. how far can C-F0 extend within the vowel) and C-F0 DIFF (i.e. the difference between F0 values after different consonant types). Only two types of C-F0 magnitudes are investigated: the one between F0[unaspirated] and F0[sonorant] and the one between F0[aspirated] and F0[sonorant]. That only these two types represents how the F0 values following the obstruents variate away from the baseline (i.e. F0[sonorant]), so they can best reveal the magnitudes of general C-F0.

(3) The third aspect is the relation between C-F0 DIFF and TONE DIFF; that is, whether the degree of tone distance can influence the magnitudes of C-F0 will be investigated. One possibility is that C-F0 magnitudes in a certain tone context may vary depending on how

crowded a tone space the tone is in. Alternatively, C-F0 magnitudes may not be related to the crowdedness of the tone space.

The general hypothesis in this chapter is that the crowdedness of the tone space (or equivalently, distance from other tones) can influence the magnitudes of C-F0 related to a particular tone. This hypothesis is based on a functional account that a more crowded tone space needs a higher degree of enhancement of tone contrast, while the C-F0 that has consistent direction across tone contexts weakens the enhancement of tone contrasts by increasing the variance of the acoustic space of the tone. Therefore, this hypothesis predicts that a shorter tone distance, which indicates stronger needs for enhancement, should be associated with weaker magnitudes of C-F0, which indicates smaller variance for the tone space. It is unclear, however, what kind of tone distance is more likely to be related to C-F0, so there are two hypotheses that will be tested in this chapter:

Hypothesis 5A: It is the acoustic distance between a tone and its acoustically-closest tone (i.e. Smallest TONE DIFF) that matters to the tone contrast and thus is related to the magnitudes of C-F0.

Hypothesis 5B: It is the mean distance between a tone and all other tones in the tone inventory (i.e. Overall TONE DIFF) that matters to the tone contrast and thus is related to the magnitudes of C-F0.

The results of Mandarin reported in Section 5.2 and Cantonese in Section 5.3 will show evidence to support Hypothesis 5A based on Smallest TONE DIFF, which correctly predicts the patterns of C-F0 LENGTH in different tone contexts. However, neither hypothesis is consistent with the C-F0 DIFF within the first 10ms.

This chapter is also interested in how competitions between tones can give rise to different magnitudes of C-F0 across languages. Mandarin has two tones that have high onset pitch: M-T1(55) and M-T4(51), while Cantonese only has one tone that starts from high pitch: C-T1(55). This indicates that the competition between tone contrast is more intense for M-T1(55) than for C-T1(55), since M-T1(55) has a competitor in the high pitch space whereas C-T1(55) has no neighbor. If higher degree of tone competition may restrict C-F0, this predicts that C-T1(55) should have a stronger C-F0 than the same high-level M-T1(55). Section 5.4 shows evidence from C-F0 LENGTH and C-F0 DIFF to support this point.

Section 5.5 provides concluding remarks and a discussion.

## **5.2 Intra-linguistic comparison in Mandarin**

This section investigates the relation between C-F0 and tone distance in Mandarin. Section 5.2.1 uses the results of Smallest TONE DIFF and Overall TONE DIFF to make different predictions (5.2A and 5.2B) about the magnitudes of C-F0. Section 5.2.2 shows that Prediction 5.2A based on Smallest TONE DIFF can correctly predict C-F0 LENGTH in different tone contexts, but not the C-F0 DIFF within the first 10ms. Section 5.2.3 shows that there was no statistically significant correlation between Smallest TONE DIFF and C-F0 DIFF. Section 5.2.4 provides a summary.

### **5.2.1 Tone distance in Mandarin**

#### **5.2.1.1 Smallest TONE DIFF in Mandarin**

Figure 23 shows the F0 difference of every possible tone pair in Mandarin. The y-axis is the normalized F0 difference within the first 10ms of the vowels in six Mandarin tone pairs within subjects. The thick horizontal lines in the boxes represent the median values of F0 difference between two tones. The upper whiskers above the box represent the range of the maximum

values and the 75<sup>th</sup> percentile, while the lower whiskers below represent the range of the 25<sup>th</sup> percentile and the minimum values. The figure clearly shows that the tone pair M-T2(35) vs. M-T3(214) has the lowest median value of F0 difference among all tone pairs, while M-T1(55) vs. M-T4(51) has the second lowest.

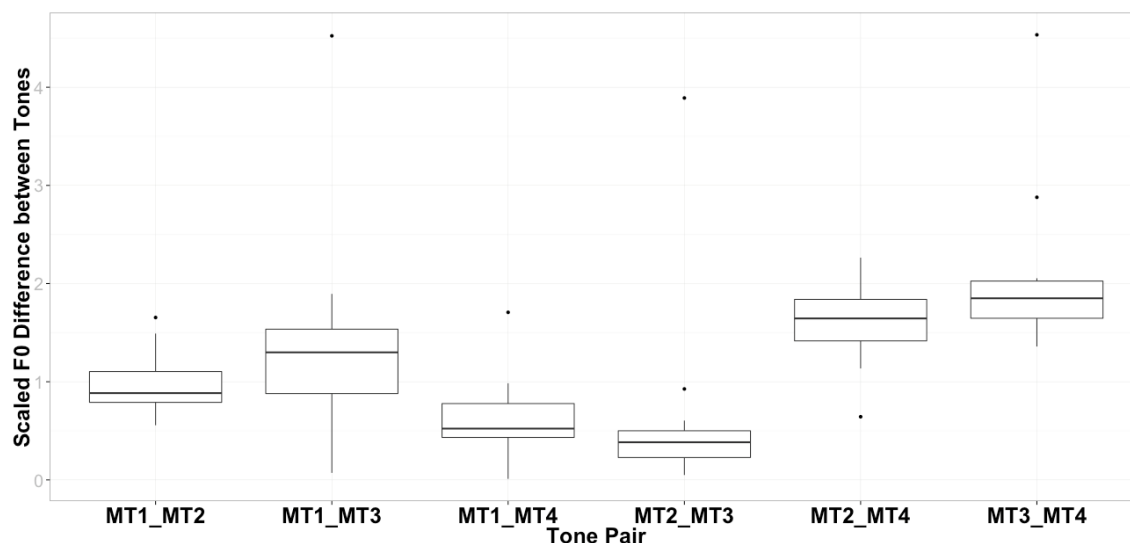


Figure 23. Scaled F0 difference in six Mandarin tone pairs (within the first 10ms). Each tick on the x-axis separates the names of the tones. E.g. ‘MT1\_MT2’ represents the tone pair M-T1(55) vs. M-T2(35).

An ANOVA, where the dependent variable was the F0 differences, and the within-subjects independent variable was TONE PAIR, revealed that there were significant differences in tone distance for the six tone pairs ( $F[5, 70]=34.4$ ,  $p<0.001$ ). The follow-up t-tests found that the F0 difference between M-T1(55) vs. M-T4(51) is significantly higher than that between M-T2(35) vs. M-T3(214) ( $t[14]=2.43$ ,  $p=0.029$ ), but significantly lower than that between M-T1(55) vs. M-T2(35) ( $t[14]=2.32$ ,  $p=0.035$ ) and M-T1(55) vs. M-T3(214) ( $t[14]=2.15$ ,  $p=0.041$ ). Since the F0 difference between M-T2(35) vs. M-T4(51) and that between M-T3(214) vs. M-T4(51) are obviously higher than other tone pairs, no t-tests were performed on these two tone pairs.



The results suggest that the Smallest TONE DIFF for both M-T1(55) and M-T4(51) is the distance between these two tones. Similarly, the Smallest TONE DIFF for both M-T2(35) and M-T3(214) is the distance between them. Considering the observed patterns for the tonal differences, the prediction based on Hypothesis 5A is as follows.

PREDICTION 5.2A: The Smallest TONE DIFF for each tone (from big to small) is  $M-T1 = M-T4 > M-T2 = M-T3$ . That is, the Smallest TONE DIFF for M-T1(55) is equal to M-T4(51) and larger than that for M-T2(35), which is equal to the Smallest TONE DIFF for M-T3(214). If the acoustic distance between a tone and its acoustically-closest tone (i.e. Smallest TONE DIFF) is positively related to the magnitudes of C-F0, this predicts that the magnitudes of C-F0 (from strong to weak) is as follows:  $C-F0(M-T1) = C-F0(M-T4) > C-F0(M-T2) = C-F0(M-T3)$ .

#### **5.2.1.2 Overall TONE DIFF in Mandarin**

Figure 24 shows the average tone distance of a target tone from other tones in the inventory, i.e. Overall TONE DIFF. In other words, the Overall TONE DIFF of each tone is the average of only 3 difference values in Mandarin that involves that relevant tone for each participant. In the figure, M-T4(51) has a much higher Overall TONE DIFF than the other three tones, while M-T1(55) has the lowest Overall TONE DIFF among all four tones.

An ANOVA, where the dependent variable was the Overall TONE DIFF, and the within-subjects independent variable was TONE, revealed that there was a significant difference for the Overall TONE DIFF of the four Mandarin tones ( $F[3, 42]=34.4, p<0.001$ ). Follow-up t-tests in Table 25 showed that the Overall TONE DIFF of M-T4(51) is significantly larger than the other three tones. No other comparisons were significant. The prediction based on Hypothesis 5B is as follows.

PREDICTION 5.2 B: The Overall TONE DIFF for each tone (from big to small) is M-T4 > M-T1 = M-T2 = M-T3. If the magnitudes of C-F0 is related to the mean distance between a tone and all other tones in the tone inventory, this predicts that the magnitudes of C-F0 (from strong to weak) is as follows: C-F0(M-T4) > C-F0(M-T1) = C-F0(M-T2) = C-F0(M-T3).

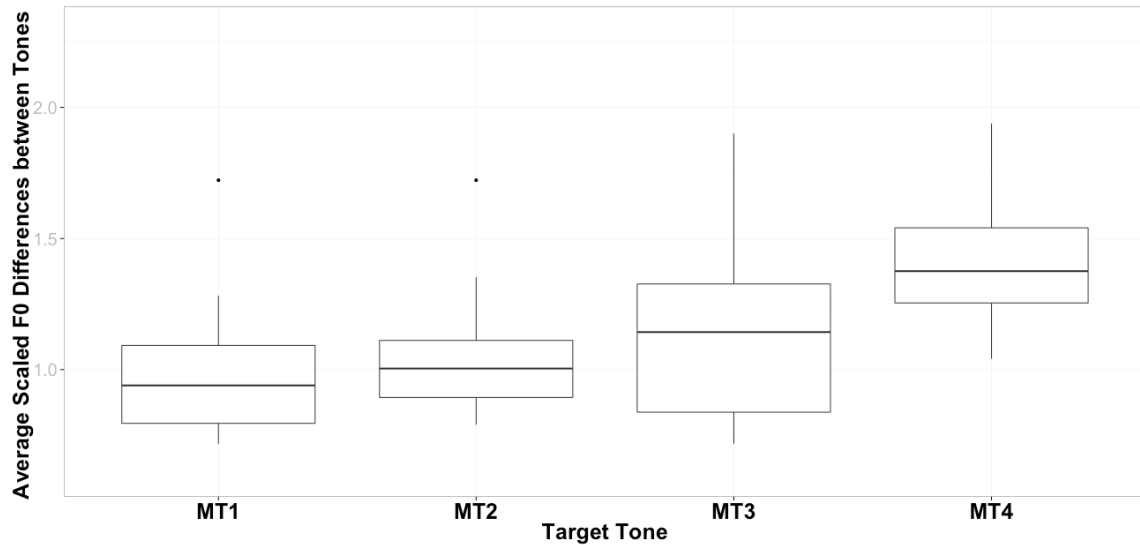


Figure 24. Overall TONE DIFF: the average values of all scaled F0 differences between a target tone and all other tones in Mandarin (within the first 10ms). E.g. Overall TONE DIFF (M-T1) = (TONE DIFF(M-T1 vs. M-T2) + TONE DIFF(M-T1 vs. M-T3) + TONE DIFF(M-T1 vs. M-T4)) ÷ 3

M-T1 vs. M-T2	t[14]=2.52 p = 0.062	M-T2 vs. M-T3	t[14]=0.28 p = 0.782
M-T1 vs. M-T3	t[14]=1.78 p = 0.108	M-T2 vs. M-T4	t[14]=6.37 p <0.001 ***
M-T1 vs. M-T4	t[14]=2.52 p <0.001 ***	M-T3 vs. M-T4	t[14]=6.37 p = 0.024 *

Table 25. T-tests comparing the Overall TONE DIFF of each Mandarin tone (within the first 10ms)

### 5.2.2 The magnitudes of C-F0 in Mandarin tones

Prediction 5.2A states that the magnitudes of C-F0 (from strong to weak) is C-F0(M-T1) = C-F0(M-T4) > C-F0(M-T2) = C-F0(M-T3), while Prediction 5.2B states that C-F0(M-T4) > C-F0(M-T1) = C-F0(M-T2) = C-F0(M-T3). To evaluate magnitudes of C-F0, we examine two dimensions: C-F0 LENGTH and C-F0 DIFF.

The results of C-F0 LENGTH (i.e., the duration that significant C-F0 extends to in the vowel) in Chapter 4 have supported Prediction 5.2A: M-T1(55) and M-T4(51) have similar large values of maximum C-F0 LENGTH (TH-N) that extends to 20 - 35ms of the vowel, whereas M-T3(214) has weak C-F0 LENGTH and C-F0 is non-significant throughout M-T2(35).

However, if we only consider the differences between F0 values after consonants, i.e. C-F0 DIFF, within the first 10ms, Figure 25 shows that C-F0 DIFF (T-N) (i.e. the difference between F0 after unaspirated obstruents and sonorants) are similar in each tone context, whereas C-F0 DIFF (TH-N) (i.e. the difference between F0 after aspirated obstruents and sonorants) are slightly lower with M-T2(35) than other tones.

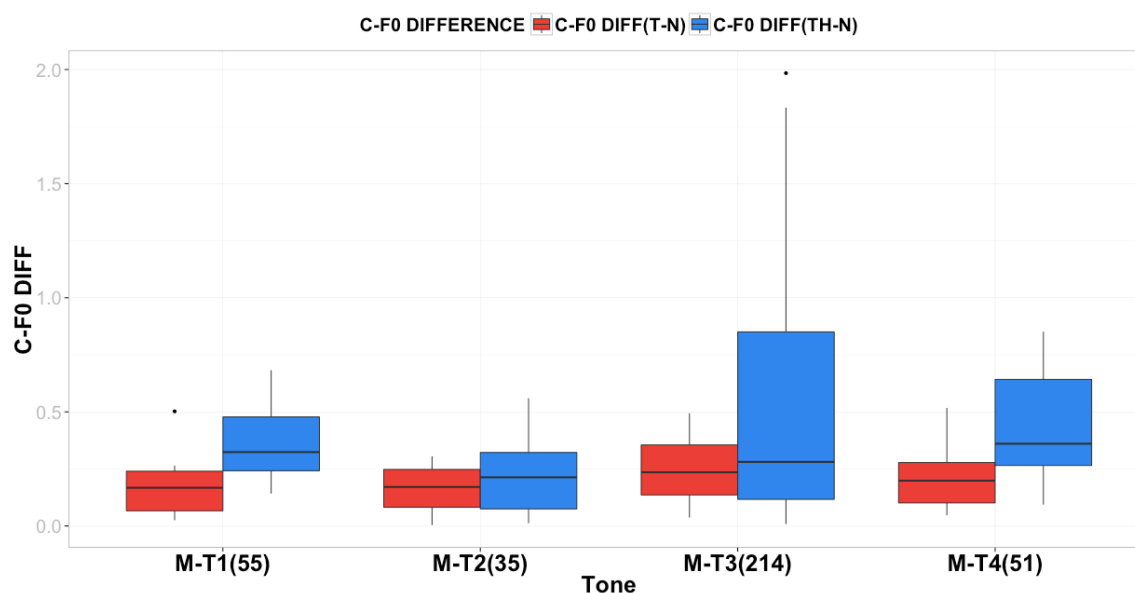


Figure 25. Average C-F0 DIFF in each Mandarin tone (within the first 10ms). The red boxes represent the difference between F0 values after unaspirated obstruents and after sonorants. The blue boxes represent the difference between F0 values after aspirated obstruents and after sonorants.

ANOVAs, where the dependent variable was the C-F0(T-N) or C-F0(TH-N), and the within-subjects independent variable was TONE, found that the difference was non-significant for C-F0 DIFF (T-N) ( $F[3, 42] = 1.12, p=0.348$ ) but significant for C-F0 DIFF (TH-N) ( $F[3, 42] = 2.85, p=0.048$ ) (Table 26). Follow-up t-tests (Table 27) found that there was no crucial difference

of C-F0 DIFF for M-T1(55), M-T3(214) and M-T4(51), but C-F0 DIFF of M-T2(35) was significantly lower than the other three tones. This result was neither predicted by Prediction 5.2A or by 5.2B. That is, the result indicates that neither Smallest TONE DIFF nor Overall TONE DIFF influences C-F0 DIFF within the first 10ms in the predicted way.

C-F0 LENGTH \ Tone	M-T1(55)	M-T2(35)	M-T3(214)	M-T4(51)
LENGTH (T- N)	5ms	Non-significant throughout	Non-significant throughout	5ms
LENGTH (TH- N)	35ms	Non-significant throughout	5ms	20ms

Table 26. Summary of C-F0 LENGTH results in Question1 for Mandarin. The values of maximum C-F0 LENGTH is the timepoint after which C-F0 becomes non-significant.

M-T1 vs. M-T2	t[14]=2.33 p = 0.035 *	M-T2 vs. M-T3	t[14]=2.42 p = 0.029 *
M-T1 vs. M-T3	t[14]=1.19 p = 0.253	M-T2 vs. M-T4	t[14]=2.98 p = 0.009 **
M-T1 vs. M-T4	t[14]=0.827 p = 0.422	M-T3 vs. M-T4	t[14]=0.923 p = 0.371

Table 27. T-tests comparing C-F0 DIFF(TH-N) in each Mandarin tone (within the first 10ms)

In sum, considering the two aspects of C-F0 magnitudes, Prediction 5.2A using Smallest TONE DIFF is consistent with the pattern of C-F0 LENGTH, but neither prediction is consistent with the pattern of C-F0 DIFF within the first 10ms.

### 5.2.3 C-F0 and Mandarin tone distance

The previous Section 5.2.2 showed that Prediction 5.2A partly correctly captures the fact that the magnitudes of C-F0 are similar for M-T1(55) and M-T4(51), based on the findings that these two tones have the same Smallest TONE DIFF. The same also applies to M-T2(35) and M-T3(214). Prediction 5.2A also rightly captures the fact that C-F0 is stronger in M-T1(55) and M-T4(51) than in M-T2(35) and M-T3(214), since Smallest TONE DIFF for the former two tones is bigger than that for the latter two tones.

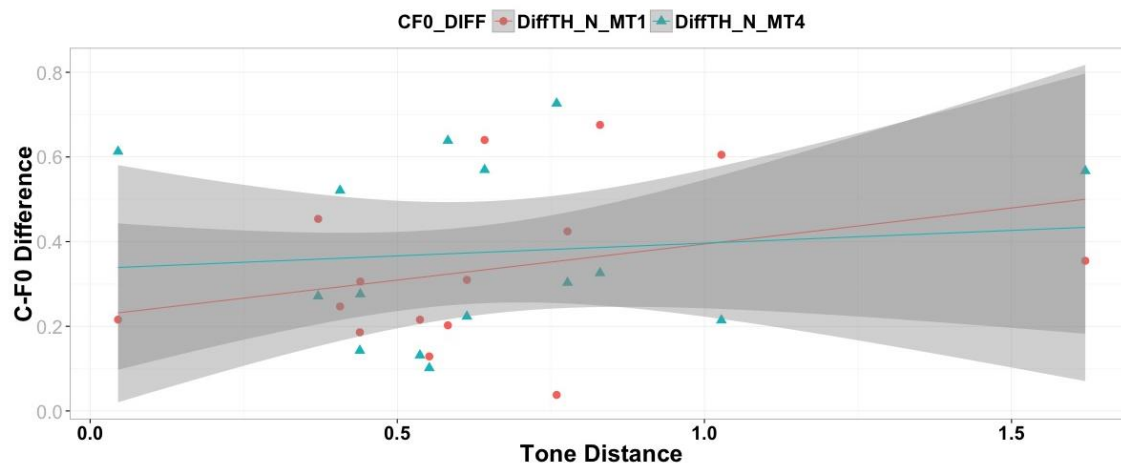


Figure 26. Relation between C-F0 DIFF(TH-N) for M-T1(55) and M-T4(51) and Smallest TONE DIFF between M-T1(55) and M-T4(51)

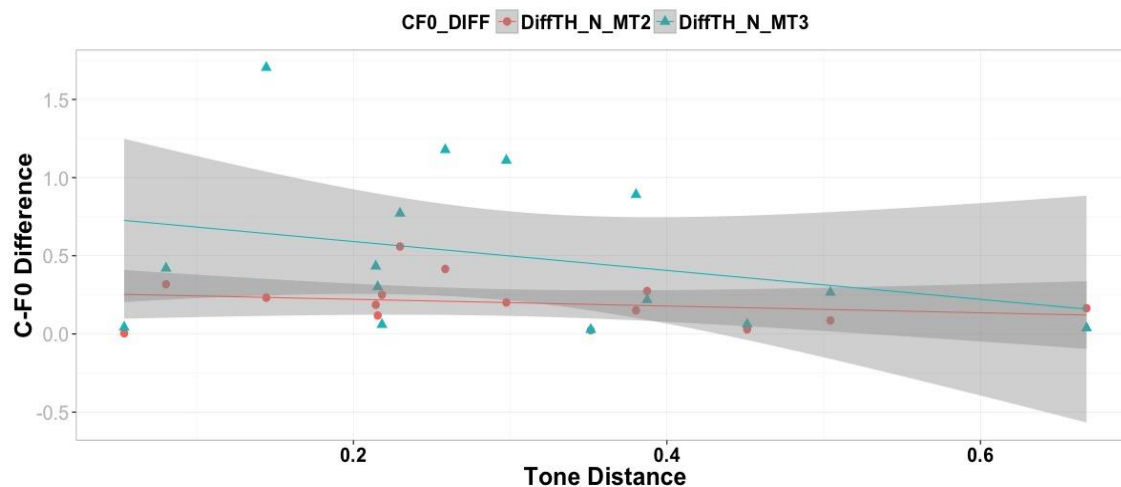


Figure 27. Relation between C-F0 DIFF(TH-N) for M-T2(35) and M-T3(214) and Smallest TONE DIFF between M-T2(35) and M-T3(214)

	$R^2$	F [1, 13]	p
M-T1(55)	0.071	0.064	0.803
M-T2(35)	0.059	0.208	0.655
M-T3(214)	0.19	4.283	0.059
M-T4(51)	0.064	0.155	0.699

Table 28. Linear Regression tests for the relation between C-F0 DIFF(TH-N) and Smallest TONE DIFF in Mandarin (within the first 10ms)

The relation between C-F0 DIFF(TH-N) and Smallest TONE DIFF is investigated in this sub-section. The Smallest TONE DIFF is used as the independent variable to predict the

dependent variable C-F0 DIFF(TH-N). Figures 26 – 27 show how C-F0 DIFF(TH-N) changes with Smallest TONE DIFF. If tone distance can condition C-F0, a positive correlation between them should be expected. A slight positive correlation is observable in Figure 26 for M-T1(55) and M-T4(51), but a slight negative correlation is observable in Figure 27 for M-T2(35) and M-T3(214). A series of linear regression tests show that none of the correlations was significant (Table 28). In sum, no evidence was found to show that Smallest TONE DIFF influenced C-F0 DIFF in Mandarin.

#### **5.2.4 Summary of intra-linguistic comparison in Mandarin**

This sub-section has investigated Question 2 on the relation between tone distance and C-F0 in Mandarin. The Smallest TONE DIFF results find the following sequence (from strong to weak): M-T1 = M-T4 > M-T2 = M-T3. On the other hand, the Overall TONE DIFF has the following sequence (from strong to weak): M-T4 > M-T1 = M-T2 = M-T3.

Prediction 5.2A based on Smallest TONE DIFF correctly predict that C-F0 LENGTH(TH-N) is similar for M-T1(55) and M-T4(51) and for M-T2(35) and M-T3(214). Prediction 5.2A also rightly predicts that C-F0 LENGTH(TH-N) is longer for M-T1(55) and M-T4(51) than for M-T2(35) and M-T3(214). However, for the other indicator of magnitudes, C-F0 DIFF within the first 10ms, Prediction 5.2A fails in predicting the findings that C-F0 DIFF of M-T2(35) is significantly lower than the other three tones. Also, no significant correlation is found between Smallest TONE DIFF and C-F0 DIFF in any of the tones.

To summarize, the results from Mandarin indicate that Smallest TONE DIFF may condition C-F0 LENGTH but not C-F0 DIFF. No evidence is found that Overall TONE DIFF is related to the magnitudes of C-F0.

### **5.3 Intra-linguistic comparison in Cantonese**

This section studies the relation between magnitudes of C-F0 and tone distance in Cantonese. Section 5.3.1 reports the results of Smallest TONE DIFF and Overall TONE DIFF and lays out different predictions (5.3A and 5.3B) about the magnitudes of C-F0. Similar to the results of Mandarin, Section 5.3.2 supports Prediction 5.3A based on Smallest TONE DIFF, which can rightly predict the pattern of C-F0 LENGTH but not the pattern of C-F0 DIFF within the first 10ms. Section 5.3.3 shows that there was no statistically significant correlation between Smallest TONE DIFF and C-F0 DIFF. Section 5.3.4 provides a summary.

#### **5.3.1 Tone distance in Cantonese**

##### **5.3.1.1 Smallest TONE DIFF in Cantonese**

Figure 28 shows the F0 difference of six tone pairs in Cantonese. In Chapter 4, it is shown that the onset pitch of C-T3(33) vs. C-T6(22) and C-T2(25) vs. C-T5(23) have merged, so it is reasonable to remove one of C-T3(33) and C-T6(22) and one of C-T2(25) and C-T5(23). Since there were more consonant categories and more data for C-T3(33) and C-T2(25), C-T6(22) and C-T5(23) were removed in this investigation. For C-T1(55), the shortest distance appears to be the F0 difference between C-T1(55) and C-T3(33) in the figure. For C-T2(25) and C-T4(21), the shortest appears to be the F0 difference between C-T2(25) and C-T4(21). For C-T3(33), it is the F0 difference between C-T3(33) and C-T4(21).

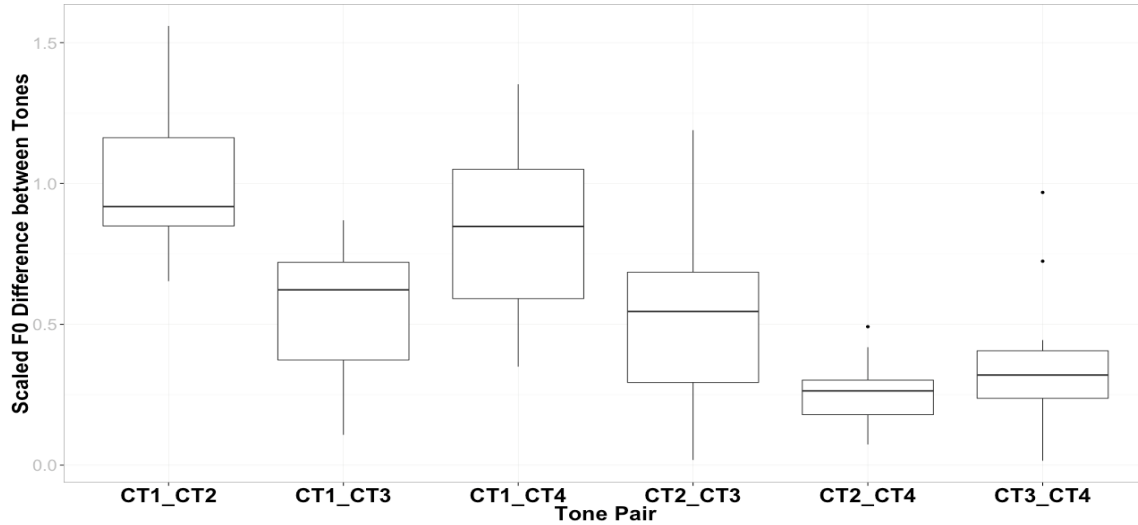


Figure 28. Scaled F0 difference in ten Cantonese tone pairs (within the first 10ms). Each tick on the x-axis separates the names of the tones. E.g. ‘CT1\_CT2’ represents the tone pair C-T1(55) vs. C-T2(25).

An ANOVA, where the dependent variable was the scaled F0 difference between tones, and the within-subjects independent variable was TONE PAIR, revealed that there are significant differences for the F0 distance between tones for the six tone pairs ( $F[5, 70]=23.9$ ,  $p<0.001$ ). The following t-tests find that the F0 difference between C-T1(55) and C-T3(33) (i.e. Smallest TONE DIFF for C-T1(55)) is significantly higher than that between C-T2(25) and C-T4(21) (i.e. Smallest TONE DIFF for C-T2(25) and C-T4(21)) ( $t[14]=3.83$ ,  $p=0.001$ ) and marginally higher than that between C-T3(33) and C-T4(21) (i.e. Smallest TONE DIFF for C-T3(33)) ( $t[14]=2.04$ ,  $p=0.054^*$ ). However, there is no significant difference between TONE DIFF for C-T2(25) and C-T4(21) and C-T3(33) and C-T4(21) ( $t[14]=1.21$ ,  $p=0.245$ ).

The results suggest that the Smallest TONE DIFF for C-T1(55) is significantly higher than that for C-T2(25), C-T3(33) and C-T4(21). Also, there is no statistically significant difference between the Smallest TONE DIFF for the latter three tones. Considering the observed patterns for the tonal differences, the prediction based on Hypothesis 5A is as follows:



PREDICTION 5.3A: The Smallest TONE DIFF for each tone (from big to small) is  $C-T1 > C-T2 = C-T3 = C-T4$ . If Smallest TONE DIFF is related to the magnitudes of C-F0, this predicts that the magnitudes of C-F0 (from strong to weak) is as follows:  $C-F0(C-T1) > C-F0(C-T2) = C-F0(C-T3) = C-F0(C-T4)$ .

### 5.3.1.2 Overall TONE DIFF in Cantonese

Figure 29 shows the mean F0 difference of each tone from other tones within the first 10ms, i.e. Overall TONE DIFF. In the figure, the median values of Overall TONE DIFF from large and small are as follows  $C-T1(55) > C-T2(25) > C-T4(21) > C-T3(33)$ .

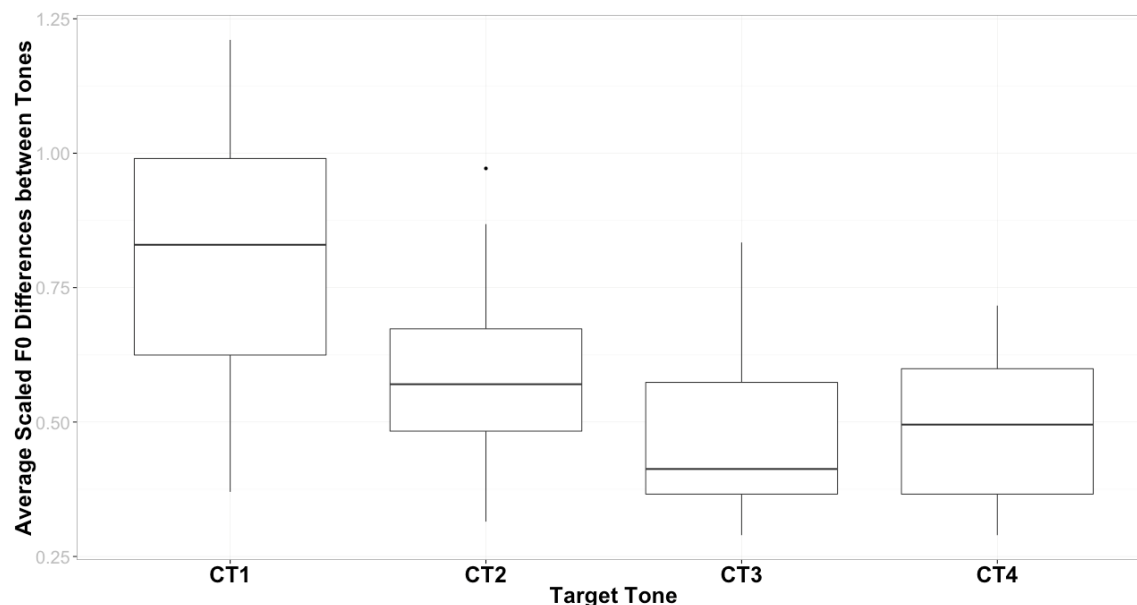


Figure 29. Overall TONE DIFF: the average values of all scaled F0 differences between a target tone and all other tones in Cantonese (within the first 10ms). E.g. Overall TONE DIFF (C-T1) = (TONE DIFF[C-T1 vs. C-T2) + TONE DIFF(C-T1 vs. C-T3) + TONE DIFF(C-T1 vs. C-T4)) ÷ 3

An ANOVA, where the dependent variable was the scaled F0 difference between tones, and the within-subjects independent variable was TONE PAIR, revealed that there are significant differences for the Overall TONE DIFF of the four Cantonese tones ( $F[3, 42]=24.2, p<0.001$ ). T-tests in Table 29 showed that C-T1(55) has the highest Overall TONE DIFF among all tones, while the second highest is C-T2(25). C-T(33) and C-T4(21) have the lowest values of Overall

TONE DIFF, and there is no significant difference between the Overall TONE DIFF for these two tones.

C-T1 vs. C-T2	t[14]=3.26 p = 0.005 *	C-T2 vs. C-T3	t[14]=4.05 p = 0.001 **
C-T1 vs. C-T3	t[14]=5.85 p <0.001 ***	C-T2 vs. C-T4	t[14]=3.21 p <0.006 **
C-T1 vs. C-T4	t[14]=7.54 p <0.001 ***	C-T3 vs. C-T4	t[14]=0.27 p = 0.787

Table 29. T-tests comparing the Overall TONE DIFF of each Cantonese tone (within the first 10ms)

Considering the observed pattern of tonal differences, the prediction based on Hypothesis 5B is as follows:

PREDICTION 5.3A: The Overall TONE DIFF for each tone (from big to small) is C-T1(55) > C-T2(25) > C-T3(33) = C-T4(21). If Overall TONE DIFF is related to the magnitudes of C-F0, this predicts that the magnitudes of C-F0 (from strong to weak) is as follows: C-F0(C-T1) > C-F0(C-T2) > C-F0(C-T3) = C-F0(C-T4).

### 5.3.2 C-F0 Difference in Cantonese Tones

Prediction 5.3A states that the magnitudes of C-F0 (from strong to weak) is C-F0(C-T1) > C-F0(C-T2) = C-F0(C-T3) = C-F0(C-T4), while for Prediction 5.3B is C-F0(C-T1) > C-F0(C-T2) > C-F0(C-T3) = C-F0(C-T4). Like Mandarin, to evaluate magnitudes of C-F0 in Cantonese, we also examine C-F0 LENGTH and C-F0 DIFF.

The results of C-F0 LENGTH between aspirated obstruents and sonorants (i.e. LENGTH(TH-N)) in Chapter 4 support Prediction 5.3A: C-T1(55) has longer C-F0 duration than the other three tones, while the three tones do not have obvious difference in C-F0 LENGTH. However, neither prediction is supported by the effects on C-F0 LENGTH between unaspirated obstruents and sonorants, since C-T2(25) has a longer C-F0 LENGTH than C-T1(55), but both predictions state that C-T1(55) should have the longest C-F0 LENGTH (Table 30).

Tone C-F0 LENGTH	C-T1(55)	C-T2(25)	C-T3(33)	C-T4(21)
LENGTH (T- N)	10ms	20ms	Non-significant throughout	N/A
LENGTH (TH-N)	35ms	Non-significant throughout	5ms	Non-significant throughout

Table 30. Summary of C-F0 LENGTH results in Question1 for Cantonese. The values of maximum C-F0 LENGTH is the timepoint after which C-F0 becomes non-significant.

For the F0 difference after consonants within the first 10ms, Figure 30 shows that C-F0 DIFF (T-N) in red of C-T2(25) is larger than C-T1(55) and C-T3(33). As for C-F0 DIFF (TH-N) in blue, C-T1(55) has the largest value and C-T2(25) has the smallest.

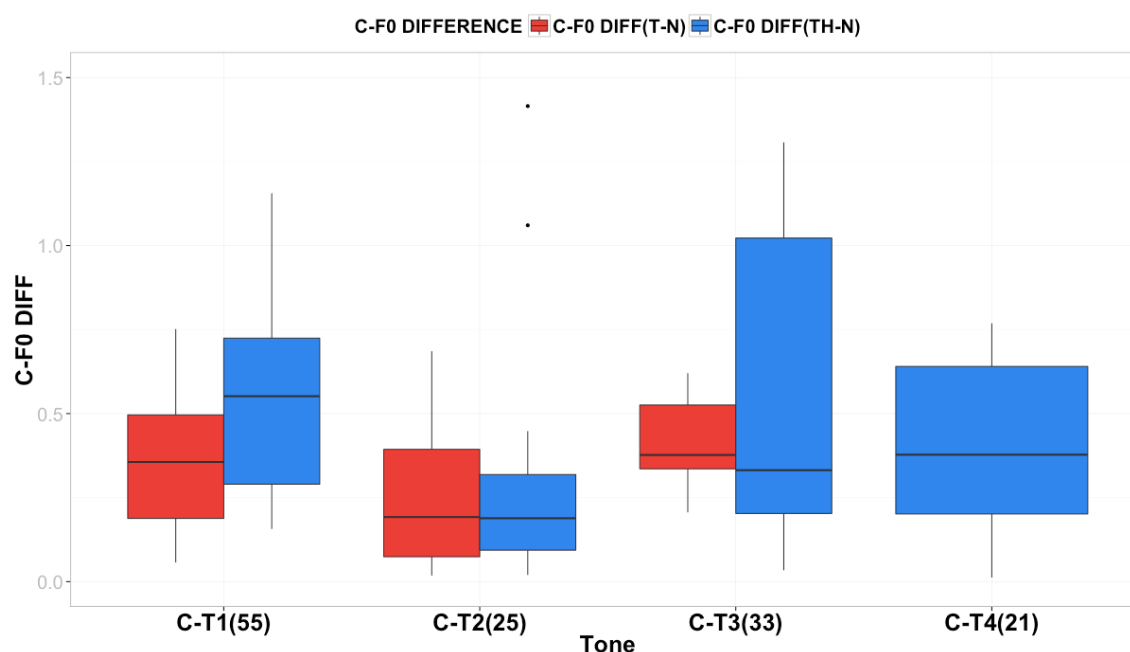


Figure 30. Average C-F0 DIFF in each Cantonese tone (within the first 10ms). The red boxes represent the difference between F0 values after unaspirated obstruents and after sonorants. The blue boxes represent the difference between F0 values after aspirated obstruents and after sonorants. C-T4(21) only has two consonant categories, unaspirated obstruents and sonorants, but no aspirated obstruent.

ANOVAs, where the dependent variable was the scaled F0 difference between tones, and the within-subjects independent variable was TONE PAIR, revealed that the difference was non-significant for C-F0 DIFF (T-N) ( $F[2, 14] = 3.40, p=0.062$ ) but significant for C-F0 DIFF (TH-

N) ( $F[3, 42] = 3.12, p=0.049$ ). T-tests (Table 31) found that there was no statistically significant difference of C-F0 DIFF among C-T1(55), C-T3(33) and C-T4(21), but C-F0 DIFF of C-T2(25) is significantly lower than the other three tones. This is consistent with neither Prediction 5.3A nor Prediction 5.3B. In other words, neither Smallest TONE DIFF nor Overall TONE DIFF influences C-F0 DIFF within the first 10ms in Cantonese.

C-T1 vs. C-T2	$t[14]=3.25$ $p = 0.026 *$	C-T2 vs. C-T3	$t[14]=3.75$ $p = 0.048 *$
C-T1 vs. C-T3	$t[14]=0.98$ $p = 0.363$	C-T2 vs. C-T4	$t[14]=1.23$ $p = 0.032 *$
C-T1 vs. C-T4	$t[14]=0.72$ $p = 0.352$	C-T3 vs. C-T4	$t[14]=0.48$ $p = 0.328$

Table 31. T-tests comparing C-F0 DIFF(TH-N) in each Mandarin tone (within the first 10ms)

In sum, similar to the findings in Mandarin, Prediction 5.3A using Smallest TONE DIFF is consistent with the pattern of C-F0 LENGTH in Cantonese, but neither prediction is consistent with the pattern of C-F0 DIFF within the first 10ms.

### 5.3.3 C-F0 and Cantonese tone distance

The previous Section 5.3.2 has shown that Prediction 5.3A correctly captures the pattern of C-F0 LENGTH (TH-N) for C-T1(55) is higher than the other three tones. However, neither Prediction 5.3A nor 5.3B correctly captures the pattern of C-F0 DIFF within the first 10ms.

The relation between C-F0 DIFF(TH-N) and Smallest TONE DIFF is further examined. The Smallest TONE DIFF is used as the independent variable to predict the dependent variable C-F0 DIFF(TH-N). If tone distance can condition C-F0 DIFF, a positive correlation between them should be expected. Figure 31 shows that there is a negative relation between C-F0 DIFF(TH-N) and the relevant Smallest TONE DIFF for the Cantonese tones except C-T4(21).

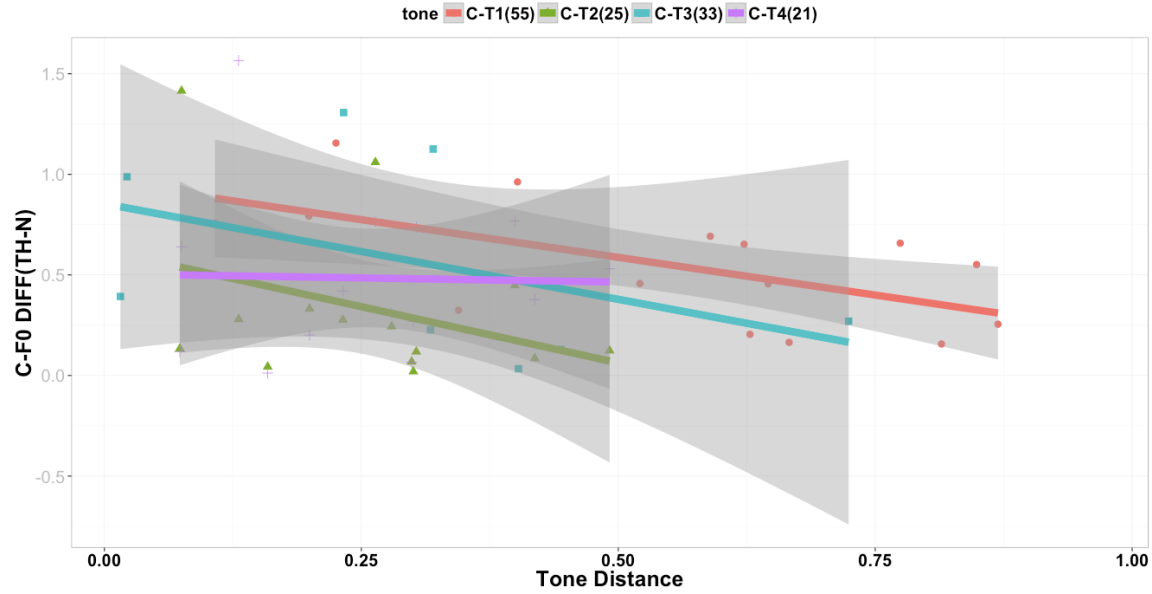


Figure 31. Relation between C-F0 DIFF(TH-N) for all four Cantonese tones and Smallest TONE DIFF of the tones

A series of linear regression tests for the relation between C-F0 DIFF(TH-N) and Smallest TONE DIFF were conducted, shown in Table 32. A negative correlation is found for C-T1(55): a longer tone distance is related to a smaller C-F0 difference in C-T1(55). No other significant relations were found. Therefore, the results suggest that no evidence is found to support that tone distance can condition C-F0 DIFF (in the expected direction).

	$R^2$	F [1, 13]	p
C-T1(55)	0.324	7.739	0.015*
C-T2(25)	0.044	1.601	0.229
C-T3(33)	0.059	1.437	0.275
C-T4(21)	0.082	0.155	0.931

Table 32. Linear Regression tests for the relation between C-F0 DIFF(TH-N) and Smallest TONE DIFF in Cantonese (within the first 10ms)

### 5.3.4 Summary of intra-linguistic comparison in Cantonese

This sub-section has investigated Question 2 on the relation between tone distance and C-F0 in Cantonese. The Smallest TONE DIFF for Cantonese tones were in the following sequence (from

big to small): C-T1 > C-T2 = C-T3 = C-T4. On the other hand, the Overall TONE DIFF were in the following sequence (from big to small): C-T1 > C-T2 > C-T3 = C-T4.

Prediction 5.3A based on the results of Smallest TONE DIFF correctly states that C-F0 LENGTH(TH-N) for C-T1(55) is longer than the three other tones i.e. C-T2(25), C-T3(33) and C-T4(21). However, the C-F0 DIFF results find the F0 difference for C-T2(25) is significantly lower than the three other tones. Neither Prediction 5.3A nor Prediction 5.3B has rightly captured the results for C-F0 DIFF. Moreover, no significant positive correlation was found between Smallest TONE DIFF and C-F0 DIFF.

To summarize, the results of Cantonese indicate that Smallest TONE DIFF may condition C-F0 LENGTH but not C-F0 DIFF. No evidence was found that Overall TONE DIFF is related to the magnitudes of C-F0.

#### **5.4 Cross-linguistic comparison**

This sub-section examines how competitions between tones can be related to different magnitudes of C-F0 across languages. Mandarin has two tones that have high onset pitch: M-T1(55) and M-T4(51), while Cantonese only has one high-level C-T1(55). This suggests that the distinctiveness of M-T1(55) may be threatened by its neighbor M-T4(51) in the onset position, whereas C-T1(55) does not have any neighbor. If a higher degree of tone competition may restrict C-F0, this predicts that C-F0 for C-T1(55) is less restricted than C-F0 for M-T1(55), and thus the former has stronger magnitudes than the latter.

Considering C-F0 LENGTH in Table 33, Cantonese and Mandarin both have the same long C-F0 LENGTH (TH-N), while Cantonese C-F0 LENGTH (T- N) is slightly longer than that of Mandarin.

C-F0 LENGTH \ Tone	C-T1(55)	M-T1(55)
LENGTH (T- N)	10ms	5ms
LENGTH (TH-N)	35ms	35ms

Table 33. Summary of C-F0 LENGTH results in Question1 for Cantonese and Mandarin. The values of maximum C-F0 LENGTH is the timepoint after which C-F0 becomes non-significant.

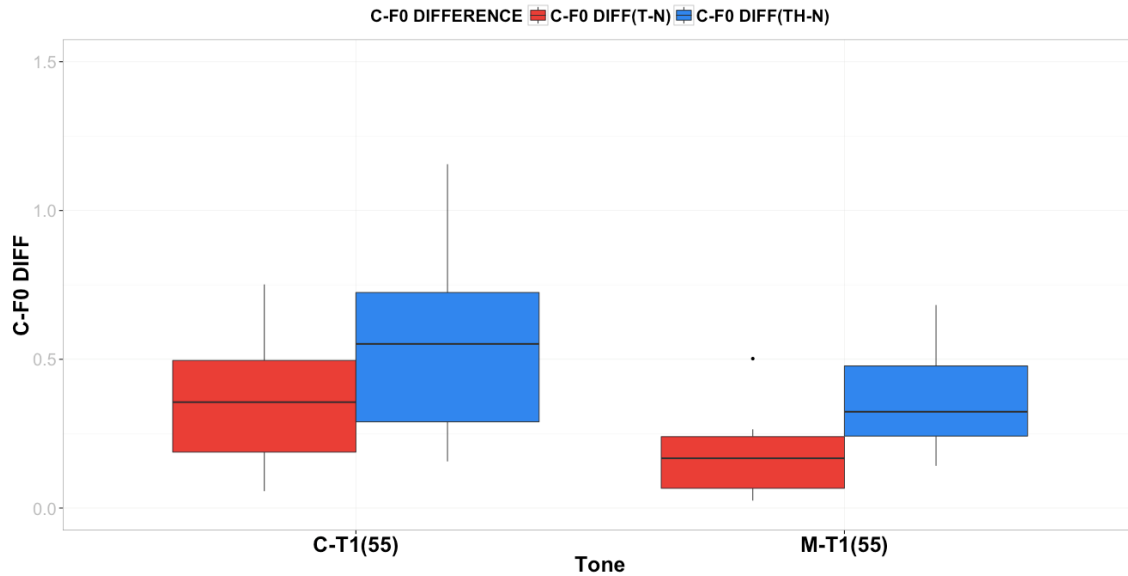


Figure 32. Average C-F0 DIFF in each C-T1(55) and M-T1(55) (within the first 10ms)

Considering C-F0 DIFF in Figure 32, both values of C-F0 DIFF(T-N) and C-F0 DIFF(TH-N) of Cantonese are higher than those of Mandarin. An independent-sample t-test was conducted to compare C-F0 DIFF for C-T1(55) and M-T1(55). There was a marginally significant difference in C-F0 DIFF(TH-N) for C-T1(55) ( $M=0.549$ ) and C-F0 DIFF(TH-N) for M-T1(55) ( $M=0.366$ ) conditions ( $t[28]=2.026$ ,  $p=0.524$ ). There was a marginally significant difference in C-F0 DIFF(T-N) for C-T1(55) ( $M=0.349$ ) and C-F0 DIFF(TH-N) for M-T1(55) ( $M=0.169$ ) conditions ( $t[28]=2.833$ ,  $p=0.008$ ).

In sum, the cross-linguistic comparison of C-F0 LENGTH and C-F0 DIFF suggests that Cantonese has stronger C-F0 magnitudes than Mandarin in the high-level tone context. This indicates that competitions between tones can be related to different magnitudes of C-F0 across languages.

## 5.5 Overall summary for results-Q2

This chapter has answered Research Question 2, which asks whether the acoustic distance between tones is responsible for the different magnitudes of C-F0.

The question is investigated by checking two types of F0 difference between tones: Smallest TONE DIFF (i.e. the F0 difference between the target tone and its closest tone in the tone inventory) and Overall TONE DIFF (i.e. the average value of a target tone's distances from all other tones in the tone inventory). Different predictions were given based on Hypothesis 5A and 5B, which predict that the pattern of C-F0 magnitudes may either follow the pattern of Smallest TONE DIFF or Overall TONE DIFF. Two indicators of C-F0 magnitudes were examined: C-F0 LENGTH and C-F0 DIFF within the first 10ms.

For the findings in Mandarin, the Smallest TONE DIFF from big to small is M-T1 = M-T4 > M-T2 = M-T3, and the Overall TONE DIFF is M-T4 > M-T1 = M-T2 = M-T3. The ranking of Smallest TONE DIFF correctly predicts that C-F0 LENGTH(TH-N) is similar for M-T1(55) and M-T4(51) and for M-T2(35) and M-T3(214). Moreover, C-F0 LENGTH(TH-N) is longer for M-T1(55) and M-T4(51) than for M-T2(35) and M-T3(214). However, for C-F0 DIFF within the first 10ms, neither Smallest TONE DIFF nor Overall TONE DIFF succeeds in predicting that C-F0 DIFF of M-T2(35) is significantly lower than the other three tones. Also, no significant correlation is found between Smallest TONE DIFF and C-F0 DIFF in any of the tones in Mandarin.

For the findings in Cantonese, the Smallest TONE DIFF from big to small is C-T1 > C-T2 = C-T3 = C-T4, and the Overall TONE DIFF is C-T1 > C-T2 > C-T3 = C-T4. The ranking of Smallest TONE DIFF correctly predicts that C-F0 LENGTH(TH-N) for C-T1(55) is longer than that of all other three tones, while there is no significant difference of the C-F0 LENGTH(TH-N)



for the three tones. However, for C-F0 DIFF within the first 10ms, neither Smallest TONE DIFF nor Overall TONE DIFF succeeds in predicting that C-F0 DIFF for C-T2(25) is significantly lower than the other three tones. Also, Smallest TONE DIFF and C-F0 DIFF are not significantly positively correlated in any of the Cantonese tones.

For the cross-linguistic comparison, C-F0 for C-T1(55) is predicted to be less restricted than C-F0 for M-T1(55), since C-T1(55) has no neighbor in the high pitch space while M-T1(5) does. The comparison shows that Cantonese has longer C-F0 LENGTH and a larger C-F0 DIFF than Mandarin in the high-level tone context. The findings indicate that different magnitudes of C-F0 across languages can be related to tonal neighborhood density.

## **6. Results for Question 3: cue trading between VOT and F0?**

### **6.1 Introduction**

This chapter addresses Research Question 3 stated in Chapter 1: is there a trading relation between VOT and onset F0 as cues for the aspiration contrast of the initial consonants? The cue trading relation is investigated in specific tone contexts, for tone can be related to not only F0 values but also VOT values.

The results for Mandarin are shown in Section 6.2. Section 6.2.1 shows that the VOT difference between aspirated and unaspirated is significant in every Mandarin tone context, while the F0 difference is significant only in M-T1(55) and M-T4(51), suggesting that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue. Moreover, VOT for the aspirated stops in the contexts of M-T2(35) and M-T3(214) is longer than that for M-T1(55) and M-T4(51). However, the findings of Sections 6.2.2 – 6.2.5 show that there is a significant negative correlation between VOT and onset F0 for the unaspirated category in M-T2(35), with a poor fit ( $R^2=0.05$ ). Positive correlations were found in the aspirated category in M-T1(55), M-T3(214) and M-T4(51). These findings do not support a cue trading relation between VOT and onset F0 in Mandarin.

The results for Cantonese are reported in Section 6.3. Section 6.3.1 showing that VOT values are significantly different between aspirated and unaspirated in each of the three Cantonese tones (i.e. C-T1(55), C-T2(25), and C-T3(33)), while the F0 values are significantly different only in C-T1(55) and C-T2(25), but non-significant in C-T3(33). The magnitudes of acoustic differences suggest that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue. Moreover, tones do not have a significant effect on VOT. Sections 6.3.2 – 6.3.4 present results showing that there was no significant correlation between VOT and F0 in any of

the three Cantonese tone contexts, which indicates that a trading relation is unlikely between VOT and F0 in Cantonese.

A summary of the results in both languages is provided in Section 6.4.

## **6.2 Cue trading between VOT and F0 in Mandarin?**

This section investigates the relation between VOT and onset F0 values in each Mandarin tone. Section 6.2.1 shows that VOT is a strong cue for the aspiration contrast of consonants, while the onset F0 difference is a relatively weak one. Sections 6.2.2 – 6.2.5 show that there is a significant negative correlation between VOT and onset F0 in the unaspirated category in M-T2(35), with a poor fit ( $R^2=0.05$ ). Positive correlations were found in the aspirated category in M-T1(55), M-T3(214) and M-T4(51). No strong evidence was found to support a cue trading relationship in any of the tone contexts in Mandarin. Section 6.2.6 provides a summary.

### **6.2.1 VOT and F0 as cues for the aspiration contrast in Mandarin tone contexts**

#### **6.2.1.1 VOT as a strong cue for the aspiration contrast in Mandarin**

Previous studies have shown that VOT can be a robust phonetic cue for aspiration contrasts (Cho & Ladefoged, 1999; Keating, 1984; Lisker & Abramson, 1964). Our results also support this. Figure 33 shows the z-scores of VOT values for aspirated and unaspirated obstruents (i.e. VOT[aspirated] and VOT[unaspirated]) in each Mandarin tone context. In each tone context, aspirated obstruents have obvious higher VOT values than the unaspirated counterparts.

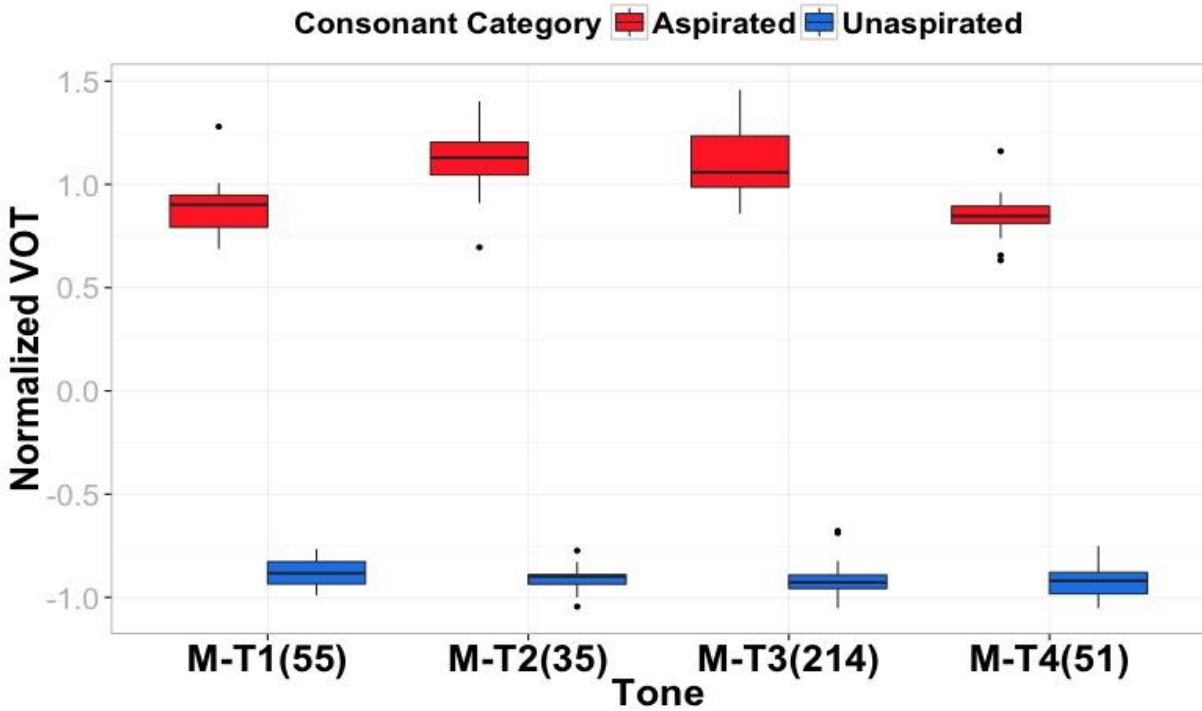


Figure 33. Normalized VOT of aspirated and unaspirated obstruents across Mandarin tones

The ANOVA (Table 34), where the dependent variable was VOT, and the within-subjects independent variable was CONSONANT and TONE, revealed that there was a significant main effect for CONSONANT, TONE and a significant interaction. The follow-up ANOVA tests found significant differences for VOT[aspirated] in the context of different tones ( $F[3, 42]=13.66$ ,  $p<0.001$ ) but not VOT[unaspirated] ( $F[3, 42]=1.68$ ,  $p=0.185$ ). The t-tests (Table 35) found that VOT[aspirated] in the contexts of M-T2(35) and M-T3(214) were significantly longer than those in M-T1(55) and M-T4(51). No other comparisons were significant.

Factor \ Value	F	P
CONSONANT	[1, 112]=69.33	<0.001 ***
TONE	[3, 112]=9.68	<0.001 ***
TONE * CONSONANT	[3, 112]=9.32	<0.001 ***

Table 34. ANOVA test for average VOT values of aspirated and unaspirated obstruents across Mandarin tones

Tone Pair \ Value	t [14]	P
M-T1 vs. M-T2	4.11	<0.001 ***
M-T1 vs. M-T3	3.66	0.002 **
M-T1 vs. M-T4	1.13	0.275
M-T2 vs. M-T3	0.15	0.886
M-T2 vs. M-T4	5.44	<0.001 ***
M-T3 vs. M-T4	4.79	<0.001 ***

Table 35. T-tests comparing normalized VOT of ASPIRATED obstruents in different tone contexts

The results show that, firstly, VOT values can clearly separate aspirated and unaspirated categories in Mandarin, and thus is a robust acoustic cue for the aspiration contrast of the obstruents. Secondly, tones can have an influence on VOT, but the effect is restricted only for aspirated stops. VOT[aspirated] in the contexts of tones with low-pitch onset (i.e. M-T2(35) and M-T3(214)) is longer than in tones with high-pitch onsets.

#### 6.2.1.2 Onset F0 as a weak cue for the aspiration contrast in Mandarin

Figure 34 shows the normalized F0 values after aspirated and unaspirated obstruents (i.e. F0[aspirated] and F0[unaspirated]) within the first 10ms of the vowels in each Mandarin tone context. In each tone context, aspirated stops have higher mean F0 values than the unaspirated counterparts, except in M-T2(35) where the difference is slight, though still in the same direction as the other tones.

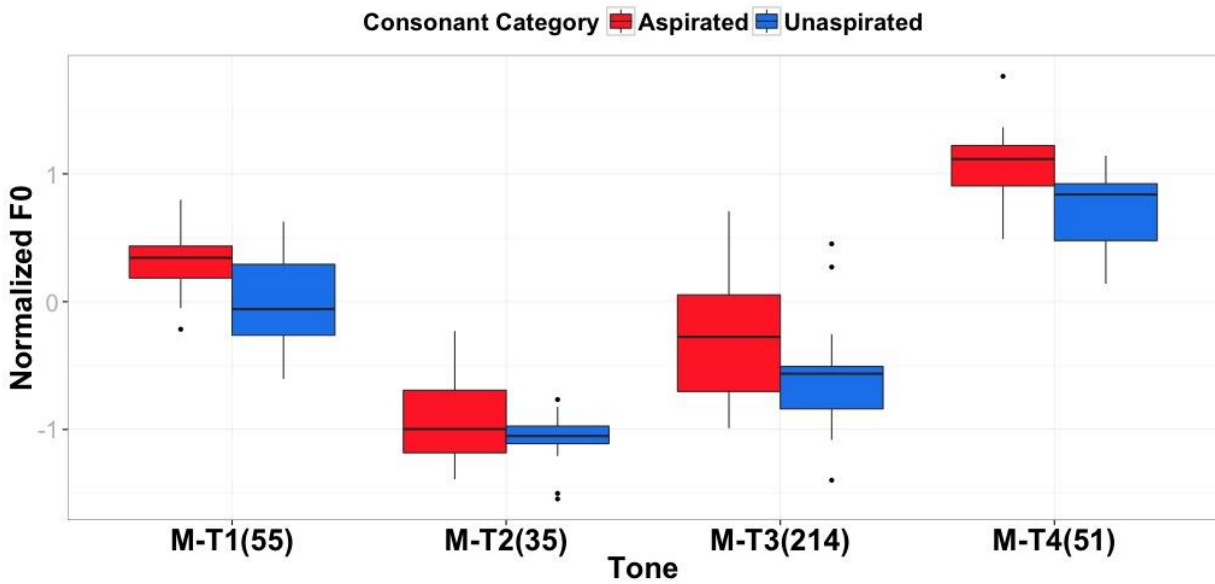


Figure 34. F0 values(within the first 10ms) after aspirated and unaspirated obstruents across Mandarin tones

The ANOVA (Table 36), where the dependent variable was F0, and the within-subjects independent variable was CONSONANT and TONE, reveals that there is a significant main effect for CONSONANT types on the F0 values and for TONE, but no significant interaction. The follow-up t-tests (Table 37) show that the F0[aspirated] is significantly higher than F0[unaspirated] in M-T1(55) and M-T4(51), but non-significant in M-T2(35) and M-T3(214), though they show the same direction of difference as the former two tones.

Factor \ Value	F	P
CONSONANT	[1, 112]=17.85	<0.001 ***
TONE	[3, 112]=150.16	<0.001 ***
TONE * CONSONANT	[3, 112]=0.54	0.653

Table 36. ANOVA test for average F0 values (the first 10ms) after aspirated and unaspirated obstruents across Mandarin tones

Tone \ Value	t[14]	p
M-T1(55)	3.36	0.005 **
M-T2(35)	1.14	0.271
M-T3(214)	1.56	0.141
M-T4(51)	3.56	0.003 **

Table 37. T-tests comparing average F0 values (the first 10ms) after aspirated and unaspirated obstruents in each Mandarin tone

The results indicate that onset F0 is a relatively weak cue for the aspiration contrast in Mandarin compared to VOT, for the difference between F0[aspirated] and F0[unaspirated] is only significant in M-T1(55) and M-T4(51), comparing that the difference between VOT[aspirated] and VOT[unaspirated] is significant across all Mandarin tones. Moreover, the difference of the z-scored values between VOT[aspirated] and VOT[unaspirated] is much larger than the difference between F0[aspirated] and F0[unaspirated]. Therefore, it is safe to conclude that VOT is a much stronger cue for the aspiration contrast than onset F0 is in Mandarin.

The findings of Section 6.2.1 also suggest that M-T2(35) and M-T3(214) have a much longer VOT(aspirated) to cue the aspiration contrast, and therefore there may be less in need to rely on onset F0 as a redundant cue in these two tone contexts. On the other hand, M-T1(55) and M-T4(51) have a shorter VOT(aspirated) to cue the aspiration contrast, and therefore onset F0 can be a supplement to cue the contrast. This hypothesis predicts that the more the reliance on VOT as a cue for the aspiration contrast, the less on onset F0 – which is a cue trading relation. Sections 6.2.2 – 6.2.5 seek to investigate whether there is such relation in each tone context.

### 6.2.2 VOT and F0 for M-T1(55)

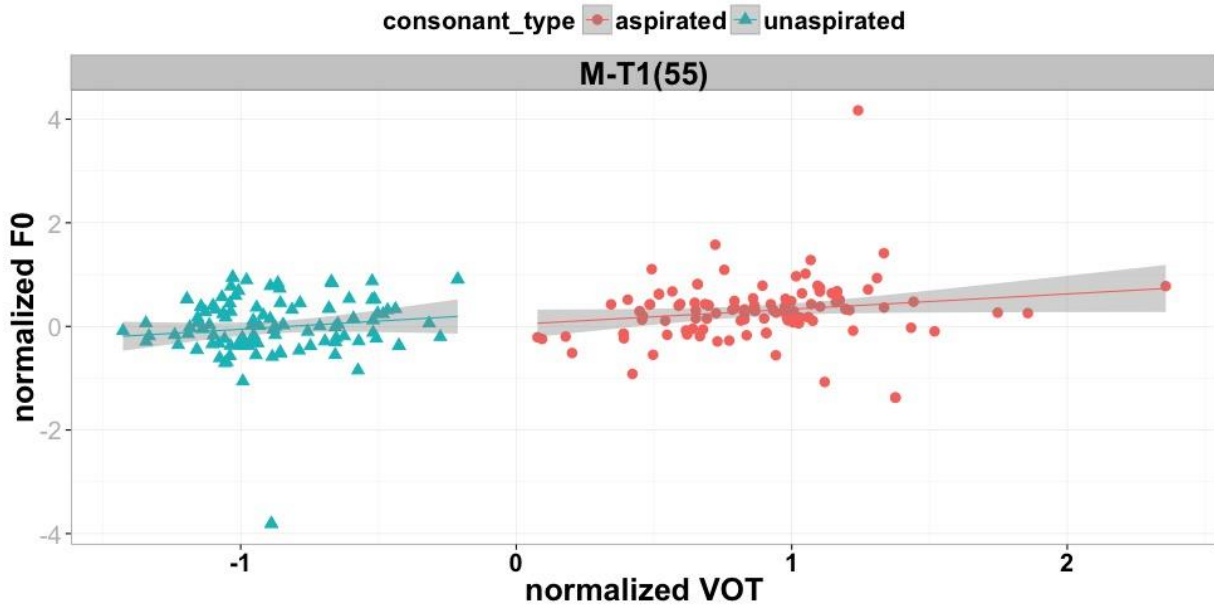


Figure 35. The relation between normalized VOT and F0 (within the first 10ms) for M-T1(55)

For the relation between VOT and F0 in M-T1(55) in the unaspirated context, a linear regression was conducted to predict F0[unaspirated] based on VOT[unaspirated]. The independent variable VOT[unaspirated] had a non-significant proportion of variance in the dependent variable F0[unaspirated] ( $R^2=0.007$ ,  $F(1, 119)=1.80$ ,  $p=0.182$ ) (left section of Figure 35).

On the other hand, for the relation between VOT and F0 in M-T1(55) in the aspirated context, the linear regression found that the independent variable VOT[aspirated] was positively correlated with the dependent variable F0[aspirated] ( $\beta=0.29$ ,  $t(118)=1.94$ ,  $p=0.052$ ). The independent variable also explained a marginally significant proportion of variance in the dependent variable ( $R^2=0.022$ ,  $F(1, 118)=3.76$ ,  $p=0.052$ ) (right section of Figure 35).

The results show no significant correlation between VOT and onset F0 for the unaspirated category in M-T1(55), but a positive correlation between the two acoustic



dimensions for the aspirated. It suggests that there is unlikely to be a cue trading relation between VOT and F0 in M-T1(55).

### 6.2.3 VOT and F0 for M-T2(35)

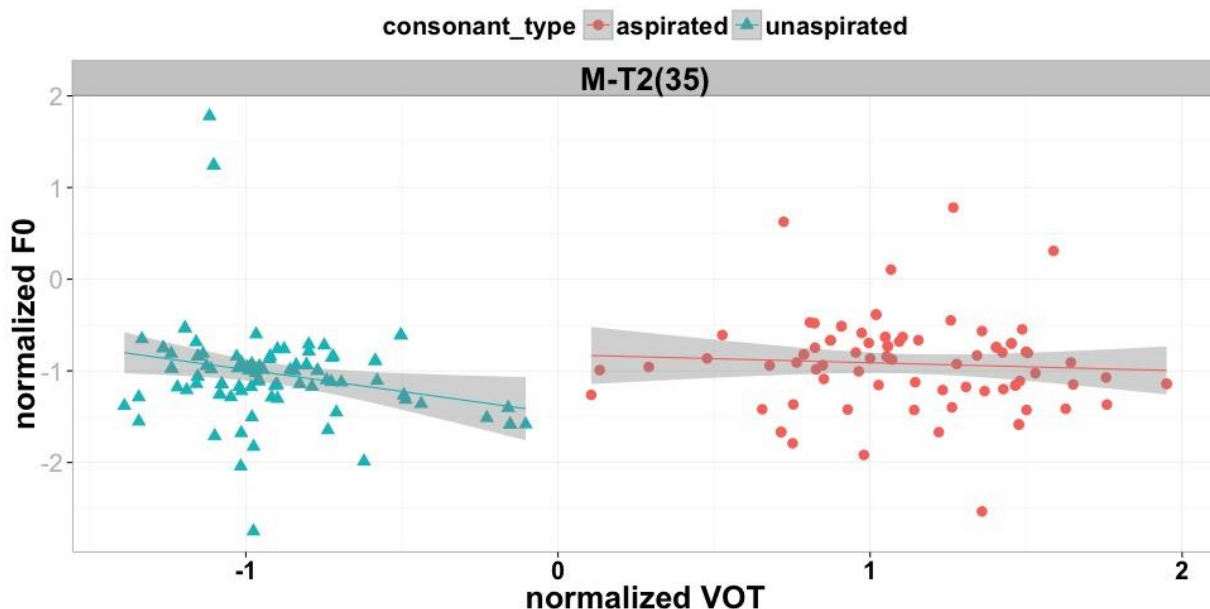


Figure 36. The relation between normalized VOT and F0 (within the first 10ms) for M-T2(35)

For the relation between VOT and F0 in M-T2(35) in the unaspirated context, a linear regression was conducted to predict F0[unaspirated] based on VOT[unaspirated] ( $\beta = -0.47$ ,  $t(94) = -2.321.94$ ,  $p = 0.022$ ). The independent variable VOT[unaspirated] explained a significant proportion of variance in the dependent variable F0[unaspirated] ( $R^2 = 0.05$ ,  $F(1, 94) = 1.80$ ,  $p = 0.022$ ) (left section of Figure 36).

For the aspirated category, for the relation between VOT and F0, the linear regression did not find any significant proportion of variance of the independent variable VOT[aspirated] in the dependent variable F0[aspirated] ( $R^2 = 0.004$ ,  $F(1, 83) = 0.37$ ,  $p = 0.544$ ) (right section of Figure 36).

The results for M-T2(35) showed a significant negative correlation between VOT and onset F0 for the unaspirated category, but no significant correlation between the two acoustic dimensions for the aspirated. The findings may suggest a cue trading relation between VOT and F0 for the unaspirated category in M-T2(35).

#### 6.2.4 VOT and F0 for M-T3(214)

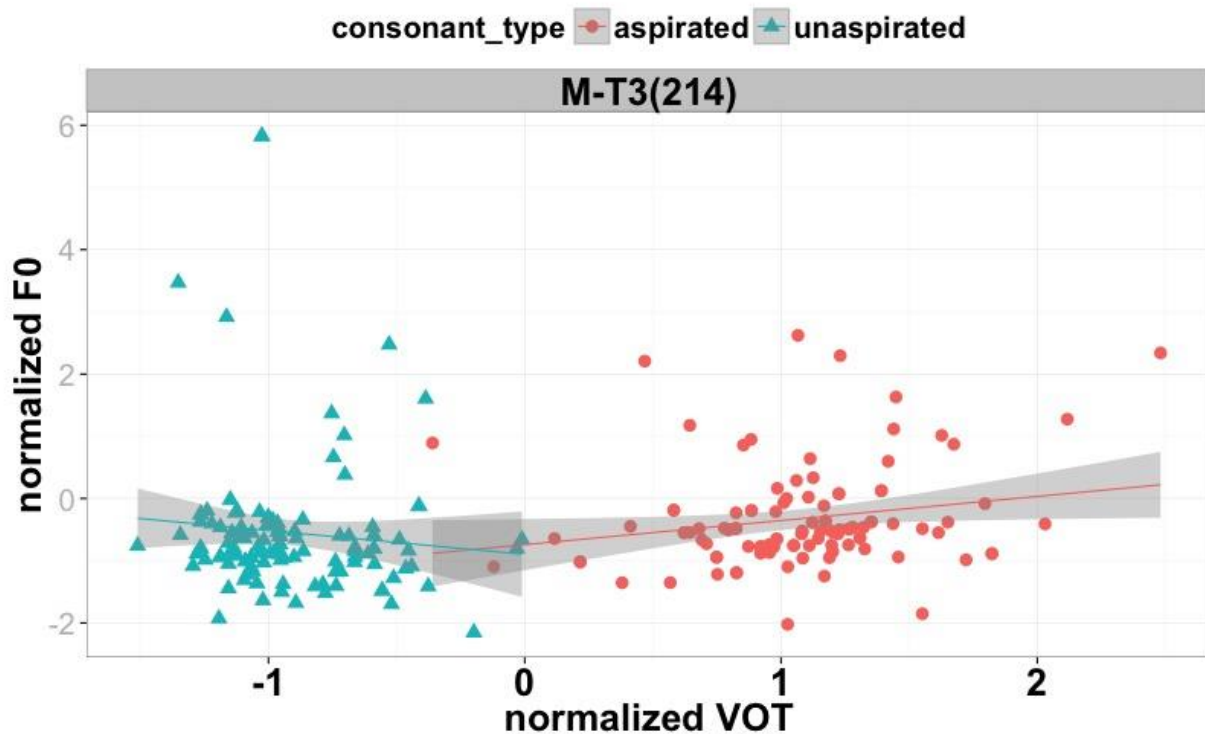


Figure 37. The relation between normalized VOT and F0 (within the first 10ms) for M-T3(214). For the relation between VOT and F0 in M-T3(214) in the unaspirated context, the linear regression showed that the independent variable VOT[unaspirated] explained a non-significant proportion of variance in the dependent variable F0[unaspirated] ( $R^2=0.009$ ,  $F(1, 120)=1.09$ ,  $p=0.297$ ) (left section of Figure 37).

On the other hand, for the relation between VOT and F0 in M-T3(214) in the aspirated context, the linear regression found the independent variable VOT[aspirated] was positively

correlated with the dependent variable  $F0[\text{aspirated}]$  ( $\hat{\beta}=0.39$ ,  $t(107)=2.14$ ,  $p=0.034$ ). A significant proportion of variance of  $F0[\text{aspirated}]$  was explained by  $VOT[\text{aspirated}]$  ( $R^2=0.032$ ,  $F(1, 107)=4.59$ ,  $p=0.034$ ) (right section of Figure 37).

The results show no significant correlation between VOT and onset F0 for the unaspirated category in M-T3(214), but a positive correlation between the two acoustic dimensions for the aspirated. It further suggests that there is unlikely to be a cue trading relation between VOT and F0 in M-T3(214).

### 6.2.5 VOT and F0 for M-T4(51)

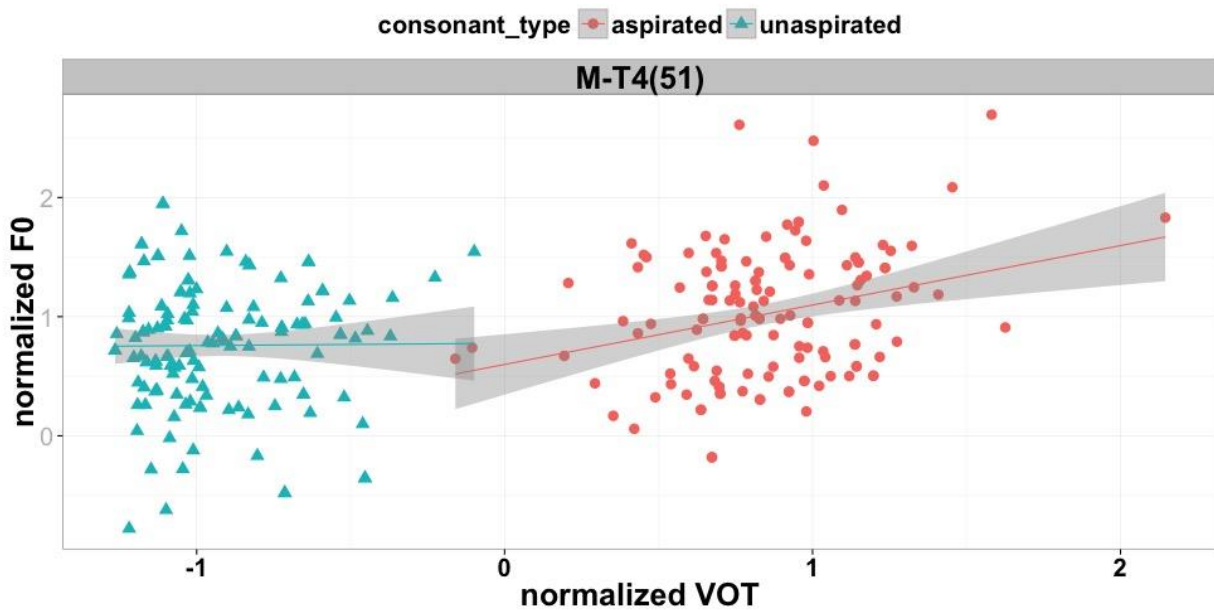


Figure 38. The relation between normalized VOT and F0 (within the first 10ms) for M-T4(51). For the relation between VOT and F0 in M-T4(51) in the unaspirated context, the linear regression found that the independent variable  $VOT[\text{unaspirated}]$  explained a non-significant proportion of variance in the dependent variable  $F0[\text{unaspirated}]$  ( $R^2=0.007$ ,  $F(1, 135)=0.011$ ,  $p=0.918$ ) (left section of Figure 38).

For the relation between VOT and F0 in M-T4(51) in the aspirated context, the linear regression found that the independent variable VOT[aspirated] explained a significant proportion of variance in the dependent variable F0[aspirated] ( $R^2=0.078$ ,  $F(1, 135)=12.52$ ,  $p<0.001$ ). VOT[aspirated] was positively correlated F0[aspirated] ( $\beta=0.50$ ,  $t(135)=3.54$ ,  $p<0.001$ ) (right section of Figure 38).

The results show no significant correlation between VOT and onset F0 for the unaspirated category in M-T4(51), but a positive correlation between the VOT and onset F0 for the aspirated. They suggests that there is no cue trading relation between VOT and F0 in M-T4(51).

#### **6.2.6 Summary of results-Q3 for Mandarin**

To summarize the Mandarin results for Question 3: section 6.2.1 shows that (1) the VOT difference between aspirated and unaspirated is significant in every Mandarin tone context, while the F0 difference is significant only in M-T1(55) and M-T4(51). The magnitudes of acoustic differences suggest that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue; (2) VOT(aspirated) for M-T2(35) and M-T3(214) is longer than that for M-T1(55) and M-T4(51).

The findings of Section 6.2.1 may suggest that M-T2(35) and M-T3(214) have a much longer VOT(aspirated) to cue the aspiration contrast, and therefore there is no need to rely on onset F0 as a redundant cue. On the other hand, M-T1(55) and M-T4(51) have a shorter VOT(aspirated) to cue the aspiration contrast, and therefore there can be more reliance on onset F0 to cue the contrast in two contexts.

However, the findings of Sections 6.2.2 – 6.2.5 (Table 38) have not found any significant negative correlation between VOT and onset F0 except in the unaspirated category in M-T2(35),

with a poor fit ( $R^2=0.05$ ). Positive correlations were found in the aspirated category in M-T1(55), M-T3(214) and M-T4(51). These findings do not support the cue trading hypothesis in any of the tone contexts in Mandarin.

Consonant Tone	Unaspirated	Aspirated
M-T1(55)	Non-significant Correlation	Significant Positive Correlation
M-T2(35)	Significant Negative Correlation	Non-significant Correlation
M-T3(214)	Non-significant Correlation	Significant Positive Correlation
M-T4(51)	Non-significant Correlation	Significant Positive Correlation

Table 38. A summary of the linear regression results for the relation between VOT and F0 in four Mandarin tones

### 6.3 Cue trading between VOT and F0 in Cantonese?

This section investigates the relation between VOT and onset F0 values in each Cantonese tone. Since three Cantonese tones, C-T1(55), C-T2(25), and C-T3(33), have the aspiration (or laryngeal) contrast in our data, while the other three Cantonese tones (i.e. C-T4(21), C-T5(23), and C-T6(22)) have only one laryngeal category, only the former three Cantonese tones are examined. Section 6.3.1 finds that the VOT is a robust acoustic cue for the aspiration contrast in Cantonese tone, while the onset F0 is a relatively weak cue. Also, Cantonese tones do not have a significant influence on VOT. Sections 6.3.2 – 6.3.4 show that there are no statistically significant correlations between VOT and onset F0 in any of the three Cantonese tone contexts, which indicates that there is no trading relation between VOT and onset F0 in Cantonese. Section 6.3.5 provides a summary.

#### 6.3.1 VOT and F0 as cues for the aspiration contrast in Cantonese tone contexts

##### 6.3.1.1 VOT as a strong cue for the aspiration contrast in Cantonese

Figure 39 shows normalized VOT[aspirated] and VOT[unaspirated] in three Cantonese tone contexts. In each tone, aspirated obstruents have obvious higher VOT values than the unaspirated

counterparts. The results suggest that VOT can clearly separate aspirated stops and unaspirated ones in Cantonese, which is supported by previous literature (Cho & Ladefoged, 1999; Keating, 1984; Lisker & Abramson, 1964), and the difference between VOT[aspirated] and VOT[unaspirated] is large in each tone context. However, unlike Mandarin, Cantonese tones do not have a differential effect on the VOT values.

The ANOVA (Table 39), where the dependent variable was VOT, and the within-subjects independent variable was CONSONANT and TONE, found that there was a significant main effect for CONSONANT types, but not significant for TONE or their interaction.

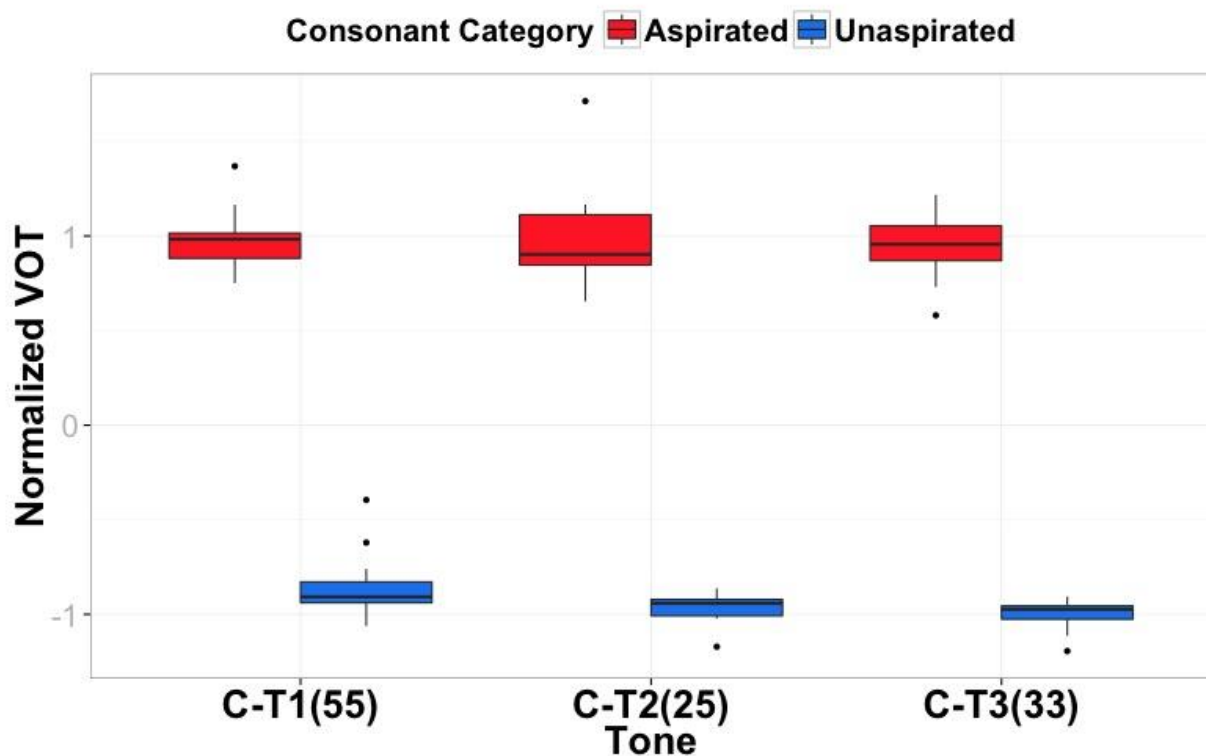


Figure 39. Normalized VOT of aspirated and unaspirated obstruents across Cantonese tones

Factor \ Value	F	P
CONSONANT	[1, 84]=32.89	<0.001 ***
TONE	[2, 84]=2.341	0.102
TONE * CONSONANT	[2, 84]=1.26	0.288

Table 39. ANOVA test for average VOT values of aspirated and unaspirated obstruents across Cantonese tones

### 6.3.1.2 Onset F0 as a weak cue for the aspiration contrast in Cantonese

Figure 40 shows the normalized F0[aspirated] and F0[unaspirated] within the first 10ms of the vowels in each Cantonese tone context. In each tone context, F0[aspirated] is higher than F0[unaspirated].

The ANOVA (Table 40) found that there was a significant main effect for CONSONANT type on the F0 values and for TONE, but no significant interaction. The follow-up t-tests (Table 41) showed that the F0[aspirated] was significantly higher than F0[unaspirated] in C-T1(55) and C-T2(25), but non-significant in C-T3(33); however, all three tones showed the same direction of difference.

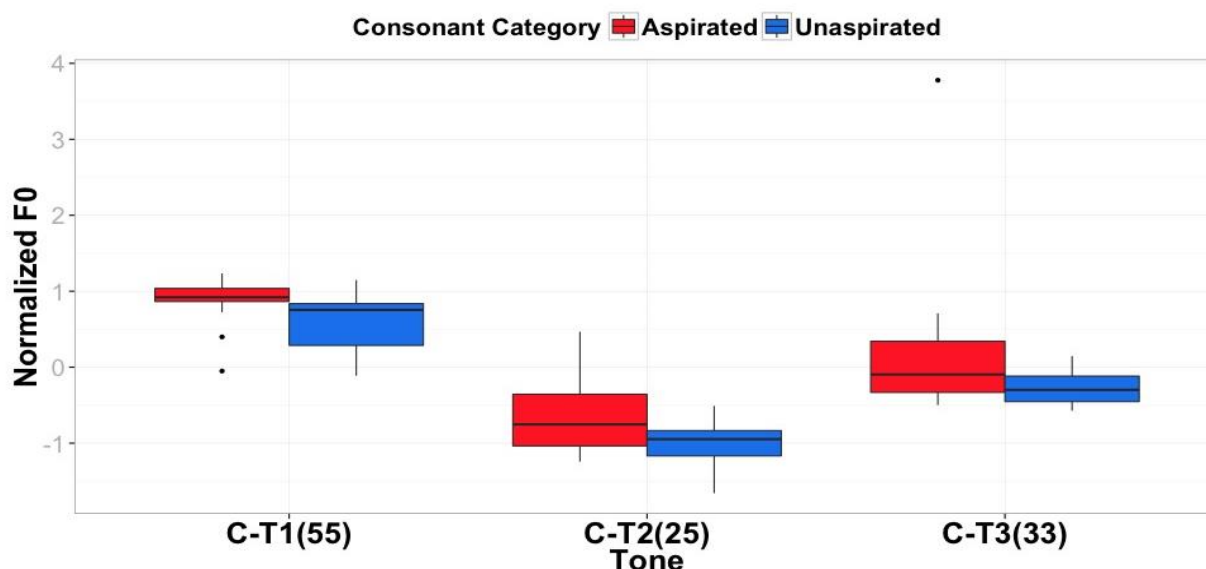


Figure 40. F0 values (within the first 10ms) after aspirated and unaspirated obstruents across Cantonese tones

Factor \ Value	F	P
CONSONANT	[1, 84]=11.69	<0.001 ***
TONE	[2, 84]=60.26	<0.001 ***
TONE * CONSONANT	[2, 84]=0.31	0.730

Table 40. ANOVA test for average F0 values (the first 10ms) after aspirated and unaspirated obstruents across Cantonese tones

Tone \ Value	t[14]	p
C-T1(55)	3.08	0.008 **
C-T2(25)	2.59	0.021 *
C-T3(33)	1.71	0.110

Table 41. T-tests comparing average F0 values (the first 10ms) after aspirated and unaspirated obstruents in each Cantonese tone

The results indicate that, like in Mandarin, onset F0 is a relatively weak cue for the aspiration contrast in Cantonese compared to VOT, for the difference between F0[aspirated] and F0[unaspirated] between is only significant in C-T1(55) and C-T2(25), whereas the difference between VOT[aspirated] and VOT[unaspirated] is significant across all the three Cantonese tones. Also, there is a bigger scaled difference between VOT[aspirated] and VOT[unaspirated] than the difference between F0[aspirated] and F0[unaspirated]. Therefore, onset F0 is a weaker cue for the aspiration contrast than VOT in Cantonese.

### 6.3.2 VOT and F0 for C-T1(55)

For the relation between VOT and F0 in C-T1(55) in the unaspirated context, the linear regression found that the independent variable VOT[unaspirated] explained a non-significant proportion of variance in the dependent variable F0[unaspirated] ( $R^2=0.006$ ,  $F(1, 149)=0.056$ ,  $p=0.814$ ) (left section of Figure 41).

Likewise, for the relation between VOT and F0 in the aspirated context, the linear regression showed that the independent variable VOT[aspirated] does not explain a significant



proportion of variance in the dependent variable F0[aspirated] ( $R^2=0.000$ ,  $F(1, 155)=0.94$ ,  $p=0.333$ ) (right section of Figure 41).

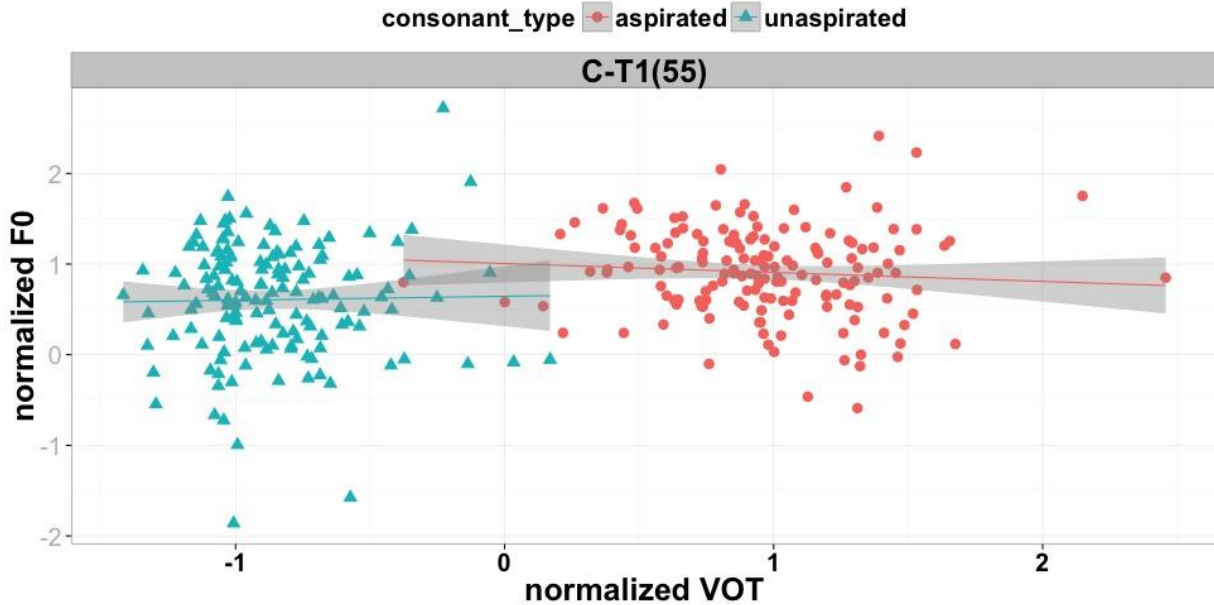


Figure 41. The relation between normalized VOT and F0 (within the first 10ms) for C-T1(55)

The results show no significant correlation between VOT and onset F0 for the unaspirated or the aspirated category in C-T1(55). They further suggest that there is no cue trading relation between VOT and F0 in C-T1(55).

### 6.3.3 VOT and F0 for C-T2(25)

For the relation between VOT and F0 in C-T2(25) in the unaspirated context, the linear regression found that the independent variable VOT[unaspirated] does not explain any significant proportion of variance in the dependent variable F0[unaspirated] ( $R^2=0.010$ ,  $F(1, 74)=0.244$ ,  $p=0.622$ ) (left of Figure 42).

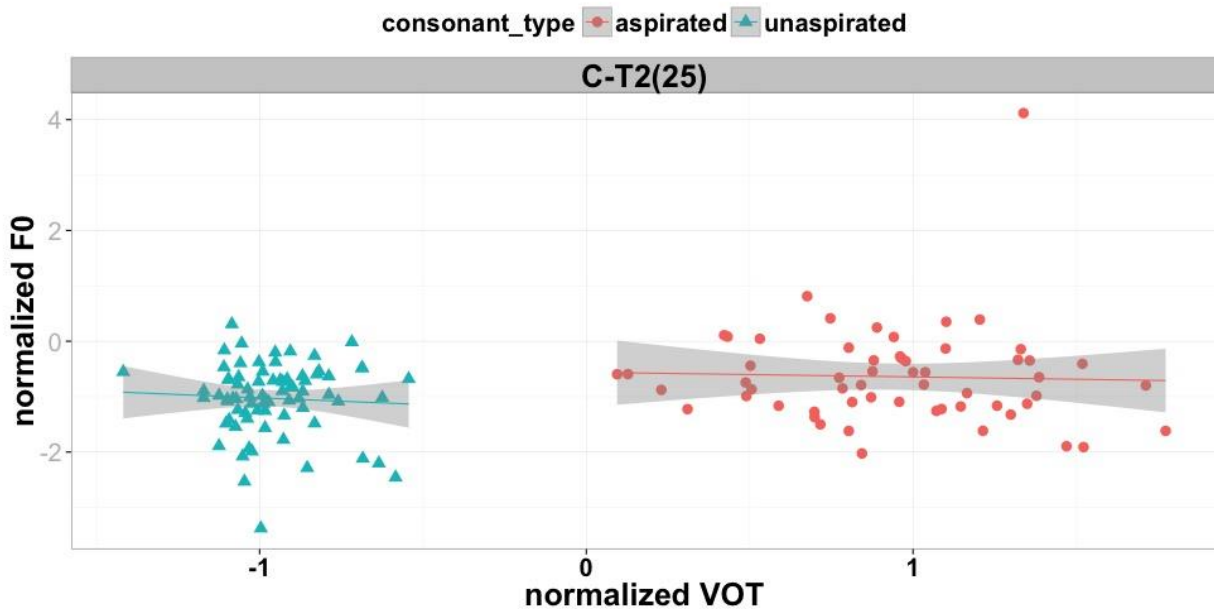


Figure 42. The relation between normalized VOT and F0 (within the first 10ms) for C-T2(25)

Similarly, for the relation between VOT and F0 in the aspirated context, the linear regression found that the independent variable VOT[aspirated] does not explain any significant proportion of variance of in the dependent variable F0[aspirated] ( $R^2=0.017$ ,  $F(1, 56)=0.07$ ,  $p=0.788$ ) (right of Figure 42).

Like C-T1(55), the results show no significant correlation between VOT and onset F0 for the unaspirated or the aspirated category in C-T2(25). They further suggest that there is no cue trading relation between VOT and F0 in C-T2(25).

#### 6.3.4 VOT and F0 for C-T3(33)

For the relation between VOT and F0 in C-T3(33) in the unaspirated context, the linear regression found that the independent variable VOT[unaspirated] does not explain any significant proportion of variance of in the dependent variable F0[unaspirated] ( $R^2=0.012$ ,  $F(1, 76)=0.092$ ,  $p=0.762$ ). The same non-significant findings also apply to the aspirated context, where the independent variable VOT[aspirated] does not explain a significant proportion of

variance in the dependent variable F0[aspirated] ( $R^2=0.004$ ,  $F(1, 67)=1.32$ ,  $p=0.253$ ) (Figure 43). The results show no significant correlation between VOT and onset F0 for the unaspirated or the aspirated category in C-T3(33). They also suggest that there is no cue trading relation between VOT and F0 in C-T3(33).

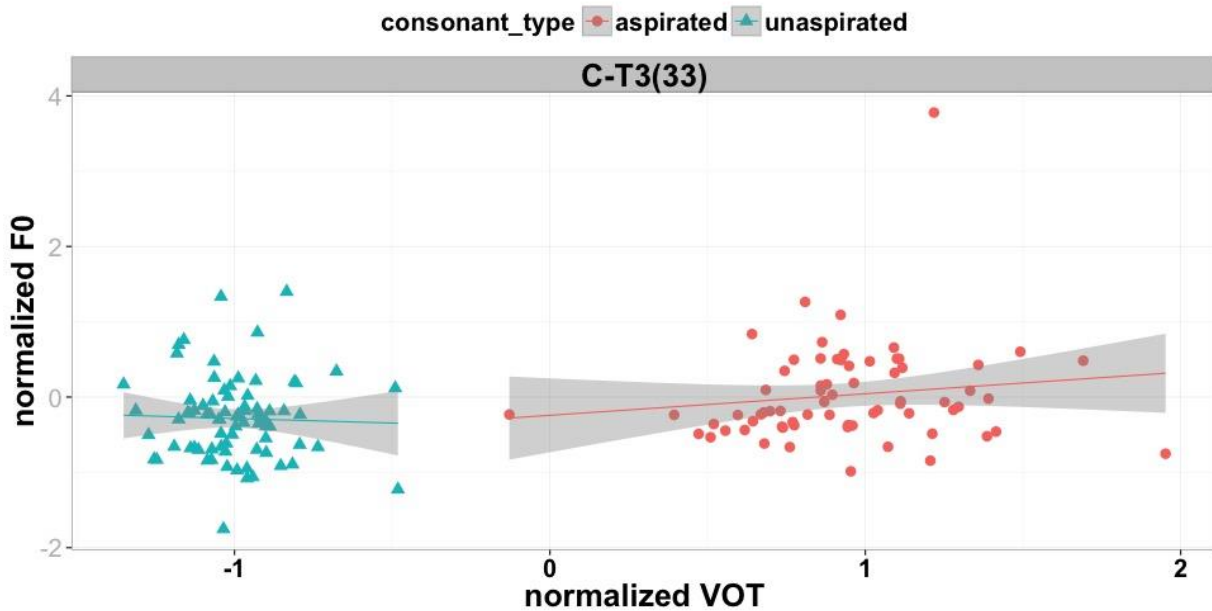


Figure 43. The relation between normalized VOT and F0 (within the first 10ms) for C-T3(33)

### 6.3.5 Summary of results-Q3 for Cantonese

To summarize the Cantonese results for Question 3, section 6.3.1 shows that (1) the VOT difference between aspirated and unaspirated consonants is significant in every Cantonese tone context, while the F0 difference is significant only in C-T1(55) and C-T2(25), but non-significant in C-T3(33). wea suggest that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue; (2) tones do not have a significant influence on VOT.

The findings of Sections 6.3.2 – 6.3.4 have not found any significant correlations between VOT and F0 in any of the three Cantonese tone contexts, which indicate that there is unlikely a trading relation between these two acoustic dimensions in Cantonese (Table 42).

Consonant Tone	Unaspirated	Aspirated
C-T1(55)	Non-significant Correlation	Non-significant Correlation
C-T2(25)	Non-significant Correlation	Non-significant Correlation
C-T3(33)	Non-significant Correlation	Non-significant Correlation

Table 42. A summary of the linear regression results for the relation between VOT and F0 in three Cantonese tones

#### 6.4 Overall summary of results-Q3

This chapter has addressed Research Question 3, which asks whether there a trading relation between VOT and F0 as cues for the aspiration contrast of the initial consonants.

The results of Mandarin show that the VOT difference between aspirated and unaspirated consonants is significant in every Mandarin tone context, while the F0 difference is significant only in M-T1(55) and M-T4(51), suggesting that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue. Moreover, VOT for the aspirated stops in the contexts of M-T2(35) and M-T3(214) is longer than that for M-T1(55) and M-T4(51), suggesting that the former two tones may have less need to use F0 as a secondary cue for the aspiration contrast. However, positive correlations were found in the aspirated category in M-T1(55), M-T3(214) and M-T4(51), while a significant negative correlation between VOT and onset F0 for the unaspirated category was found in M-T2(35), with a poor fit ( $R^2=0.05$ ). These findings do not provide strong evidence for a cue trading relation between VOT and onset F0 in Mandarin.

The results of Cantonese show that VOT values are significantly different between aspirated and unaspirated consonants in each of the three Cantonese tones tested (i.e. C-T1(55), C-T2(25), and C-T3(33)), while the F0 values are significantly different only in C-T1(55) and C-T2(25), but non-significant in C-T3(33). The magnitudes of acoustic differences suggest that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue. Moreover, no significant

correlation between VOT and F0 is found in any of the three Cantonese tone contexts, which indicates that a trading relation between VOT and F0 is unlikely in Cantonese.

## **7. General discussion and conclusion**

### **7.1 Summary of major findings and discussion**

This dissertation has asked three questions about C-F0 in tonal languages:

- (1) Questions 1: Is C-F0 conditioned by lexical tones?
- (2) Questions 2: Is C-F0 conditioned by F0 difference between tones?
- (3) Questions 3: Is there a cue trading relation between F0 and VOT for laryngeal contrast?

For Research Question 1, the major findings are that tones can influence C-F0 LENGTH (the vowel duration that C-F0 can extend) and C-F0 DIFF (the difference between F0 following different consonants) in Mandarin and Cantonese. There was a consistent C-F0 DIRECTION across all Mandarin tones and C-T1(55) and C-T3(33): F0[aspirated] was higher than F0[unaspirated] and F0[sonorant]. The trajectory of F0[unaspirated] started as the second highest and converged with the lowest baseline F0[sonorant]. The results indicate a robust aspiration raising and a weaker voiceless unaspirated raising effect in both Mandarin and Cantonese.

For Research Question 2, the pattern of C-F0 LENGTH was found to follow the pattern of Smallest TONE DIFF (i.e. the F0 difference between the target tone and its closest tone in the tone inventory). However, the C-F0 DIFF within the first 10ms did not follow the patterns of Smallest TONE DIFF or Overall TONE DIFF (i.e. the average value of a target tone's distances from all other tones in the tone inventory). Finally, the cross-linguistic comparison provides support to the hypothesis that higher degree of tone competition may restrict C-F0.

For Research Question 3, the results show that VOT is a strong cue for the aspiration contrast while onset F0 is a weak cue. However, the findings do not provide evidence for a cue trading relation between VOT and onset F0 in Mandarin or Cantonese.

### **7.1.1 Discussion: C-F0 can be conditioned by lexical tones**

#### **7.1.1.1 C-F0 is generally stronger in tones with high onset pitch**

The findings for Question 1 support the hypothesis that the length that C-F0 can extend is conditioned by tones in both Mandarin and Cantonese. To be more specific, the length that C-F0 extends is longer in tones associated with high pitch than low pitch: the effect was longer in M-T1(55) and in M-T4(51) than in M-T3(214). Non-significant C-F0 was found in M-T2(35). In Cantonese, C-F0 in C-T1(55) was longer than C-T2(25), which was longer than in C-T3(33) and C-T5(23). No significant C-F0 was found in C-T4(21). Since C-F0 LENGTH is an aspect of C-F0 magnitudes, these findings show that C-F0 is generally stronger in tones with high onset pitch.

Hombert (1977a, 1977b) also found that C-F0 was weaker for vowels bearing low pitch than for those bearing high pitch from Yoruba speakers and American English speakers. Hombert argued that these findings could not be explained by the tongue pull theory that assumed the correlation of tongue and pitch height (Ladefoged, 1968): the tongue pull theory is based on the assumption that the tongue is high for producing high vowels, which exerts extra tension to the larynx, and thus increases the tension of the vocal folds, which gives rise to higher pitch. By this logic, since the larynx is already in higher position for high tones than for low tones, the tension exerted by the tongue will be less in the high tone context. Assuming a linear relationship between tension and larynx elevation, as well as a correlation between larynx height and F0 (Ewan & Krones, 1974; Ohala & Ewan, 1973), less tension gives rise to less larynx elevation, which predicts that F0 difference would be smaller in a high tone context than in a low tone one, meaning C-F0 would be weaker in vowels with a high tone. However, Hombert's and our findings in this dissertation have shown the opposite from this prediction.

The stronger C-F0 in high tones can also be explained by the saliency of high tones compared to low tones, and therefore high tones allow more F0 variability caused by C-F0 than low tones. There are observations that speakers of tone languages actively minimize C-F0 to make each tone maximally perceptually distinct (Francis et al., 2006; Hombert, 1977a). If high tones are more salient than low tones, there will be less need for speakers to minimize C-F0 for high tones than for low tones. The following evidence can show that high tones are more salient than low tones: phonologically, Chen (2000:191-192) found that for disyllabic compounds in New Chongming and Xining, there was a tendency for preserving a high tone and deleting a non-high tone in a tonal deletion process. Chen analyzed that this was a consequence of the ‘uncontroversial’ tonal saliency of high tones. Perceptually, studies have shown that non-tonal listeners have a higher accuracy rates in perceiving M-T1(55) and M-T4(51) than M-T2(35) and M-T3(214) (Ding et al., 2011; Hao, 2012; Kiriloff, 1969; Lee et al., 2010; Shi, 2007; So & Best, 2010). Hao (2012) also revealed that L2 non-native speakers performed better in mimicking M-T1(55) and M-T4(51) than the other two tones, which may be due to the perceptual saliency of these two tones with high onset pitch. Given that high tones are more salient than low tones phonologically and phonetically, stronger C-F0 in high tone context can be seen as less restricted force on F0 variance by high tones due to their high saliency.

#### **7.1.1.2 C-F0 DIRECTION is generally consistent across tones**

The general C-F0 DIRECTION is consistent in Mandarin and Cantonese tone contexts, except in C-T2(25). A general voiceless raising effect was found across all Mandarin tones and C-T1(55) and C-T3(33): F0[aspirated] started the highest and did not converge with F0[unaspirated] and F0[sonorant] until a relatively late stage. F0[unaspirated] started the second highest and quickly converged with the lowest F0[sonorant] trajectory.



The raising effect of voiceless obstruents has been accounted in a few studies: Halle & Stevens (1971) proposed that for the raising effect of voiceless consonants is due to the increasing stiffness of the vocal cords, which makes the coupling between the upper and lower edges of the vocal cords larger. Since the vocal folds are associated with larynx muscles, this causes vertical larynx movements, in that the larynx raising and lowering are concomitants of the stiffening and slackening of the vocal folds. This would predict a rise in the frequency of any glottal vibration that occur at the onset of a following vowel, and thus raise the F0 values. The vertical larynx movement is later shown to influence the stiffening of the vocal folds and give rise to the rotation of the cricoid cartilage and vocal fold tension changes (Honda, 1995; Honda et al., 1999). Honda et al. (1999) found that in the high F0 range, the larynx height remained relatively constant. In the low F0 range, the entire larynx moved vertically, and the cricoid cartilage rotated along the cervical lordosis, which indicates an effective F0 lowering instead of raising. Löfqvist et al. (1989), on the other hand, have explored the control of voicelessness related to the role of changes in the longitudinal tension of the vocal folds, as indicated by cricothyroid (CT) muscle activity through electromyographic recordings. CT activity associated with the voiced and voiceless consonants indicates a higher level during the closure for both voiceless unaspirated stops and voiceless aspirated stops than for their voiced cognates. Consideration of the relative timing of this gesture suggests that the differences most likely reflect control of vocal-fold tension for suppressing phonatory vibrations, which result in difference in F0 following voiced and voiceless consonants.

As for why F0[aspirated] is higher than F0[unaspirated] in Mandarin and Cantonese, aspiration can be associated with higher transglottal pressure with which F0 varies proportionally, which results in a higher F0 after an aspirated stop than an unaspirated cognate

(Hombert, 1975). Ewan (1979) found a positive correlation between F0 and larynx height. Also, the larynx and F0 were higher after aspirated stops than after unaspirated stops. Ladefoged (1974) reported that the airflow rate is faster after aspirated stops than after unaspirated cognates, which suggests a decrease of vertical pressure perpendicular to the vocal folds and gives rise to tightening of the vocal folds, and thus raises F0 after aspirated stops. Zee (1980), on the other hand, found the intensity of voicing after aspirated stops is lower than after unaspirated stops, which indicates lower subglottal pressure after aspirated stops. Zee suggested a higher F0 may be produced even with lower subglottal pressure. He also suspected that a higher F0 after aspirated can be related to airflow, larynx height, glottal aperture, and vocal fold length at the onset of the following vowel.

### **7.1.2 Discussion: C-F0 can be restricted by competition of tones**

Question 2 asked whether the pattern of C-F0 magnitude followed the pattern of Smallest TONE DIFF (i.e. the F0 difference between the target tone and its closest tone in the tone inventory) or Overall TONE DIFF (i.e. the average value of a target tone's distances from all other tones in the tone inventory).

For Mandarin and Cantonese, the pattern of Smallest TONE DIFF was consistent with the pattern of C-F0 LENGTH(TH-N) but not C-F0 LENGTH(T-N). Neither the pattern of Smallest TONE DIFF nor Overall TONE DIFF predicted the pattern of C-F0 DIFF. No significant positive correlation was found between Smallest TONE DIFF and C-F0 DIFF in any of the tones in Mandarin or Cantonese, except that a negative correlation was found in C-T1(55).

For the cross-linguistic comparison, C-F0 for C-T1(55) is predicted to be less restricted than C-F0 for M-T1(55), since C-T1(55) has no neighbor in the high pitch space while M-T1(55) does. The comparison shows that Cantonese has longer C-F0 LENGTH and a larger C-F0 DIFF

than Mandarin in the high-level tone context. The findings indicate that magnitudes of C-F0 can be restricted by how intense the competition of tone contrasts is.

As mentioned in Chapters 1 and 5, the enhancement account (Kingston, 2007; Kluender et al., 1988; Van Summers, 1987) predicts that C-F0 should be restricted to enhance tone contrasts, since large variance of F0 values caused by consonantal effects in one tone context may trespass the F0 range of another tone. Based on this, the finding that the magnitudes of C-F0 followed the patterns of Smallest TONE DIFF instead of Overall TONE DIFF may be due to that Smallest TONE DIFF serves as a better indicator for the crucial tone contrast at the onset position than Overall TONE DIFF. That is, the crucial tone contrast that may restrict C-F0 in a certain tone is between that target tone and another tone that is the least distant from it, rather than between the target tone and all other tones in the tone inventory.

The finding that only C-F0 LENGTH(TH-N) followed Smallest TONE DIFF may also support the enhancement account of a restriction force on C-F0. Recall that the C-F0 direction is generally consistent across tones: F0[aspirated] started the highest and did not converge with F0[sonorant] until a relatively late stage. F0[unaspirated] started the second highest and quickly converged with the lowest F0[sonorant] trajectory. This shows that the high F0[aspirated] is more of a ‘threat’ for invading another tone territory than the relatively lower F0[unaspirated]. According to a controlled articulation account, in a tone context that has another tone rival that is close, F0[aspirated] of the tone in question should be restricted from the closet tone rival (so that trespassing can be avoided), or C-F0 LENGTH(TH-N) should be restricted according to that tone rival (so that any trespassing will not be too long to harm the tone contrast). By contrast, since F0[unaspirated] is less of a threat to the tone category, neither the C-F0 DIFF or C-F0 LENGTH needs to be as controlled as F0[aspirated]. This account can be further tested in a non-tonal

language that has an aspiration contrast, where tone contrast is not a factor restricting C-F0, so F0[aspirated] is predicted to not vary in different pitch environments.

Though our results support that C-F0 can be restricted by competitions of tones, they also show that this restricting force is not strong, as only C-F0 LENGTH but not C-F0 DIFF followed the pattern of Smallest TONE DIFF. Also, no significant positive correlation was found between Smallest TONE DIFF and C-F0 DIFF in any of the tones in Mandarin or Cantonese. One possible reason can be that competition of tones at the beginning of the vowel is not strong or crucial enough to inhibit C-F0. First of all, F0 values in other positions of the tone are also crucial for cuing tone contrasts. Barry & Blamey (2004) defined tone space as  $F0 \text{ offset} \times F0 \text{ onset}$  for a series of tokens for each of the six tones in Cantonese, allowing them to capture more dynamic factors, such as direction and slope, to depict the tone contrasts. Furthermore, though F0 height and F0 contour have been argued to be the most important cues over other acoustic features for native speakers to judge Mandarin tones or Cantonese tones (Gandour & Harshman, 1978; Khouw & Ciocca, 2007; Massaro et al., 1985), other factors are also considered crucial for tone contrasts, including amplitude contour (Whalen & Xu, 1992), voice quality (Garding et al., 1986; Mok et al., 2013), and duration (Blicher et al., 1990). This can explain that competitions of tones in the onset position are probably not strong or crucial enough to result in a significant inhibition force on C-F0 DIFF or give rise to significant positive correlations between Smallest TONE DIFF and C-F0 DIFF.

### **7.1.3 Discussion: C-F0 and cue trading for the consonant laryngeal contrast**

Question 3 asked whether there was a trading relation between VOT and F0 as cues for the aspiration contrast of the initial consonants. Positive correlations were found in the aspirated category in M-T1(55), M-T3(214) and M-T4(51), while a significant negative correlation

between VOT and onset F0 for the unaspirated category was found in M-T2(35), with a poor fit ( $R^2=0.05$ ). No significant correlation between VOT and F0 was found in any of the Cantonese tone contexts. These findings did not provide strong evidence for a cue trading relation between VOT and onset F0 in Mandarin or Cantonese.

Our results are in line with that in Kirby & Ladd (2016), who found no significant negative covariates between VOT and F0 within each voicing category in French and Italian, with considerable individual differences in the direction of the correlation observed. Our findings are also compatible with those in Dmitrieva et al. (2015), where within-category VOT is found uncorrelated with onset F0 in English and Spanish. The findings that F0 and VOT are not significantly correlated within the same voicing or aspiration category can shed light on the enhancement account for C-F0, which assumes that (1) onset F0 variation is governed by phonological contrast enhancement, aiming at making contrasting categories more perceptually distinct and (2) cues to the phonological contrast exist in a trading relation, in which a decrease of the value of one phonetic cue for a phonological contrast can be offset by an increase in the value of another cue (Kirby, 2010, 2013; Repp, 1982). Thus, the enhancement account predicts that the secondary cue for the voicing or aspiration contrast, such as onset F0, will be strengthened to compensate the weakening of the primary cue VOT. If this enhancement strategy is implemented by the speakers across the range of VOT values within the same voicing/aspiration category, we would expect to see a negative correlation between VOT and onset F0 in voiceless stops: as the positive VOT decreases, onset F0 is predicted to raise to offset the reduction of VOT. However, our results do not show such a negative correlation, which provides support against the enhancement approach for C-F0.

Our results have implications for experience-based theories of cue perception and integration. An experience-based account would suggest that multiple acoustic properties (e.g. F0 and VOT) are learned by listeners and integrated into cues to a specific contrast (e.g. the aspiration contrast) because those properties covary in the language (Dmitrieva et al., 2015; Holt et al., 2001). The experience-based account can be strengthened if the finding in production is consistent with that in perception, i.e. if the acoustic properties are not integrated in the perception of a contrast, the experience-based explanation is more likely if the speakers do not covary the cues in production, so that they are unlikely to learn to integrate the cues in perception from their language experience (Dmitrieva et al., 2015; Llanos et al., 2013). In other words, the experience-based account predicts that the usage of cues in perception is originated from the cue usage in production, i.e. the ‘experience’. Therefore, the experience-based account predicts that the findings of the trading relation of VOT and F0 should be consistent in perception and production. On the other hand, the experience-based account will face problems if speakers do not trade the cues in production and thus lack the experience for exposure in the language, but still have cue-trading in perception. In other words, if cue-trading is found in perception but not in production, it indicates that the source of cue-trading in perception is not originated from the experience (because there is none produced in speaking).

Our results have shown that there is no strong evidence showing that Cantonese or Mandarin speakers exploit onset F0 to distinguish aspirated and unaspirated stops, nor is there any significant trading relation between VOT and onset F0. However, some perception studies have shown that listeners can employ VOT and onset F0 to perceive aspiration distinction: Francis et al. (2006) found that Cantonese listeners could use differences in onset F0 to cue perception of the voicing contrast, but the minimum extent of F0 perturbation necessary for this

was greater than C-F0 that usually found in tonal languages. That is, their Cantonese participants could only hear the aspiration distinction when C-F0 extended to the longest duration into the vowel (=80ms) in their experiment setting, whereas C-F0 usually extends to approximately 30-50ms into the vowels according (Francis et al., 2006; Gandour, 1974; Hombert, 1977a). Likewise, Yu (2017) recently presented his results of a perception study on how Mandarin listeners integrated onset F0 and VOT to distinguish aspiration contrast of Mandarin stops. He found a negative correlation between these two cues in the perception test: as the VOT cue weakens, listeners put more weight on onset F0 to offset the weakening of the VOT values. Taking both these perception findings and our production results into consideration: while the listeners can exploit VOT and onset pitch in perceiving the aspiration distinction in Mandarin and Cantonese, our production experiments suggest that the speakers of these two languages actually do not employ onset F0 in production and thus there is a lack of experience of using onset F0 as a cue for the aspiration contrast in these languages.

## **7.2 Implications for the enhancement account and the automatic account for C-F0**

As introduced in the first chapter, there are at least two possible interpretations for C-F0: the enhancement account and the automatic account: the feature enhancement hypothesis regards C-F0 as a consequence of controlled phonetic implementation for the functional purpose of reinforcing phonological contrastive features (Kingston & Diehl, 1994). However, it does not predict trading relations between cues for the contrast, such as onset F0 and VOT to the relevant voicing specification. On the other hand, the probabilistic enhancement hypothesis (Kirby, 2010, 2013: 232) can predict the trading relation between cues. It claims that phonetic cues, such as VOT and onset F0, can be seen as probabilistic functions of contrast precision and the degrees of cue weights. The contrast precision is based on the statistical distribution of acoustic-phonetic

cues to the contrast and does not directly get categorically enhanced by the values of cues. A negative correlation between the weights of cues can be expected, when the contrast is maintained in a trading relationship between cues, i.e. the less informative one cue is, the more informative the other competing cue.

The automatic account, on the other hand, denies a teleological explanation for C-F0. Instead, C-F0 is an unintended by-product arising from articulatory factors (such as larynx activities and cricothyroid activities) and aerodynamic factors (such as the degree of transglottal air flow, transglottal pressure, and subglottal pressure).

Question 1 investigated the consistency of C-F0 in different tone contexts. An enhancement account that only focuses on the enhancement of voicing feature of the initial consonants would predict that the magnitudes of C-F0 should be similar in all pitch context, based on the rationale that speakers intentionally control onset F0 to enhance the laryngeal feature of the consonants (Chen, 2011). An enhancement account that considers both enhancing the voicing features of the consonants and the tone contrast at the onset position (i.e. High, Mid and Low) would predict that the magnitudes of C-F0 vary in different tone contexts, but tones that share the same tone feature should have similar C-F0. The automatic account, on the other hand, would allow more possible C-F0 magnitudes as the articulatory and aerodynamic consequence of stop and tone production. It predicts that F0[aspirated] should be higher than F0[unaspirated], as a result of higher larynx height, higher transglottal pressure and faster airflow for producing aspirated stops than for unaspirated stops. There is also a possibility for a hybrid account: both enhancement and automatic mechanisms have contributed to C-F0.

Our results provide some supports for the enhancement account in a certain tonal context: the magnitudes of C-F0 are not consistent across all tone contexts, but are generally stronger in



tones with high onset pitch, although the direction of C-F0 was very consistent. For Mandarin, the effect was longer in M-T1(55) and in M-T4(51) than in M-T3(214). C-F0 was not significant in M-T2(35). For Cantonese, C-F0 in C-T1(55) was longer than C-T2(25), C-T3(33) and C-T5(23). No significant C-F0 was found in C-T4(21). As discussed above, stronger C-F0 in high pitch environment can be explained as consequence of greater saliency of High pitch, which requires weaker restrictions of variability of F0 values caused by C-F0, comparing to the less salient mid or low pitch, which needs to restrict F0 values to enhance the Mid or Low feature. However, an automatic account cannot be fully ruled out: for example, it can be because larynx is elevated when Mandarin and Cantonese speakers are producing high pitch, which can give rise to greater tension and longer effect time. This hypothesis needs further direct investigation on specific physical properties in question.

The results of Question 2 have implications for whether the tone enhancement can influence C-F0. As discussed above, the tone enhancement account is supported by the findings that the patterns of C-F0 LENGTH(TH-N) followed the patterns of Smallest TONE DIFF and that Cantonese has longer C-F0 LENGTH and a larger C-F0 DIFF than Mandarin in the high-level tone context. Nevertheless, the restricting force for enhancing the tone contrast turns out to be weak, as only C-F0 LENGTH but not C-F0 DIFF followed the pattern of Smallest TONE DIFF, and no significant positive correlation was found between Smallest TONE DIFF and C-F0 DIFF. One possible reason is that the enhancement of tone contrast at the beginning of the vowel is not crucial enough to restrict C-F0. That is, unlike contrast of the entire consonants that can be enhanced by the following onset F0, tones are dynamic with other factors like direction and slope, so onset F0 may not be enough for cuing the contrast of the entire tone. Furthermore, it is unclear how the automatic account can explain our results, for the production of high-level M-

T1(55) and C-T1(55) share the same articulation setting, but facing different rivals in their own tone inventory, i.e. M-T1(55) needs to compete with M-T4(51) in Mandarin while there is no other tone with high onset as C-T1(55) in Cantonese. Therefore, our results for Question 2 support the enhancement of tone contrast account for C-F0 rather than the automatic account.

Question 3 can shed light on the probabilistic enhancement hypothesis, which hypothesizes that when the precision of a contrast along one acoustic dimension is reduced, other dimensions may be enhanced to compensate (Kirby 2010; 2013). Assuming a statistical distribution of acoustic-phonetic cues to the phonological contrast, such a model of trading relation claims that the contrast does not get categorically enhanced by the values of cues, but through the degree of precision influenced by the number of cues competing over some acoustic-phonetic space. Therefore, the probabilistic enhancement hypothesis predicts negative correlations between onset F0 and VOT as cues for voicing specification, whereas the feature enhancement hypothesis does not make such predictions. The automatic account, on the other hand, may predict positive correlations between onset F0 and VOT, for the longer the release, the longer time for transglottal airflow to boost F0 values. Our production results for Question 3 show some support for the automatic account but not the probabilistic enhancement hypothesis: only positive correlations were found in the aspirated category in M-T1(55), M-T3(214) and M-T4(51). While a significant negative correlation between VOT and onset F0 for the unaspirated category was found in M-T2(35), and it had a poor fit ( $R^2=0.05$ ). Nevertheless, as discussed before, some perception studies have shown that listeners can exploit VOT and onset F0 to perceive aspiration distinction in Mandarin and Cantonese (Francis et al., 2006; Yu, 2017). It is possible that listeners only employ the enhancement strategy in perception but not in production.

Further research that incorporate both production and perception tests is needed to investigate the observed asymmetry.

### 7.3 Conclusion

Table 43 has summarized the predictions respectively made by the enhancement account and the automatic account, as well as the major findings in this dissertation.

	Predictions by the Enhancement Account	Predictions by the Automatic Account	Major findings
Question 1: C-F0 conditioned by tones?	Enhance consonant contrast only: consistent C-F0 in all pitch context Enhance both consonant and tone contrasts: consistent C-F0	More possible patterns of C-F0 magnitudes as the articulatory and aerodynamic consequence of stop and tone production are expected	The magnitudes of C-F0 are not consistent across all tone contexts, but are generally stronger in tones with high onset pitch, although the direction of C-F0 was very consistent.
Question 2: C-F0 conditioned by tone contrasts?	C-F0 should be restricted to enhance tone contrasts, for which F0 differences between tones are used as indicators	The magnitudes of C-F0 should be similar in high-level M-T1(55) and C-T1(55), which share the same articulation setting	The pattern of Smallest TONE DIFF was consistent with the pattern of C-F0 LENGTH(TH-N); Cantonese has longer C-F0 LENGTH and a larger C-F0 DIFF than Mandarin in the high-level tone context
Question 3: a trading relation between F0 and VOT as cues for laryngeal contrasts?	The probabilistic enhancement hypothesis: negative correlations between onset F0 and VOT as cues for voicing specification	Positive correlations between onset F0 and VOT are expected	Positive correlations between onset F0 and VOT were found in the aspirated category in M-T1(55), M-T3(214) and M-T4(51); negative correlations were found for the unaspirated category in M-T2(35) with a poor fit ( $R^2=0.05$ )

Table 43. A summary of predictions by the enhancement account and the automatic account respectively and the major findings

Our major findings for Question 1 have provided evidence against the enhancement account that considers enhancing voicing contrast as the only teleological cause for C-F0. The enhancement account that considers both voicing contrast and tone contrast is supported by the results that C-F0 are not consistent across all tone contexts but are generally stronger in tones with high onset pitch. However, a hybrid account of both the functional explanation and the automatic account cannot be ruled out by our results.

The results for Question 2 have given some support for the tone enhancement account that expects C-F0 to be restricted for enhancing tone contrasts. The tone automatic account that assumes only the tone articulation matters is argue against by the cross-linguistic comparison where Cantonese has longer C-F0 LENGTH and a larger C-F0 DIFF than Mandarin in the high-level tone context.

The results for Question 3 have provided support for the consonant automatic account, which predicts positive correlations between onset F0 and VOT, but against the consonant probabilistic enhancement account.

To conclude, this dissertation has offered a new angle that few previous studies have dealt with for the topics of consonantal effects on F0 of adjacent vowel in tonal languages: previous studies have only considered whether it is the enhancement of voicing contrast or the physical properties of producing voicing that give rise to C-F0. This dissertation has introduced a new contrast, i.e. the contrast of lexical tones, to explore the question. Furthermore, the results for all three questions confirm a hybrid account: the tone enhancement account and the consonant automatic account. The dissertation also provides implications for future studies to further investigation on what physical properties or other factors, such as cognitive factors, are responsible for the reported major findings.

## APPENDICES

## APPENDIX A

### Participant Information

Participant No.	Native Language	From where	Age	Gender
M1	Standard Mandarin	Beijing	19	Female
M2	Standard Mandarin	Beijing	19	Female
M3	Standard Mandarin	Beijing	24	Female
M4	Standard Mandarin	Beijing	26	Female
M5	Standard Mandarin	Beijing	29	Female
M6	Standard Mandarin	Beijing	30	Female
M7	Standard Mandarin	Beijing	19	Female
M8	Standard Mandarin	Shenzhen	29	Female
M9	Standard Mandarin	Shenzhen	27	Female
M10	Standard Mandarin	Shenzhen	33	Female
M11	Standard Mandarin	Shenzhen	25	Female
M12	Standard Mandarin	Shenzhen	25	Female
M13	Standard Mandarin	Shenzhen	24	Female
M14	Standard Mandarin	Shenzhen	29	Female
M15	Standard Mandarin	Shenzhen	24	Female
C1	Cantonese	Shenzhen	26	Female
C2	Cantonese	Shenzhen	32	Female
C3	Cantonese	Zhuhai	25	Female
C4	Cantonese	Shenzhen	24	Female
C5	Cantonese	Shenzhen	25	Female
C6	Cantonese	Shenzhen	27	Female
C7	Cantonese	Shenzhen	26	Female
C8	Cantonese	Shenzhen	20	Female
C9	Cantonese	Shenzhen	22	Female
C10	Cantonese	Shenzhen	20	Female
C11	Cantonese	Shenzhen	23	Female
C12	Cantonese	Shenzhen	28	Female
C13	Cantonese	Shenzhen	30	Female
C14	Cantonese	Shenzhen	35	Female
C15	Cantonese	Shenzhen	31	Female

Table 44. Participant information

# APPENDIX B

## Mandarin Stimuli

POA	Tone	Unaspirated	Aspirated	Nasal	Liquid
Bilabial	1(55)	pi 逼, pa 巴	p <sup>hi</sup> 批, p <sup>ha</sup> 趴, p <sup>hu</sup> 铺	mi 咪, ma 吗	
	2(35)	pi 鼻, pa 拔, pu 不	p <sup>hi</sup> 皮, p <sup>ha</sup> 爬, p <sup>hu</sup> 葡	mi 迷, ma 麻	
	3(214)	pi 比, pa 把, pu 补	p <sup>hi</sup> 痞, p <sup>hu</sup> 谱	mi 米, ma 马, mu 母	
	4(51)	pi 必, pa 爸, pu 布	p <sup>hi</sup> 屁, p <sup>ha</sup> 怕, p <sup>hu</sup> 瀑	mi 密, ma 骂, mu 木	
Alveolar	1(55)	ti 低, ta 搭, tu 嘟	t <sup>hi</sup> 踢, t <sup>ha</sup> 它, t <sup>hu</sup> 秃	ni 呢	li 哩, la 拉, lu 噜
	2(35)	ti 迪, ta 达, tu 读	t <sup>hi</sup> 题, t <sup>hu</sup> 图	ni 泥, na 拿, nu 奴	li 离, lu 庐
	3(214)	ti 底, ta 打, tu 赌	t <sup>hi</sup> 体, t <sup>ha</sup> 塔, t <sup>hu</sup> 土	ni 你, na 哪	li 李, lu 鲁
	4(51)	ti 地, ta 大, tu 肚	t <sup>hi</sup> 替, t <sup>ha</sup> 踏, t <sup>hu</sup> 兔	ni 腻, na 纳, nu 怒	li 力, la 辣, lu 路
Velar	1(55)	kə 鸽, ku 姑	k <sup>hə</sup> 科, k <sup>hu</sup> 哭		
	2(35)	kə 隔,	k <sup>hə</sup> 壳		
	3(214)	kə 葛, ku 古	k <sup>hə</sup> 渴, k <sup>hu</sup> 苦		
	4(51)	kə 各, ku 顾	k <sup>hə</sup> 客, k <sup>hu</sup> 裤		

Table 45. Mandarin stimuli

# APPENDIX C

## Cantonese Stimuli

POA	Tone	Unaspirated	Aspirated	Nasal	Liquid
Bilabial	1(55)	pa 巴, pɔ 波, pei 悲	p <sup>h</sup> a 趴, p <sup>h</sup> ɔ 坡, p <sup>h</sup> ei 呔	ma 吗, mɔ 魔, mei 眯	
	2(25)	pa 把, pei 比	p <sup>h</sup> a 扒, p <sup>h</sup> ɔ 颇, p <sup>h</sup> ei 鄙	mɔ 摸	
	3(33)	pa 霸, pɔ 播 pei 臂	p <sup>h</sup> a 怕, p <sup>h</sup> ɔ 破, p <sup>h</sup> ei 屁		
	4(21)		p <sup>h</sup> a 爬, p <sup>h</sup> ɔ 婆, p <sup>h</sup> ei 皮	ma 麻, mɔ 磨, mei 微	
	5(23)		p <sup>h</sup> ei 婢	mei 美	
	6(22)	pa 罢, pei 鼻		ma 骂, mei 味	
Alveolar	1(55)	ta 叻, tɔ 多, tɔu 都	t <sup>h</sup> a 它, t <sup>h</sup> ɔ 拖, t <sup>h</sup> ɔu 涛		la 啦, lɔ 咯
	2(25)	tɔ 躲, tɔu 赌	t <sup>h</sup> ɔu 土		lɔ 裸
	3(33)	tɔu 到	t <sup>h</sup> ɔu 套		
	4(21)		t <sup>h</sup> ɔ 驼, t <sup>h</sup> ɔu 逃	nɔ 挪, nou 奴, nei 尼	lɔ 罗, lei 厘, lou 牢
	5(23)		t <sup>h</sup> ɔu 肚	nou 恼, nei 你	lei 李, lou 老
	6(22)	tɔ 堕, tɔu 盗, tei 地		nɔ 糯, nou 怒, nei 腻	lei 利
Velar	1(55)	ka 家, kɔ 哥, kei 机	k <sup>h</sup> a 卡	nga 鸦, ngo 厠	
	2(25)	ka 假		nga 哑	
	3(33)	ka 嫁		nga 亚	
	4(21)				
	5(23)				
	6(22)				

Table 46. Cantonese stimuli



## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Alexander, J. A. (2010). *The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems* (Doctoral Dissertation). Northwestern University, Evanston, IL.
- Barrie, M. (2007). Contour tones and contrast in Chinese languages. *Journal of East Asian Linguistics*, 16(4), 337–362.
- Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *The Journal of the Acoustical Society of America*, 116(3), 1739–1748.
- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin; New York: Mouton de Gruyter.
- Beckman, J., Jessen, M., & Ringen, C. (2009). German fricatives: Coda devoicing or positional faithfulness? *Phonology*, 26(02), 231–268.
- Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, 49(02), 259–284.
- Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18(1), 37–49.
- Boersma, P., & Weenink, D. (2015). *Praat, a system for doing phonetics by computer* [Version 6.0.05]. Retrieved from <http://www.praat.org/>
- Caisse, M. (1982). *Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops* (MA Thesis). University of California, Berkeley, CA.
- Chao, H.-J. (1992). *Aspiration in Chinese* (Doctoral Dissertation). University of Illinois at Urbana-Champaign, Champaign, IL.
- Chao, K.-Y., & Chen, L. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing*, 13(2), 215–232.
- Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, 45, 24–27.
- Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.
- Chen, M. (2011). *Prevocalic Aspiration and Mandarin Tones* (MA Thesis). National Chiao Tung University, Hsinchu, Taiwan.

- Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.
- Chen, Y. & Ng, M. (2005). Examination of voicing onset time during Mandarin tone productions. *The Journal of the Acoustical Society of America*, 117(4), 2459–2459.
- Chen, Y. (2011). How does phonology guide phonetics in segment–f<sub>0</sub> interaction? *Journal of Phonetics*, 39(4), 612–625.
- Cheung, K. H. (1986). *The present day phonology of Cantonese* (Doctoral Dissertation). University of London, London, United Kingdom.
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27(2), 207–229.
- Collier, R. (1974). Laryngeal muscle activity, subglottal air pressure, and the control of pitch in speech. *Haskins Laboratory Status Report on Speech Research*, 137–170.
- Connell, B. (2002). Tone languages and the universality of intrinsic F<sub>0</sub>: evidence from Africa. *Journal of Phonetics*, 30(1), 101–129.
- Deterding, D., & Nolan, F. (2007). Aspiration and voicing of Chinese and English plosives. In *Proceedings of the 16th International Congress of Phonetic Sciences, Universität des Saarlandes Saarbrücken, Germany*, 385–388.
- Diehl, R. L. (2008). Acoustic and auditory phonetics: the adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1493), 965–978.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, 1(2), 121–144.
- Ding, H., Hoffmann, R., & Jokisch, O. (2011). An investigation of tone perception and production in German learners of Mandarin. *Archives of Acoustics*, 36(3), 509–518.
- Dixit, R. P., & MacNeilage, P. F. (1980). Cricothyroid activity and control of voicing in Hindi stops and affricates. *Phonetica*, 37(5–6), 397–406.
- Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f<sub>0</sub> as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49, 77–95.
- Duanmu, S. (1994). Against contour tone units. *Linguistic Inquiry*, 25(4), 555–608.
- Duanmu, S. (1999). Metrical structure and tone: evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics*, 8(1), 1–38.
- Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford: Oxford University Press.

- Durvasula, K., Huang, H.-H., Uehara, S., Luo, Q., & Lin, Y.-H. (2018). Phonology Modulates the Illusory Vowels in Perceptual Illusions. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 9(1), 7.
- Dutta, I. (2007). Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt (Doctoral Dissertation). University of Illinois at Urbana-Champaign, Champaign, IL.
- Erickson, D., & Abramson, A. S. (2013). F0, EMG and Tonogenesis in Thai. *Journal of Nagoya Gakuin University (Language and Culture): collected papers in honor of Prof. Katsumasa Shimizu*, 24, 1-13.
- Ewan, W. G. (1979). Laryngeal Behavior in Speech. *Report of the Phonology Laboratory* (3). University of California, Berkeley.
- Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyrohyoidmeter. *Journal of Phonetics*, 2, 327–335.
- Ewan, W. G. (1976). *Laryngeal behavior in speech* (Doctoral Dissertation). University of California, Berkeley, CA.
- Fischer-Jørgensen, E. (1968). Les occlusives françaises et danoises d'un sujet bilingue. *Word*, 24, 112–153.
- Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *Journal of Acoustical Society of America*, 120, 2884–2895.
- Fung, R. S. Y., & Wong, C. S. P. (2010). Mergers and near-mergers in Hong Kong Cantonese tones. *The Fourth European Conference on Tone and Intonation (TIE4)*. Stockholm, Sweden.
- Gandour, J. (1974). Consonant types and tone in Siamese. *Journal of Phonetics*, 2, 337–350.
- Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, 21(1), 1–33.
- Garding, E., Kratochvil, P., Svantesson, J.-O., & Zhang, J. (1986). Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech*, 29(3), 281–293.
- Halle, M., & Stevens, K. (1971). A Note on Laryngeal Features. *MIT-RLE Quarterly Progress Report*, 101, 198–214.
- Hanson, H. M., & Stevens, K. N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLSyn. *The Journal of the Acoustical Society of America*, 112(3), 1158–1182.

- Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269–279.
- Harris, J. (1994). *English sound structure*. Oxford: Blackwell Publishing.
- Hashimoto, O. Y. (1972). *Phonology of Cantonese*. Cambridge: Cambridge University Press.
- Hirose, H., Lee, C.-Y., & Ushijima, T. (1974). Laryngeal control in Korean stop production. *Journal of Phonetics*, 2, 145–152.
- Hirose, H., Lisker, L., & Abramson, A. S. (1973). Physiological aspects of certain laryngeal features in stop production. *The Journal of the Acoustical Society of America*, 53(1), 294–295.
- Hirose, H., & Ushijima, T. (1978). Laryngeal control for voicing distinction in Japanese consonant production. *Phonetica*, 35(1), 1–10.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *The Journal of the Acoustical Society of America*, 109(2), 764–774.
- Hombert, J. M. (1978). Consonant types, vowel quality, and tone. *Tone: A Linguistic Survey*, 77–111.
- Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic Explanations for the Development of Tones. *Language*, 55(1), 37–58.
- Hombert, J. M. (1975). *Towards a theory of tonogenesis: an empirical, physiologically and perceptually based account of the development of tonal contrasts in languages* (Doctoral Dissertation). University of California, Berkeley, CA.
- Hombert, J. M. (1977a). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, 8(2), 173.
- Hombert, J.-M. (1977b). Tone space and universals of tone systems. *The Journal of the Acoustical Society of America*, 61(1), 89.
- Honda, K. (1995). Laryngeal and extra-laryngeal mechanisms of F0 control. *Producing Speech: Contemporary Issues*, 215–232.
- Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech*, 42(4), 401–411.
- Honeybone, P. (2005). Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. *The Internal Organization of Phonological Segments*, 319, 54.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

- Hyman, L. M. (1976). Phonologization. In A. Juilland (ed.), *Linguistic studies presented to Joseph H. Greenberg* (407-418). Saratoga, CA: Anna Libri.
- Hyman, L. M. (2013). Enlarging the scope of phonologization. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (3-28). Oxford: Oxford University Press.
- Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, 12(03), 369-396.
- Jeel, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. *Annual Report of the Institute of Phonetics, University of Copenhagen*, 9, 191-211.
- Jessen, M. (2001). Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. *Distinctive Feature Theory*, 2, 237.
- Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, 19(02), 189-218.
- Jun, S.-A. (1996). Influence of Microprosody on Macroprosody: A Case of Phrase Initial Strengthening. *University of California Working Papers in Phonetics*, 92, 97-116.
- Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 286-319.
- Keating, P., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11(3), 277-290.
- Keyser, S. J., & Stevens, K. N. (2001). Enhancement revisited. In M. Kenstowicz & K. Hale (Eds.): *A life in language* (271-291). Cambridge, MA: MIT Press.
- Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language*, 82(1), 33-63.
- Khouw, E., & Ciocca, V. (2007). An acoustic and perceptual study of initial stops produced by profoundly hearing impaired adolescents. *Clinical Linguistics & Phonetics*, 21(1), 13-27.
- Kim, M.-R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30(1), 77-100.
- Kingston, J. (1985). *The linguistic use of vertical larynx movement* (MA Thesis). University of Texas at Austin, Austin, TX.
- Kingston, J. (2007). Segmental influences on F0: Automatic or controlled? In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes* (171-210). Berlin: Mouton de Gruyter.
- Kingston, J. (2011). Tonogenesis. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (2304-2333). Oxford: Blackwell Publishing.

- Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language*, 70(3), 419–454.
- Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, 36(1), 28–54.
- Kirby, J. P. (2010). *Cue selection and category restructuring in sound change* (Doctoral Dissertation). The University of Chicago, Chicago, IL.
- Kirby, J. P. (2013). The role of probabilistic enhancement in phonologization. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (228–246). Oxford: Oxford University Press.
- Kirby, J. P., & Ladd, B. (2015). Stop voicing and F0 perturbations: evidence from French and Italian. *The Scottish Consortium for ICPHS*.
- Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F 0: Evidence from “true voicing” languages a. *The Journal of the Acoustical Society of America*, 140(4), 2400–2411.
- Kirby, J. P., & Yu, A. C. (2007). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. In *Proceedings of the International Congress of the Phonetic Sciences XVI*, 1389–1392.
- Kirilloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, 20(2–4), 63–67.
- Kjellin, O. (1977). Observations on consonant types and “tone” in Tibetan. *Journal of Phonetics*, 5, 317–338.
- Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16, 153–169.
- Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, 41(3), 150–174.
- Kratochvil, P. (1984). Phonetic tone sandhi in Beijing dialect stage speech. *Cahiers de Linguistique-Asie Orientale*, 13(2), 135–174.
- Ladefoged, P. (1968). *A phonetic study of West African languages: An auditory-instrumental survey*. Cambridge: Cambridge University Press.
- Ladefoged, P. (1972). *Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech* (Vol. 15). Oxford: Oxford University Press.
- Ladefoged, P. (1974). *Respiration, laryngeal activity and linguistics*. In *Proceedings of the international symposium on Ventilatory and Phonatory Control Systems*. Oxford: Oxford University Press.

- Lai, Y., Huff, C., & Jongman, A. (2009). The Raising Effect of Aspirated Prevocalic Consonants on F0 in Taiwanese. In *Proceedings of the 2nd International Conference on East Asian Linguistics*.
- Lampp, C., & Reklis, H. (2004). Effects of coda voicing and aspiration on Hindi vowels. *The Journal of the Acoustical Society of America*, 115(5), 2540–2540.
- Lee, C.-Y., Tao, L., & Bond, Z. S. (2010). Identification of acoustically modified Mandarin tones by non-native listeners. *Language and Speech*, 53(2), 217–243.
- Lee, J. L. (2012). The representation of contour tones in Cantonese. In *Annual Meeting of the Berkeley Linguistics Society*, 38, 272–286.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839–862.
- Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge: Cambridge University Press.
- Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental Phonology*, 13–44.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422.
- Liu, H., Ng, M. L., Wan, M., Wang, S., & Zhang, Y. (2008). The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice*, 22(2), 210–218.
- Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, 134, 2214–2224.
- Löfqvist, A. (1975). Intrinsic and extrinsic Fo variations in Swedish tonal accents. *Phonetica*, 31(3–4), 228–247.
- Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America*, 85(3), 1314–1321.
- Löfqvist, A., McGarr, N. S., & Honda, K. (1984). Laryngeal muscles and articulatory control. *The Journal of the Acoustical Society of America*, 76(3), 951–954.
- Maddieson, I. (1984). The effects on F0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics*, 12, 9–15.
- Maddieson, I., & Gandour, J. (1976). Vowel length before aspirated consonants. *UCLA Working Papers in Phonetics*, 31, 46–52.
- Massaro, D. W., Cohen, M. M., & Tseng, C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, 267–289.



- Mielke, J. (2005). Ambivalence and ambiguity in laterals and nasals. *Phonology*, 22(02), 169–203.
- Mok, P. P.-K., & Wong, P. W.-Y. (2010). Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables. In *Speech Prosody 2010 International Conference*.
- Mok, P. P., Zuo, D., & Wong, P. W. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, 25(3), 341–370.
- Ohala, J. J. (1980). Phonological features of Hindi stops. *Report of the Phonology Laboratory (UC, Berkeley)*, 5, 96–105.
- Ohala, J. J. (1970). Aspects of the Control and Production of Speech. *UCLA Working Papers in Phonetics*, 15.
- Ohala, J. J. (1974). A mathematical model of speech aerodynamics. In *Proceedings of the Speech Communication Seminar, Stockholm*, 65–72.
- Ohala, J. J. (1976). A model of speech aerodynamics. *Report of the Phonology Laboratory (Berkeley)*, 1, 93–107.
- Ohala, J. J. (1978). Production of tone. *Tone: A Linguistic Survey*, 5–39.
- Ohala, J. J., & Ewan, W. G. (1973). Speed of pitch change. *The Journal of the Acoustical Society of America*, 53(1), 345–345.
- Ohala, M., & Ohala, J. J. (1972). The problem of aspiration in Hindi phonetics. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo*, 6, 39–46.
- Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, 75(1), 224–230.
- Peirce, J. W. P. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162.1, 8–13.
- Perkins, J. (2014). Non-Local Consonant-Tone Interaction in Thai. *Proceedings of the 4th International Symposium on Tonal Aspects of Languages*, 112–115.
- Pike, E. V. (1948). Problems in Zapotec tone analysis. *International Journal of American Linguistics*, 14(3), 161–170.
- R Core Team. (2015). *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92(1), 81.

- Rochet, B. L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, 19(4), 105–106.
- Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, 6(4), 343–352.
- Shi, J. (2007). On teaching tone three in Mandarin. *Journal of Chinese Language Teachers Association*, 42(2), 1.
- Shih, C. (1987). *The phonetics of the Chinese tonal system*. AT&T Bell Labs Technical Memo.
- Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, 132(2), 95–101.
- Simada, Z., & Hirose, H. (1971). Physiological correlates of Japanese accent patterns. *Annual Bulletin*, 5.
- Slis, I. H. (1970). Articulatory measurements on voiced, voiceless and nasal consonants. *Phonetica*, 21(4), 193–210.
- So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: effects of native phonological and phonetic influences. *Language and Speech*, 53(2), 273–293.
- Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). Cambridge, MA: MIT press.
- Tang, K. E. (2008). *The Phonology and Phonetics of Consonant-Tone Interaction* (Doctoral Dissertation). University of California at Los Angeles, Los Angeles, CA.
- Tsui, Y. H., & Valter, C. (2000). Perception of aspiration and place of articulation of Cantonese initial stops by normal and sensorineural hearing-impaired listeners. *International Journal of Language & Communication Disorders*, 35(4), 507–525.
- Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in Empirical Prosody Research*, 3, 1–28.
- Van Summers, W. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America*, 82(3), 847–863.
- Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica*, 34(2), 93–107.
- Wan, I.-P. (2007). On the phonological organization of Mandarin tones. *Lingua*, 117(10), 1715–1738.
- Wan, I.-P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology*, 15(3), 417–461.

- Wang, W. S.-Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33(2), 93–105.
- Wang, X. (2013). Perception of Mandarin Tones: The Effect of L1 Background and Training. *Modern Language Journal*, 97(1), 144–160.
- Wang, Y., Sereno, J. A., & Jongman, A. (2006). L2 acquisition and processing of Mandarin tone. In E. Bates, L. Tan, & O. Tzeng (Eds.), *Handbook of Chinese Psycholinguistics* (250–256). Cambridge: Cambridge University Press.
- Wetzels, W. L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language*, 207–244.
- Whalen, D.H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23(3), 349–366.
- Whalen, Douglas H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25–47.
- Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, 33(2), 165–181.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61–83.
- Yip, M. (1980). *The tonal phonology of Chinese* (Doctoral Thesis). Cambridge, MA: Massachusetts Institute of Technology.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yu, A. C. L. (2017). *The role of feature-general categorization gradiency in individual differences in speech processing*. Presented at the LSA 2017 Annual Meeting, Austin, TX.
- 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. (1980). The effect of aspiration on the F0 of the following vowel in Cantonese. *UCLA Working Papers in Phonetics*, 49, 90–97.
- Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(1), 153–201.